The Large-Scale Genomic Organization of Repetitive DNA Families at the Telomeres of Rye Chromosomes

Alexander V. Vershinin,[1] Trude Schwarzacher, and John S. Heslop-Harrison

John Innes Centre, Colney Lane, Norwich, NR4 7UH, United Kingdom

Repetitive DNA sequences in the terminal heterochromatin of rye (*Secale cereale*) chromosomes have consequences for the structural and functional organization of chromosomes. The large-scale genomic organization of these regions was studied using the telomeric repeat from Arabidopsis and clones of three nonhomologous, tandemly repeated, subtelomeric DNA families with complex but contrasting higher order structural organizations. Polymerase chain reaction analysis with a single primer showed a fraction of the repeat units of one family organized in a "head-to-head" orientation. Such structures suggest evolution of chromosomes by chromatid-type breakage-fusion-bridge cycles. In situ hybridization and pulse field gel electrophoresis showed the order of the repeats and the heterogeneity in the lengths of individual arrays. After XbaI digestion and pulse field gel electrophoresis, the telomeric and two subtelomeric clones showed strong hybridization signals from 40 to 100 kb, with a maximum at 50 to 60 kb. We suggest that these fragments define a basic higher order structure and DNA loop domains of regions of rye chromosomes consisting of arrays of tandemly organized sequences.

INTRODUCTION

Large heterochromatic blocks, staining as C bands, are a major structural characteristic of the rye genome and are present near the telomere of most chromosome arms. These bands, which consist largely of highly repetitive DNA sequences (Appels et al., 1978), distinguish chromosomes of species of rye (*Secale*) from those of evolutionarily close genera, including wheat and barley. The size of the heterochromatic blocks varies between different rye accessions (Weimark, 1975; Lelley et al., 1978), indicating that the copy number of the DNA sequences may change extensively and, in evolutionary terms, rapidly. Four families of heterochromatic sequences have been isolated (Bedbrook et al., 1980; Appels et al., 1981; Jones and Flavell, 1982; McIntyre et al., 1990) and shown to represent most of the DNA present in the heterochromatin (Bedbrook et al., 1980). The authentic chromosomal end—the telomere— of most (but not all) eukaryotes consists of a short tandem repeat of the consensus sequence $(d[T/A]_{1-4}dG_{1-8})$ (Richards and Ausubel, 1988; reviewed in Richards et al., 1993; Biessmann and Mason, 1994). In situ hybridization using the telomeric oligomer $(TTTAGGG)_n$ as a probe to rye chromosomes showed that the repeats hybridize at the outermost ends of rye chromosomes, with few interstitial and no centromeric locations (Schwarzacher and Heslop-Harrison, 1991).

Pulse field gel electrophoresis (PFGE), which allows separation of large DNA fragments up to several megabases, has been used for genetic mapping of telomeres of tomato (Ganal et al., 1992), rice (Wu and Tanksley, 1993a, 1993b), and barley (Roder et al., 1993). In the genomes of all these species, it

was shown that tandemly repeated satellite DNA sequences have subtelomeric locations very close to the telomeric repeats. As a rule, these subtelomeric sequences are species specific and found only in the genomes of very closely related species. In these studies, it was shown that the subtelomeric repeats are located within 50 kb of the telomeric repeats but not immediately following them. Using PFGE, it is possible to separate and identify the individual arrays of tandemly organized DNA sequences, and sizes have been estimated for different species and sequence families from several kilobases to several megabases (for review of such sequences in mammals, see Manuelidis, 1990). PFGE has also been used for mapping individual centromeric α-satellite arrays and their size estimation in humans (Wevrick and Willard, 1989; Mahtani and Willard, 1990) and Arabidopsis (Murata et al., 1994).

The telomeric regions of chromosomes, including the telomeric repeat and proximal (subtelomeric) sequences, are important to study because they are responsible for the stability of chromosomes (Richards et al., 1993), must be localized to fix end points on genetic and physical maps, commonly associate with the nuclear envelope, and perhaps have critical roles in nuclear organization (Heslop-Harrison et al., 1993) and meiotic pairing (Schwarzacher and Heslop-Harrison, 1995). In rye, one chromosome pair behaves as though it has a terminal centromere, so the distal region may have neocentric activity (Prakken and Muntzing, 1942). The simple telomeric repeat alone is not sufficient to determine the function of the telomere; the whole telomeric structure includes the simple telomeric repeats, complex proximal sequences (Richards et al., 1993), and various interacting proteins (Biessmann and Mason, 1994), whereas the telomeric repeat

---

[1] To whom correspondence should be addressed.

itself is often located elsewhere in the genome (Richards et al., 1991; Regard et al., 1994). An important step toward understanding the organization of the various subtelomeric and telomeric DNA sequences can be made by the physical mapping approaches of fluorescent in situ hybridization and PFGE, which together bridge the gap between the level of a few kilobases in a clone and C bands along metaphase chromosomes that are undefined at the molecular level. Telomeric regions of rye chromosomes are of particular interest in studying structural and functional organization of telomeres because of the enrichment of different repetitive DNA families.

In this work, we investigated the large-scale organization of highly repetitive telomeric and subtelomeric DNA sequences of rye, using the short telomeric repeat from Arabidopsis (Richards and Ausubel, 1988) and clones representing three nonhomologous highly repetitive DNA families from rye. In particular, we wanted to examine the chromosomal location, genomic organization, and length of individual arrays of these sequences and to determine whether the arrangements of different sequence arrays are similar. As with other aspects of genetic analysis, comparisons with other species, including barley, tomato, and humans, provide important data about universal features of genomic organization and the evolutionary formation of chromosomal regions enriched for many repetitive DNA sequences.

RESULTS

Primary Structure and Genomic Organization

The nonhomologous sequences pSc200 and pSc250, containing inserts of 521 and 476 bp, respectively, are shown in Figures 1A and 1B. A homology search of the EMBL (release 84), GenBank, and DDBJ data bases did not reveal any sequences similar to pSc250. Within pSc200, the regions between positions −1 and −140 and between −380 and −521 have 97% homology, indicating that the clone represents part of an array of tandemly organized sequences with a monomer length of 379 bp. Data base comparisons showed that pSc200 is 94% homologous with the "350 family" (May and Appels, 1987). Within a third sequence, pSc119.2 (originally obtained by Bedbrook et al. [1980] and subcloned by McIntyre et al. [1990]; not homologous with pSc250 or sequences in the data bases), the four monomeric units (sequenced by McIntyre et al. [1990], although not included in major data bases) show between 93 and 96% homology. Figure 1 highlights various features within the sequences.

