# *De Novo* Evolution of Satellite DNA on the Rye B Chromosome

**Tim Langdon,* Charlotte Seago,* R. Neil Jones,* Helen Ougham,† Howard Thomas,† John W. Forster*,‡ and Glyn Jenkins***

**Institute of Biological Sciences, University of Wales, Penglais, Aberystwyth, Ceredigion SY23 3DD, United Kingdom, †Institute of Grassland and Environmental Research, Aberystwyth SY23 3EB, United Kingdom and ‡Plant Biotechnology Centre, Agriculture Victoria, La Trobe University, Burdoora, Victoria 3083, Australia*

Manuscript received July 26, 1999
Accepted for publication October 18, 1999

ABSTRACT

The most distinctive region of the rye B chromosome is a subtelomeric domain that contains an exceptional concentration of B-chromosome-specific sequences. At metaphase this domain appears to be the physical counterpart of the subtelomeric heterochromatic regions present on standard rye chromosomes, but its conformation at interphase is less condensed. In this report we show that the two sequence families that have been previously found to make up the bulk of the domain have been assembled from fragments of a variety of sequence elements, giving rise to their ostensibly foreign origin. A single mechanism, probably based on synthesis-dependent strand annealing (SDSA), is responsible for their assembly. We provide evidence for sequential evolution of one family on the B chromosome itself. The extent of these rearrangements and the complexity of the higher-order organization of the B-chromosome-specific families indicate that instability is a property of the domain itself, rather than of any single sequence. Indirect evidence suggests that particular fragments may have been selected to confer different properties on the domain and that rearrangements are frequently selected for their effect on DNA structure. The current organization appears to represent a transient stage in the evolution of a conventional heterochromatic region from complex sequences.

CEREALS, like other higher eukaryotes, often contain very large amounts of repetitive DNAs, which make up the bulk of the genome in species such as maize, barley, wheat, and rye. It is still not clear whether these nongenic DNAs play useful roles in genome function. The majority of repetitive DNA in these species is derived from long terminal repeat (LTR) retrotransposons (SanMiguel *et al.* 1996; Bennetzen *et al.* 1998; Panstruga *et al.* 1998), which are potentially capable of explosive amplification and are intrinsically likely to cause genome expansion. This may be an irrevocable process, as there does not appear to be an efficient mechanism to reverse such colonization, which occurs predominantly at dispersed intergenic sites (Bennetzen and Kellog 1997). The proportion of a genome composed of these elements is, therefore, as likely to reflect the species' evolutionary history as any current selective pressures. A more direct reflection of current pressures or processes may be provided by the distribution of the satellite DNA families that are usually concentrated in discrete heterochromatic domains and that may be highly variable between genomes. These domains are capable of rapid quantitative changes under experimental conditions (see, for example, Karp *et al.* 1992), and phylogenetic distributions of specific satellite families indicate that extensive amplification and elimination also occur in natural populations. Some satellites may be species or family specific, while others may show patchy distribution across wide phylogenetic distances (Grebenstein *et al.* 1995; Vershinin *et al.* 1996; Nagaki *et al.* 1998). To some extent, these distributions fit the library hypothesis (Salser *et al.* 1976), in which a range of satellite sequences are present in related genomes, with species-specific patterns of family amplification occurring. However, while most characterized satellites do appear to