The nucleotide composition of pSc200 is 44.3% GC and that of pSc250 is 48.0% GC, whereas pSc119.2 has a higher GC content of 53.3%. For genomic DNA of wheat and rye, the GC content is estimated at 43% by buoyant density centrifugation and thermal denaturation (Bendich and McCarthy, 1970; they considered this value to be an underestimation because of the presence of 5-methylcytosine) or 47% by chemical
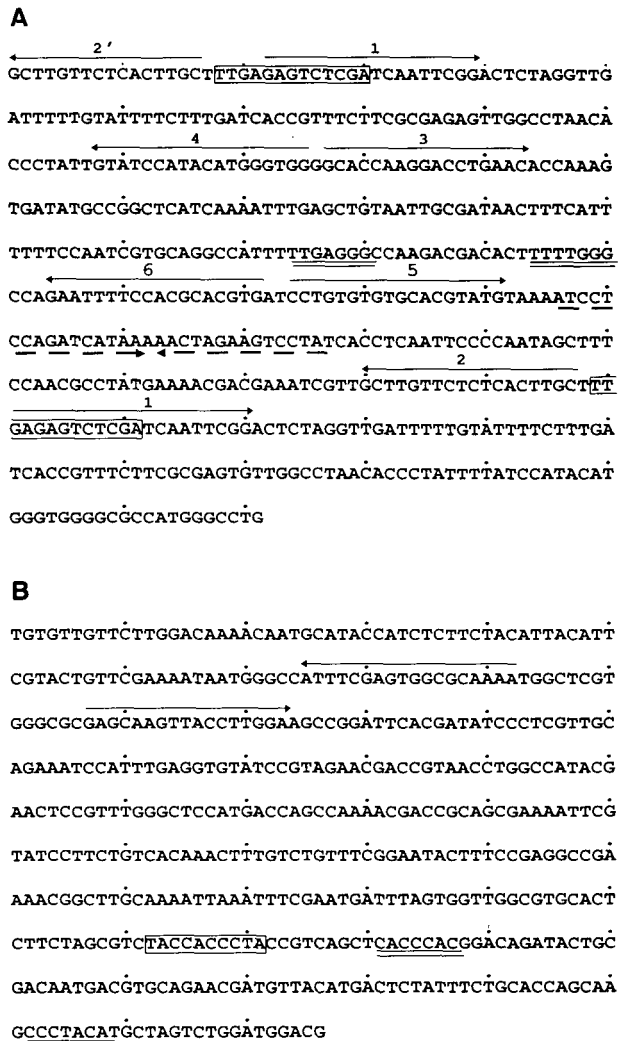
A

```
    2'                              1
←————————————————  ←——————————————————→
GCTTGTTCTCACTTGCT[TTGAGAGTCTCGA]TCAATTCGGACTCTAGGTTG

ATTTTTGTATTTTCTTTGATCACCGTTTCTTCGCGAGAGTTGGCCTAACA
                    4                    3
←————————————————  ◄——  ——————————————————→
CCCTATTGTATCCATACATGGGTGGGGCACCAAGGACCTGAACACCAAAG

TGATATGCCGGCTCATCAAAATTTGAGCTGTAATTGCGATAACTTTCATT

TTTTCCAATCGTGCAGGCCATTTTTGAGGGCCAAGACGACACTTTTTGGG
          6                    5
    ←——————————  ——————————————————
CCAGAATTTTCCACGCACGTGATCCTGTGTGTGCACGTATGTAAAATCCT

CCAGATCATAAAAACTAGAAGTCCTATCACCTCAATTCCCCAATAGCTTT
— — — — — — — — — — — — — — — — — — — — — 2
CCAACGCCTATGAAAACGACGAAATCGTTGCTTGTTCTCTCACTTGCTTT
——————————→
 1
GAGAGTCTCGA]TCAATTCGGACTCTAGGTTGATTTTTGTATTTTCTTTGA

TCACCGTTTCTTCGCGAGTGTTGGCCTAACACCCTATTTTATCCATACAT

GGGTGGGGCGCCATGGGCCTG
```

B

```
TGTGTTGTTCTTGGACAAAACAATGCATACCATCTCTTCTACATTACATT

                    ←———————————————————————
CGTACTGTTCGAAAATAATGGGCCATTTCGAGTGGCGCAAAATGGCTCGT
——————————————————→
GGGCGCGAGCAAGTTACCTTGGAAGCCGGATTCACGATATCCCTCGTTGC

AGAAATCCATTTGAGGTGTATCCGTAGAACGACCGTAACCTGGCCATACG

AACTCCGTTTGGGCTCCATGACCAGCCAAAACGACCGCAGCGAAAATTCG

TATCCTTCTGTCACAAACTTTGTCTGTTTCGGAATACTTTCCGAGGCCGA

AAACGGCTTGCAAAATTAAATTTCGAATGATTTAGTGGTTGGCGTGCACT

CTTCTAGCGTC[TACCACCCTA]CCGTCAGCTCACCCACGGACAGATACTGC
                                     —————————————
GACAATGACGTGCAGAACGATGTTACATGACTCTATTTCTGCACCAGCAA

GCCCTACATGCTAGTCTGGATGGACG
————————
```

Figure 1. Sequence of Subtelomeric Repeat Families.

(A) pSc200. A conserved region of dyad symmetry is boxed. Arrows show the positions of primers. Positions 2 and 2' of primer 2 differ by 2 bp because of the differences in monomers. The long inverted repeat is indicated with dashed underlining; variants of telomeric repeats are double underlined. Numbers indicate primers; dots indicate every 10 bp.

(B) pSc250. The region of high homology with the site for binding of the HMG1 chromatin proteins is boxed. Arrows show the positions of primers; variants of telomeric repeats are double underlined.

The EMBL accession number for pSc200 is Z50039, and that for pSc250 is Z50049.

methods (Schildkraut et al., 1962). Slot hybridization and calculation according to Rivin et al. (1986) show that the sequences studied represent ∼2.5% (pSc200) and ∼1% (pSc250) of the S. cereale genome. Bedbrook et al. (1980) previously showed that pSc119 represents 1 to 2% of the genome,

so the subclone pSc119.2 presumably represents a somewhat lower proportion.

The genomic organization of pSc119.2, pSc250, and pSc200 was studied by DNA gel blot hybridization to different restriction enzyme digests of *S. cereale* (Figure 2). We also examined hybridization to wild rye species to determine the species specificity of the sequences and compare their organization in close relatives of *S. cereale*. pSc119.2 yielded a ladder pattern of hybridization fragments, typical for simple tandemly organized DNA sequences (Figure 2A). The length of a monomeric unit, 120 bp, corresponds to the analysis of primary structure (McIntyre et al., 1990). After HaeIII treatment, the most intense fragment was 120 bp, and the 360-bp fragment was strongest after TaqI treatment.

The pattern of hybridization fragments of pSc250 after EcoRV digestion corresponds to a ladder pattern, with a monomeric unit of ~550 bp (Figure 2B). Because the analysis of the pSc250 sequence showed one EcoRV site (Figure 1), we concluded that pSc250 is a representative of a family of tandemly

organized DNA sequences with an unusually extended monomer length. After RsaI or HaeIII digestions, additional fragments were obtained.

Complex patterns were revealed after hybridization of pSc200 to rye genomic DNA digests (Figure 2C). Sites for common restriction enzymes such as BamHI, BglII, EcoRI, EcoRV, HindIII, and others were absent in the sequence and its forward complement. HaeIII digests showed fragments of 125 and 240 bp (plus very small fragments), consistent with the sequence data, and 379 bp, consistent with the single site in the forward complementary sequence. As predicted from the sequence, most DNA of the pSc200 family was not cut by RsaI or EcoRV, and strong hybridization was seen in the upper part of the autoradiograph after RsaI and EcoRV digestions (Figure 2C). A part of the sequences includes RsaI or EcoRV sites, and their basic periodicity corresponds to 760 bp (a dimer of 380 bp) with additional fragments. Thus, we found higher order structures, including a double monomer unit of 379 bp, in the pSc200 family. The broad band near the 210-bp marker
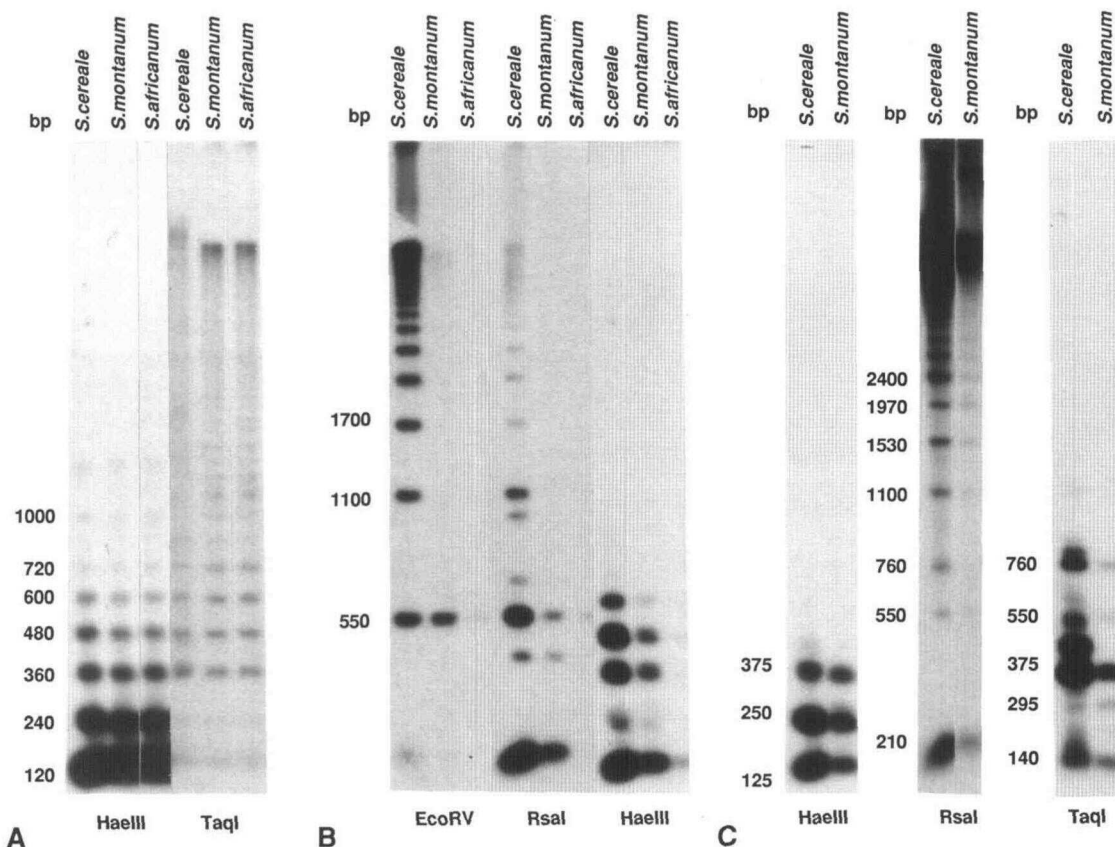

Figure 2. DNA Gel Blot Hybridization of Subtelomeric Sequences to Genomic DNA of *S. cereale* and Wild Rye Species.

(A) pSc119.2.

(B) pSc250.

(C) pSc200.

Numbers at left indicate lengths in base pairs. Restriction enzymes used for digests are shown below the relevant lanes.
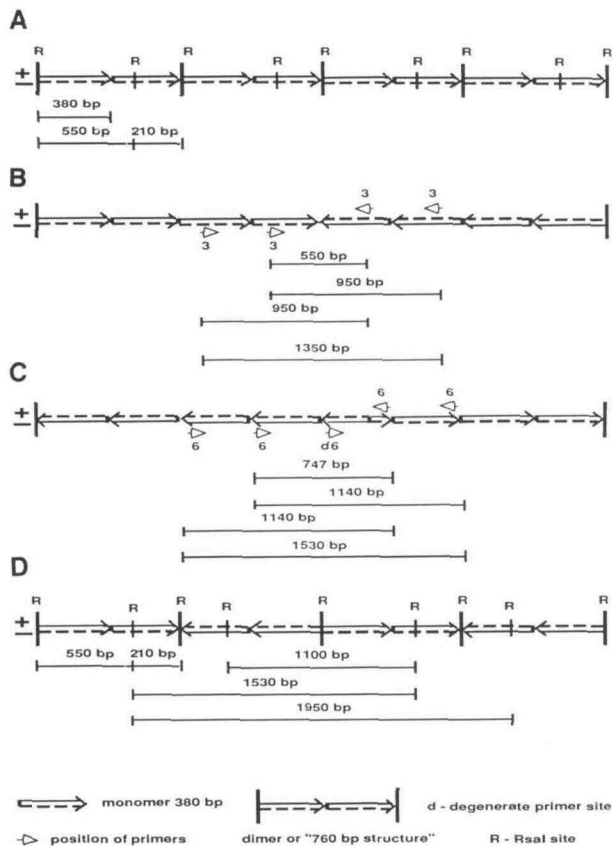
Figure 3. The Organization of Repeat Monomers of pSc200.

(A) A model showing head-to-tail organization of the monomer and RsaI sites. The model explains neither restriction fragments (Figure 2) nor the PCR products (Figure 4).

(B) and **(C)** A proposed model with head-to-head and tail-to-tail junctions within the head-to-tail array. Positions of PCR primers are indicated; the lengths of fragments correspond to actual PCR products (Figure 4). The 747-bp fragment corresponds to a sequenced product.

(D) A model accounting for both DNA gel blot hybridization results (Figure 2) and the PCR results (Figure 4 and **[B]** and **[C]**).

(+) and (−) indicate the two DNA strands; d, degenerate primer sites; R, RsaI sites.

obtained. Primer 3 (reverse complementary orientation with respect to primer 6) gave major fragments of ~550, 950, and 1350 bp (Figure 4, lane 4). Thus, some tandem arrays with an inverted "760-bp structure" organized head to head and tail to tail are present. Primer 3 is located near the center of the monomer unit, and primer 6 is found near the junction of the inverted sequence (Figures 3B and 3C). When primers were used in pairs (1 plus 2, 3 plus 4, and 5 plus 6), each yielded a ladder with a monomeric unit of 379 bp (e.g., Figure 4, lane 3), as expected from the sequence and hybridization data (Figures 1 and 2). In combination with the single primer data, we concluded that the 760-bp structure has two monomers of 379 bp situated head to tail and that there are higher order units where the 760-bp structure occurs in tail-to-tail and head-to-head orientations. PCR products from the 750- and 550-bp products (Figure 4, lanes 4 and 6) were cloned and sequenced. One sequence is 747 bp long (EMBL accession number Z54189). The sequence includes two complete monomers (inverted with respect to Figure 1A) and two inverted RsaI sites 379 bp apart, showing the occurrence of the complete monomer in both forward and inverted orientations in the genome. The 747 bp sequence and positions of two copies of primer 3 and two copies of primer 6 (plus a degenerate) support the DNA gel blot hybridization (Figure 2C) and the PCR data (Figure 4) and are incorporated into the scheme of the 760-bp structure genomic organization in Figures 3B and 3C.

Primer 5 is situated between primer 6 (reverse complementary orientation) and the large region of the inverted repeat. Ending at nucleotide 498 (and twice in the 747-bp PCR product
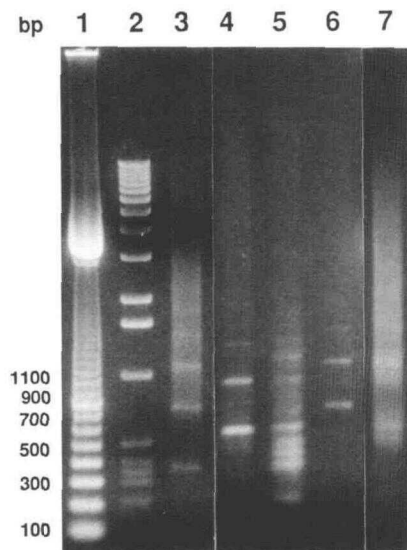
most probably represents two fragments derived from cutting the monomer into two nearly equal fragments at RsaI sites present in a few family members. Some fragments, such as those of 430 and 1100 bp, cannot be explained by the typical and simple tandem repeat organization with a head-to-tail arrangement of monomer units (Figure 3A).

We used polymerase chain reaction (PCR) analysis to aid in interpretation of the genomic organization of pSc200 obtained from DNA gel blot hybridization. PCR results using single primers and primer pairs constructed from the sequence (see Figure 1A) with rye genomic DNA as a template are shown in Figure 4. When primer 6 was used alone (Figure 4, lane 6), three major products of ~760, 1150, and 1530 bp were

Figure 4. The DNA Fragments Obtained by PCR of *S. cereale* with Single and Pairs of Primers.

Lane 1 contains markers of a 100-bp ladder; lane 2, markers of a 1-kb ladder; lane 3, primers 1 and 2 for pSc200; lane 4, primer 3; lane 5, primer 5; lane 6, primer 6; and lane 7, a pair of primers for pSc250.

discussed above), seven of eight are reverse complementary to the 3' end of the primer, presumably yielding the 210-bp product (at least from variants not sequenced). Other products are represented by the 380-bp ladder.

When we used single primers from the pSc250 sequence, no PCR products were obtained from genomic DNA. When primer pairs were used (Figure 4, lane 7), the basic ladder pattern was obtained (see Figure 2), with a considerable smear. A fragment of 550 bp was revealed by both DNA gel blot hybridization and PCR analysis, although only the latter generated a shorter 500-bp fragment. This fragment and the strong smear indicate the presence of the pSc250 family monomers with different unit lengths and sequence variants, but the lack of single primer products indicates that the monomers virtually never occur in a head-to-head orientation.

The genomic organizations of pSc119.2, pSc200, and pSc250 were compared in S. cereale and in the wild rye species S. montanum, S. silvestre, and S. africanum (see Figure 2). pSc119.2 showed very similar hybridization patterns in all four species. No hybridization of pSc200 and pSc250 to DNA from S. silvestre, or of pSc200 to S. africanum, was detected (data not shown). Hybridization showed considerably weaker signal from pSc200 and pSc250 to the DNA of the wild species. Although the monomeric units were similar in all species, various other fragments were missing or of considerably reduced intensity in the wild species.

Chromosomal Location of pSc119.2, pSc200, and pSc250

After performing in situ hybridization using clones pSc119.2, pSc200, and pSc250 (Figure 5), strong hybridization signals were detected at many subtelomeric locations corresponding to positively fluorescing heterochromatic C bands by staining with 4',6-diamidino-2-phenylindole. Figures 5A and 5B show a prophase with pSc119.2 hybridization at many minor intercalary sites and multiple sites near, but never at, the ends of all short and many long chromosome arms. Using microscopy of high-resolution extended prophase chromosomes, a total of ∼100 discrete hybridization sites were visualized as bands or double dots. In contrast, sequences homologous with pSc200 and pSc250 were located in mainly single, prominent bands very close to the ends of ∼18 chromosome arms and did not hybridize significantly at intercalary positions (Figures 5D and 5F). Double target hybridization (Figures 5G and 5H) showed that both sequences occur together on most chromosomes and that pSc250 (yellow) is always more proximal than pSc200 (red), although one telomere has only a prominent pSc200 site.

Large-Scale Organization

Gel blot hybridization of highly repetitive DNA sequences to DNA separated by PFGE usually shows a strong hybridization

smear to the lower part of the filter (in a region from 50 to 200 kb), preventing identification of separate fragments. We used additional high-stringency washes of the filters so that the hybridization would detect only fragments with high homology with the subtelomeric sequences and not fragments homologous with diverged members of the repeat families.

From preliminary experiments using a wide range of restriction enzymes and examination of the primary sequences, BglII, EcoRV, and XbaI were chosen for comparison of hybridization patterns of subtelomeric and telomeric DNA sequences. These restriction enzymes cut frequently in genomic DNA but, with the exception of EcoRV for pSc250, had no sites within pSc200, pSc250, pSc119.2, or pLT11, the telomeric sequence from Arabidopsis. The patterns of hybridization of these four sequences consisted of up to 20 discrete fragments from 30 to 800 kb (Figure 6).

The fewest fragments were revealed after hybridization with pLT11. After BglII digestion, eight fragments between 30 and 290 kb were generated. The fragment of 120 kb showed strong hybridization. EcoRV digests yielded a similar overall range, but with different fragments, including one strongly hybridized fragment at 200 kb. All hybridization signal after XbaI digestion was concentrated in two broad bands, 50 to 80 kb and 120 to 150 kb. The strong bands in these limited zones indicated high homogeneity of length between different chromosomal copies of the repeat array and little degeneracy of sequence giving XbaI sites.

All subtelomeric sequences hybridized to substantially more restriction fragments than did the telomeric repeats, and the patterns of each enzyme–probe combination were unique. pSc119.2 was the only sequence showing the hybridization fragments of the higher molecular weight in all three digests. EcoRV yielded many fragments with all three probes (pSc200 yielded 20, and pSc250 and pSc119.2 yielded 13 each), with a nearly uniform intensity. Hybridization patterns with BglII showed considerable variations in intensities of fragments.

Comparing hybridization patterns of these sequences after XbaI digestion was of particular interest. pSc119.2 yielded a set of 14 fragments (the highest number of the three restriction enzymes) from 45 to 670 kb; the highest intensity fragment was 230 kb long. The patterns of pSc200 and pSc250 were similar to each other and included strong smears of signals ranging from 40 to 100 kb, with most at 50 to 60 kb. Longer exposures revealed a few longer fragments with significantly lower intensity ranging from 170 to 425 kb. Such a concentration of a signal in a limited zone indicated a relatively constant distance between XbaI sites in all tandem arrays of the sequences, whereas the wide variation in fragment lengths, characteristic of pSc119.2, indicated either heterogeneity of individual array lengths or the interspersion of other sequence families in the arrays.

Hybridization patterns of subtelomeric DNA sequences of tomato (Ganal et al., 1991) and barley (Roder et al., 1993), as well as subtelomeric rye sequences, have a wide set of fragment lengths. The longest fragments are 850 kb and longer. We obtained similar fragments when we did not use high-stringency washes or used longer film exposures.
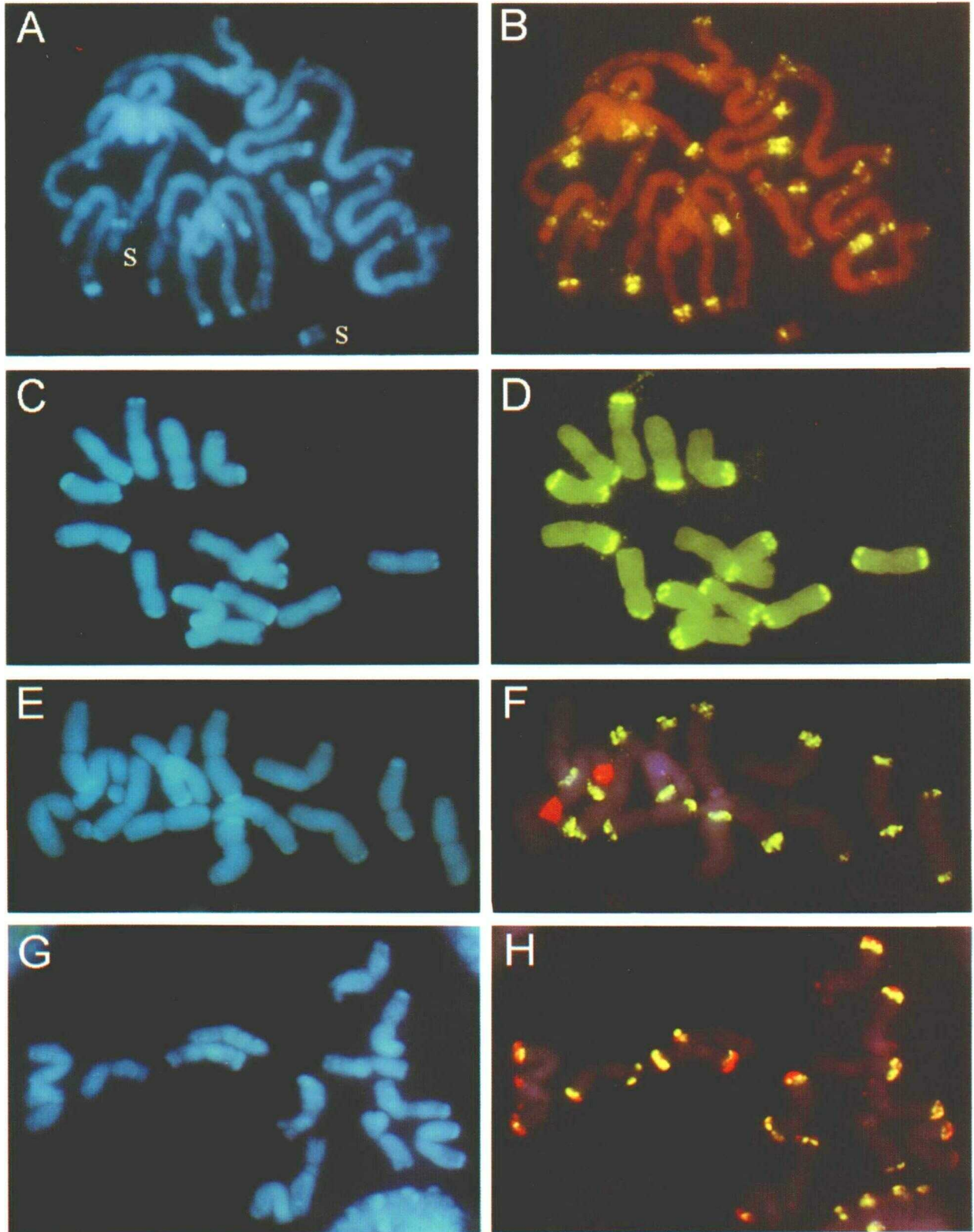
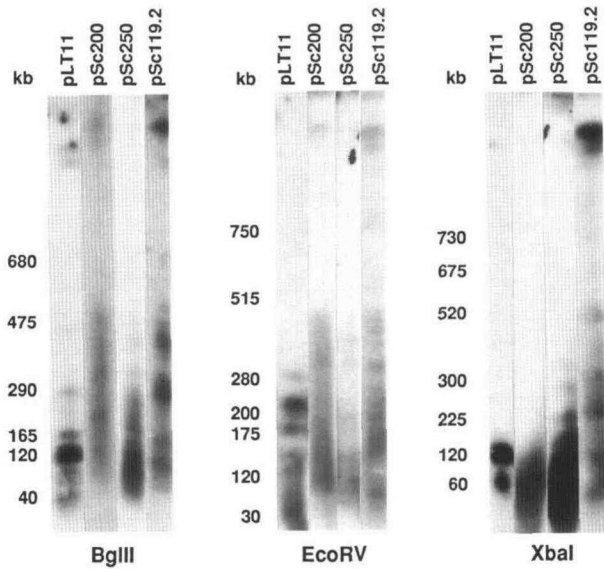**Figure 5.** Fluorescent in Situ Hybridization to Root Tip Chromosomes of Rye.

Figure 6. Hybridization Pattern after PFGE.

The same filter was hybridized sequentially to each of the four telomeric and subtelomeric probes. Numbers at left indicate lengths in kilobases. Restriction enzymes used for digests are shown below the relevant lanes.

In summary, we can conclude that the hybridization pattern and fragment lengths for each of the different rye subtelomeric sequences are basically different. In particular, no conspicuous fragments contain any pair of sequences. pSc200 and pSc250 have more similar large-scale organization following from XbaI digests. Arrays of pSc119.2 are generally on larger fragments, indicating that they are the most proximal fragment, a result in agreement with in situ hybridization data.

DISCUSSION

The structural organization of the distal parts of chromosomes relates to many aspects of their stability, activity, and evolution. The combination of conventional and pulse field electrophoresis with PCR and in situ hybridization analysis enabled the complex and interrelated genomic organization of three major nonhomologous subtelomeric DNA sequences to be compared with the telomeric sequence in the rye genome.

The pSc119.2 family has the shortest monomer length of the sequences investigated, 120 bp (see Figure 2A). A higher order structure is revealed with TaqI: fragments of 120 and 240 bp are virtually absent, but fragments in a 120-bp ladder are present from 360 bp. The lack of monomers and dimers and strong trimer, tetramer, and pentamer fragments indicates that the repeat is present in larger structural units, because random variation in the restriction site would lead to the presence of monomers and dimers. Such multimeric structures are also characteristic of centromeric α-satellite DNA in humans (reviewed in Willard and Waye, 1987) and mouse (Dod et al., 1989), with variants often highly specific to single chromosome types.

Tandem repeats with a monomer length equal to the length that wraps around a nucleosome core and linker (170 to 180 bp or double) have been described in many cereals (Peacock et al., 1981; De Kochko et al., 1991; Vershinin et al., 1994). The length of the pSc250 monomer, 550 bp, is longer than most; the pSc119.2 monomer, at 120 bp, is shorter. Using the model of tandem repetitive DNA formation proposed by Stephan and Cho (1994), it follows that repeat length is increased with decreasing rates of (unequal) crossing over. The immediately subtelomeric regions of rye chromosomes, where pSc250 is located, has a frequency of crossing over that is lower than average, corresponding to cytological data (Jones, 1978). The general conclusion from theoretical models by Smith (1976) and Stephan and Cho (1994) is that natural selection can push random sequences toward order by controlling sequence

Figure 5. (continued).

In (A), (C), (E), and (G), the chromosomes were counterstained with the DNA-specific dye 4',6-diamidino-2-phenylindole to show their morphology. Strongly fluorescing, large heterochromatic C bands are visible at many chromosome ends.

(A) and (B) A pro-metaphase chromosome preparation probed with pSc119.2 labeled with digoxigenin and detected by fluorescein isothiocyanate (FITC) fluorescence. The extended chromosomes allow high-resolution localization at many intercalary and subtelomeric sites. Satellite regions (s) of chromosome 1R are separated at the nucleolus organizing region.

(C) and (D) pSc200, labeled with digoxigenin and detected with FITC, is located in 18 large blocks at the ends of most chromosomes.

(E) and (F) pSc250, labeled with digoxigenin and detected with FITC, together with direct rhodamine-labeled rDNA probe. After use of a UV-reduced excitation beam and a triple-band pass filter, strong signal of both probes is visible and DAPI fluorescence (blue) is reduced. pSc250 is located in 16 large and two small subtelomeric sites; many of the large blocks seem to be composed of several smaller bands. Two large red rDNA sites on chromosome 1R are visible.

(G) and (H) Simultaneous in situ hybridization of pSc200 (labeled with biotin and detected by red Cy3 fluorescence) and pSc250 (labeled with digoxigenin and detected with green FITC fluorescence) shows that most heterochromatic blocks, visible as strongly fluorescing bands after 4',6-diamidino-2-phenylindole staining, contain both repeated families, whereas only a few minor sites show hybridization with one probe. Where both probes are located together, there is considerable overlap of two domains evidenced by yellow fluorescence (green plus red), but it is clear that pSc200 is more distal in all cases. Magnification, ×2000.

length, although the forces controlling the sizes of tandem repeats are not yet known.

Many short direct, inverted repeats and stems (short palindromes) are present in both pSc250 and pSc200. Such structures are characteristic features of repetitive DNA sequences and can provide a potential source for homologous recombination allowing creation of rearrangements. We note that pSc250 has a region of high homology with the motif TCAATTAAAT (see Figure 1B), which is the site for binding of the high-mobility group I chromatin proteins (Reeves and Nissen, 1990). The occurrence of the protein binding sites supports the possibility that pSc200 (see also Appels et al., 1986) and pSc250 have a structural role that may relate to telomere functions.

The majority of tandem repeat monomers are arranged in a head-to-tail orientation. Analysis of the primary structure and DNA gel blot hybridization results shows that the rye subtelomeric sequences have such a type of tandem organization. PCR amplification of genomic DNA with primer pairs from pSc250 (see Figure 4, lane 7) showed the existence of pSc250 tandem arrays with units shorter than the 550-bp monomers of basic tandem arrays revealed by DNA gel blot hybridization. These minor subsets could be specific for a definite chromosome (or chromosomes) and coexist with the basic monomer within a single array. Both possibilities have been demonstrated for the organization of different subsets of human α-satellite DNA (Willard et al., 1986). Strachan et al. (1985) proposed that the polymorphic higher order repeats represent transition stages in the ongoing homogenization of tandem arrays.

The contrasting head-to-head type of tandem organization resulting from evolutionary mechanisms involving breakage-fusion-bridge cycles or stem–loop formation and crossing over has been postulated, but there is little experimental evidence for this organization in eukaryotic genomes. Although the major fragments hybridizing with pSc200 in the DNA gel blots (see Figure 2) can arise from head-to-tail organization, some fragments cannot arise from this type of organization in a 379-bp monomer, even if assumptions are made about sequence divergence, concerted evolution, and different arrangements of sequence blocks. PCR amplification of genomic DNA showed that single primers generate strong products (see Figure 4); all of the major PCR fragments can be obtained if the basic monomer occurs in a head-to-head and tail-to-tail organization with changes in direction at defined points (close to the position of primer 6; Figures 3B and 3C). In long tandem arrays in humans and Chinese hamster, the direction of the monomers changes at sites of large inversions (Toledo et al., 1992; Wevrick et al., 1992), and in contrast to the pSc200 organization, the mammalian inversions appear to be isolated events. In rye, the presence of a head-to-head organization of some monomers suggests that the formation of chromosomes during evolution might have been accompanied by the breakage-fusion-bridge cycles of the chromatid type first described by McClintock (1941), which explains the origin of inverted repeats as well as over-replication. Furthermore, as indicated by the in situ hybridization results (see Figure 5), such inverted repeat organization implies that the arrays with head-to-head monomers probably lie closer to the telomeres than do arrays with head-to-tail organization. Lukaszewski (1995) has demonstrated active breakage-fusion-bridge cycles of the chromatid type initiated by reverse tandem duplications in wheat chromosomes after crossing with rye. Chromosome-type cycles, giving rise to additional reverse tandem duplications, were also detected, along with dicentric and ring chromosomes.

When the pulse field blots were probed with the Arabidopsis telomere clone pLT11, restriction fragments were revealed between 30 and 300 kb (see Figure 6). These fragment lengths greatly exceeded the length of real tandem arrays of telomeric repeats. Calculations from Bal31 digests (Schwarzacher and Heslop-Harrison, 1991) indicate that the length of the telomeric repeat on rye chromosomes varies from 1 to 6 kb, similar to estimates involving Arabidopsis (Richards and Ausubel, 1988), humans (Cross et al., 1989), and tomato (Ganal et al., 1991). The discrete number of restriction fragments and the presence of some stronger bands imply the existence of higher order DNA structure near the ends of the chromosomes. XbaI discriminates only two major classes of DNA in the telomeric region, implying the presence of XbaI sites at a conserved distance from each telomere. BglII and EcoRV show more polymorphism (up to eight bands of the 14 maximum), which might make the fragments suitable for mapping some telomeres, as in tomato (Ganal et al., 1991).

The arrays of the tandemly organized subtelomeric sequences are much longer than those of the telomeric repeats. With the pLT11 probe and XbaI digest, the fragments concentrate in two broad bands, 50 to 80 kb and 120 to 150 kb. The length of the first band coincides with strong signals after hybridization with pSc200 and pSc250, whereas pSc119.2 yields 14 fragments from 45 to 670 kb. The considerable discrepancy in the patterns of hybridization implies that pSc119.2 differs from the other three sequences in large-scale organization and, as revealed by in situ hybridization (see Figures 5A and 5B), shows the most proximal location and many intercalary sites. The number of the pSc119.2 fragments (>20 in the EcoRV digest) correlates with the number of bands detected by high-resolution in situ hybridization (~50 on a haploid set; using a lower stringency than that of the PFGE), suggesting that many pulse field fragments correspond to one of four individual chromosomal sites.

Our PFGE experiments (Figure 6) indicate copy numbers from 100 to 2000 (pSc200) and 60 to 900 (pSc250), assuming that the entire fragment is a tandem array of the repeat sequence. The sites localized by in situ hybridization (Figures 5C to 5H) are much larger, and each must include several enzyme sites. In situ hybridization shows that the overall size of arrays varies over a two- to fourfold range from chromosome to chromosome; on average, each of the nine physical sites of pSc200 represents ~20 Mbp or ~50,000 monomers; pSc250 represents 10 Mbp or 20,000 monomers. Comparisons with the size of Arabidopsis chromosomes (on average, 20 Mbp) and the increased signal strength compared with α-satellite hybridization to human chromosomes (5000 to 20,000 monomers; Mahtani and Willard, 1990) support these estimates.

Following PFGE, the peak of the signals after XbaI treatment and hybridization with pSc200, pSc250, and partially with pLT11 is concentrated around 50 to 60 kb, the size of stable DNA fragments obtained in chromatin from different plant tissue by Espinas and Carballo (1993) and in mammals (for review, see Manuelidis, 1990). Several higher order levels of DNA organization have been identified in chromatin, including the nucleosome, the 30-nm solenoid, and the DNA loop domain (Nelson et al., 1986). The loop domain defines a basic unit of higher order DNA structure in eukaryotic cells, including replicons (for review, see Nelson et al., 1986). XbaI sites might thus define an element of the higher order DNA structures or loop domains of the heterochromatic regions of rye chromosomes. It is noteworthy that the telomeric sequence, pLT11, also shows the 60-kb fragment length. Knowledge about the higher order structural organization of eukaryotic chromosome ends is restricted. However, as with the telomeric repeat, some aspects of the complex organization of the subtelomeric DNA sequences at the ends of rye chromosomes most probably will be found in other species. A comparative approach will enable identification of general features defining the structure and function of the distal regions of large chromosomes.

METHODS

DNA Probes and Labeling

We used four cloned DNA sequences. The telomeric probe pLT11 represents the 400-bp insert of pAtT4 from Arabidopsis (Richards and Ausubel, 1988) subcloned in pUC19. pLT11 was provided by L. Turner and N. Ellis (John Innes Centre) with the kind permission of E. Richards (Washington University, St. Louis, MO) and F. Ausubel (Massachusetts General Hospital, Boston). pSc119.2 from *Secale cereale* was obtained from R. Appels (McIntyre et al., 1990) and is a subclone of pSc119 (Bedbrook et al., 1980). Two new clones were made by sonicating total genomic DNA from *S. cereale* (Russian variety Onokhoiskaya) to an average length of 200 to 600 bp. The fragments were denatured at 100°C in 0.1 M sodium phosphate buffer, pH 7.5, for 10 min and incubated at 60°C to C_0t of 0.02 (Britten and Kohne, 1968). The samples were treated with S1 nuclease. The resulting highly repeated double-stranded DNA sequences were cloned into the pUC18 at the PstI site. From several thousand clones, 300 were chosen randomly for dot hybridization with ^{32}P-labeled total DNA from rye, wheat, and barley. Two clones, pSc200 and pSc250, gave very strong signals with rye DNA and no signals with wheat and barley DNA; they were selected for further studies.

For in situ hybridization, the rye clones were labeled with digoxigenin-11-dUTP (Boehringer Mannheim) or biotin-11-dUTP (Sigma) by polymerase chain reaction (PCR). Inserts of all probes amplified by PCR were labeled with ^{32}P-dCTP (Amersham Corp.) by random priming (Feinberg and Vogelstein, 1983).

DNA Isolation, Blot and Slot Hybridization, and Sequencing

Total genomic DNA for gel blot hybridization was extracted from seedlings of *S. cereale* (var Petkus), *S. montanum* (John Innes Centre accession No. PBI R15), *S. africanum* (PBI R102), and *S. silvestre* (PBI

R52), according to Ausubel et al. (1987). DNA was digested with restriction enzymes, separated on 1% agarose gels, and transferred to a Hybond-N+ (Amersham Corp.) membrane. Hybridization was performed in 5 × SSPE (1 × SSPE is 0.15 M NaCl, 0.01 M NaH$_2$PO$_4$, 0.02 M EDTA, pH 7.4), 5% SDS, 2% BSA at 65°C, followed by washes in 2 × SSPE, 0.1% SDS twice and 1 × SSPE, 0.1% SDS at 65°C. To allow reuse of the filters, probes were removed by incubating the filters in 0.1 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.1% SDS for 15 min at 100°C or in 0.4 M NaOH for 20 min at 45°C.

The copy numbers of pSc200 and pSc250 were determined by slot hybridization under the same conditions as those for DNA gel blot hybridization. The radioactivity in each spot was determined by liquid scintillation counting, and the copy number was calculated as described by Rivin et al. (1986). pSc200 and pSc250 were sequenced by automated DNA sequencer (ALF; Pharmacia). Computer analysis of sequence data was performed using the Genetics Computer Group (Madison, WI) package (Devereux et al., 1984).

PCR Analysis

Six primers contained in the insert of pSc200 were synthesized and are as follows: primer 1, 5'-GAGTCTCGATCAATTTCGG-3'; primer 2, 5'-GCAAGTGAGAGAACAAGC-3'; primer 3, 5'-GCACCAAGGACCTG-AAC-3'; primer 4, 5'-CCACCCATGTATGGATAC-3'; primer 5, 5'-CTG-TGTGTGCACGTATG-3'; and primer 6, 5'-CACGTGCGTGGAAAATTC-3'. Two primers from pSc250 were made, as indicated in Figure 1B. PCR was performed with total genomic DNA in a thermal cycler (Perkin-Elmer Cetus, Norwalk, CT) using the following conditions: 5 min at 94°C for 30 cycles of 30 sec at 94°C, 30 sec at 56°C, and 90 sec at 72°C, followed by 5 min at 72°C. PCR products were separated on a 1% agarose gel, and 100-bp and 1-kb ladders were used as length standards. PCR products were cloned in pUC18 using a T4 DNA polymerase as described by Wang et al. (1994) and sequenced using a Pharmacia ALF sequencing apparatus.

In Situ Hybridization

Methods for chromosome slide preparation and in situ hybridization were adapted from protocols developed by Schwarzacher et al. (1989) and Heslop-Harrison et al. (1991). Seedling root tips from *S. cereale* (var Petkus) were treated with ice water, fixed in ethanol/glacial acetic acid (3:1), digested with enzymes, and squashed in 45% acetic acid. Slides were pretreated with RNase, pepsin, and fresh paraformaldehyde, washed in 2 × SSC, dehydrated in an ethanol series, and air dried. The hybridization mixture consisted of 0.5 to 1.5 ng/μL DNA probe, 50% formamide, 10% dextran sulfate, 0.1% SDS, and 300 ng/μL sheared salmon sperm DNA in 2 × SSC. The hybridization mixture was added to the preparations, and they were denatured together at 70 to 75°C in a modified thermocycler for 5 to 8 min. Hybridization was performed overnight at 37°C. Slides were washed in 20% formamide in 0.1 × SSC at 42°C, followed by several washes. Sites of probe hybridization were detected via fluorescein isothiocyanate–conjugated sheep anti-digoxigenin antibody (Boehringer Mannheim) and 5 μg/mL Cy3 conjugated to streptavidin (Sigma) in 4 × SSC, 0.1% Tween 20, 5% BSA. Slides were counterstained with 4',6-diamidino-2-phenyl-indole and occasionally probidium iodide and mounted in antifade solution (AF1; Citifluor, London, UK). Preparations were analyzed on a Leitz (Leica, Wetzlar, Germany) epifluorescence microscope single-band pass filters and an Omega (Brattleboro, VT) triple-band pass filter set. Photographs were taken on SuperHG400 color film (Fuji, Tokyo, Japan) and digitized with the PhotoCD (Eastman Kodak) system. Final

prints were prepared with Adobe (Mountain View, CA) Photoshop (applying the same adjustments of contrast, brightness, and color balance to the whole image).

DNA Plug Preparation and Pulse Field Gel Electrophoresis

The isolation of intact protoplasts and the subsequent mixing with 2% low-melting-point agarose, washing the blocks, and treating high molecular weight DNA with restriction enzymes were performed essentially as described by Cheung and Gale (1990). Gels were run with a CHEF-DRII pulse field electrophoresis system (Bio-Rad) in 1% agarose (Fast-Lane; FMC BioProducts, Rockland, ME) at 4.5 V/cm and 75-sec pulse time, optimizing separation of fragments between 50 and 800 kb. The transfer of DNA to a filter, hybridization, and washing were performed as for conventional DNA gel blot hybridization.

REFERENCES

Appels, R., Driscoll, C., and Peacock, W.J. (1978). Heterochromatin and highly repeated DNA sequences in rye (*Secale cereale*). Chromosoma **70**, 67–89.

Appels, R., Dennis, E.S., Smith, D.R., and Peacock, W.J. (1981). Two repeated DNA sequences from the heterochromatic regions of rye *Secale cereale* chromosomes. Chromosoma **84**, 265–277.

Appels, R., Moran, L.B., and Gustafson, J.P. (1986). Rye heterochromatin. I. Studies on clusters of the major repeating sequence and the identification of a new dispersed repetitive sequence element. Can. J. Genet. Cytol. **28**, 645–657.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds (1987). Current Protocols in Molecular Biology. (New York: Greene Publishing Associates/Wiley Interscience).

Bedbrook, J.R., Jones, J., O'Dell, M., Tompson, R., and Flavell, R. (1980). A molecular description of telomeric heterochromatin in *Secale* species. Cell **19**, 545–560.

Bendich, A.J., and McCarthy, B.J. (1970). DNA comparisons among barley, oats, rye and wheat. Genetics **65**, 545–566.

Biessmann, H., and Mason, J.M. (1994). Telomeric repeat sequences. Chromosoma **103**, 154–161.

Britten, R.J., and Kohne, D.E. (1968). Repeated sequences in DNA. Science **161**, 529–541.

Cheung, W.Y., and Gale, M.D. (1990). The isolation of high molecular weight DNA from wheat, barley and rye for analysis by pulse-field gel electrophoresis. Plant Mol. Biol. **14**, 881–888.

Cross, S.H., Allshire, R.S., McKay, S.J., McGill, N.I., and Cooke, H.J. (1989). Cloning of human telomeres by complementation in yeast. Nature **338**, 771–774.

De Kochko, A., Kiefer, M.C., Cordesse, F., Reddy, A.S., and Delseny, M. (1991). Distribution and organization of a tandemly repeated 352-bp sequence in the *Oryzae* family. Theor. Appl. Genet. **89**, 217–225.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**, 387–395.

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.

Espinas, M.L., and Carballo, M. (1993). Pulse-field gel electrophoresis analysis of higher-order chromatin structures of *Zea mays*. Highly methylated DNA in the 50 kb chromatin structure. Plant Mol. Biol. **21**, 847–857.

Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**, 6–13.

Ganal, M.W., Lapitan, N.L.V., and Tanksley, S.D. (1991). Macrostructure of the tomato telomeres. Plant Cell **3**, 87–94.

Ganal, M.W., Broun, P., and Tanksley, S.D. (1992). Genetic mapping of tandemly repeated telomeric DNA sequences in tomato. Genomics **14**, 444–448.

Heslop-Harrison, J.S., Schwarzacher, T., Anamthawat-Jonsson, K., Leitch, A.R., Shi, M., and Leitch, I.J. (1991). In situ hybridization with automated chromosome denaturation. Technique **3**, 109–116.

Heslop-Harrison, J.S., Leitch, A.R., and Schwarzacher, T. (1993). The physical organization of interphase nuclei. In The Chromosome, J.S. Heslop-Harrison and R. Flavell, eds (Oxford: BIOS Scientific Publisher), pp. 221–232.

Jones, G.H. (1978). Giemsa C-banding of rye meiotic chromosomes and the nature of "terminal" chiasmata. Chromosoma **66**, 45–57.

Jones, J.D.G., and Flavell, R. (1982). The mapping of highly-repeated DNA families and their relationship to C-bands in chromosomes of *Secale cereale*. Chromosoma **86**, 595–612.

Lelley, T., Josifek, K., and Kaltsikes, P.J. (1978). Polymorphism in the giemsa C-banding pattern of rye chromosomes. Can. J. Genet. Cytol. **20**, 307–312.

Lukaszewski, A.J. (1995). Chromatid and chromosome type breakage-fusion-bridge cycles in wheat (*Triticum aestivum* L.). Genetics **140**, 1069–1085.

Mahtani, M.M., and Willard, H.F. (1990). Pulsed-field analysis of α-satellite DNA at the human X chromosome centromere: High-frequency polymorphisms and arrays size estimate. Genomics **7**, 607–613.

Manuelidis, L. (1990). A view of interphase chromosomes. Science **250**, 1533–1540.

May, C.E., and Appels, R. (1987). The molecular genetics of wheat: Toward an understanding of 16 billion base pairs of DNA. In Wheat and Wheat Improvement, E.G. Heyne, ed (Madison, WI: American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America), pp. 165–198.

McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. Genetics **26**, 234–282.

McIntyre, C.L., Pereira, S., Moran, L.B., and Appels, R. (1990). New *Secale cereale* (rye) DNA derivatives for the detection of rye chromosome segments in wheat. Genome **33**, 317–323.

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

Nelson, W.G., Pienta, K.J., Barrack, E.R., and Coffey, D.S. (1986). The role of the nuclear matrix in the organization and function of DNA. Annu. Rev. Biophys. Chem. **15**, 457–475.

Peacock, W.J., Dennis, E.S., Rhoads, M.M., and Pryor, A.J. (1981). Highly repeated DNA sequences limited to knob heterochromatin in maize. Proc. Natl. Acad. Sci. USA **78**, 4490–4494.

Prakken, R., and Muntzing, A. (1942). A meiotic peculiarity in rye, simulating a terminal centromere. Hereditas **28**, 441–482.

Reeves, R., and Nissen, M.S. (1990). The AT-DNA-binding domain of mammalian high mobility group I chromosomal proteins. J. Biol. Chem. **265**, 8573–8582.

Regard, F., Lebas, M., and Lescure, B. (1994). Interstitial telomeric repeats within the *Arabidopsis thaliana* genome. J. Mol. Biol. **239**, 163–169.

Richards, E.J., and Ausubel, F.M. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. Cell **53**, 127–136.

Richards, E.J., Goodman, H.M., and Ausubel, F.M. (1991). The centromere region of *Arabidopsis thaliana* chromosome I contains telomere-similar sequences. Nucleic Acids Res. **19**, 3351–3357.

Richards, E.J., Vongs, A., Walsh, M., and Chao, S. (1993). Substructure of telomere repeat arrays. In The Chromosome, J.S. Heslop-Harrison and R. Flavell, eds (Oxford: BIOS Scientific Publisher), pp. 103–114.

Rivin, C.J., Cullis, C.A., and Walbot, V. (1986). Evaluating quantitative variation in the genome of *Zea mays*. Genetics **36**, 899–905.

Roder, M.S., Lapitan, N.L.V., Sorrells, M.E., and Tanksley, S.D. (1993). Genetic and physical mapping of barley telomeres. Mol. Gen. Genet. **238**, 294–303.

Schildkraut, C.L., Marmur, J., and Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. J. Mol. Biol. **4**, 430–433.

Schwarzacher, T., and Heslop-Harrison, J.S. (1991). In situ hybridization to plant telomeres using synthetic oligomers. Genome **34**, 317–323.

Schwarzacher, T., and Heslop-Harrison, J.S. (1995). Molecular cytogenetic investigations of meiosis. In Kew Chromosome Conference IV, P.E. Brandham, ed (London: HMSO), in press.

Schwarzacher, T., Leitch, A.R., Bennett, M.D., and Heslop-Harrison, J.S. (1989). In situ localization of parental genomes in a wide hybrid. Ann. Bot. **64**, 315–324.

Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. Science **191**, 528–535.

Stephan, W., and Cho, S. (1994). Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. Genetics **136**, 333–341.

Strachan, T., Webb, D., and Dover, G.A. (1985). Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. EMBO J. **4**, 1701–1708.

Toledo, F., Le Roscouet, D., Buttin, G., and Debatisse, M. (1992). Co-amplified markers alternate in megabase long chromosomal inverted repeats and cluster independently in interphase nuclei at early steps of mammalian gene amplification. EMBO J. **11**, 2665–2673.

Vershinin, A., Svitashev, S., Gummenson, P.-O., Salomon, B., von Bothmer, R., and Bryngelsson, T. (1994). Characterization of a family of tandemly repeated DNA sequence in *Triticeae*. Theor. Appl. Genet. **89**, 217–225.

Wang, K., Koop, B.F., and Hood, L. (1994). A simple method using T4 DNA polymerase to clone polymerase chain reaction products. BioTechniques **17**, 236–238.

Weimark, A. (1975). Heterochromatin polymorphism in the rye karyotype as detected by the giemsa C-banding technique. Hereditas **79**, 293–300.

Wevrick, R., and Willard, H.F. (1989). Long-range organization of tandem arrays of satellite DNA at the centromere of human chromosomes: High-frequency array-length polymorphism and meiotic stability. Proc. Natl. Acad. Sci. USA **86**, 9394–9398.

Wevrick, R., Willard, V., and Willard, H.F. (1992). Structure of DNA near long tandem arrays of α-satellite DNA at the centromere of human chromosome 7. Genomics **14**, 912–923.

Willard, H.F., and Waye, J.S. (1987). Hierarchical order in chromosome-specific human α-satellite DNA. Trends Genet. **3**, 192–198.

Willard, H.F., Waye, J.S., Skolnick, M.H., Schwartz, C.E., Powers, V.E., and England, S.B. (1986). Detection of restriction fragment length polymorphisms at the centromeres of human chromosomes by using chromosome-specific satellite DNA probes: Implications for development of centromere-based genetic linkage maps. Proc. Natl. Acad. Sci. USA **83**, 5611–5615.

Wu, K.-S., and Tanksley, S.D. (1993a). Genetic and physical mapping of telomeres and macrosatellites of rice. Plant Mol. Biol. **22**, 861–872.

Wu, K.-S., and Tanksley, S.D. (1993b). PFGE analysis of the rice genome: Estimation of fragment sizes, organization of repetitive sequences and relationships between genetic and physical distances. Plant Mol. Biol. **23**, 243–254.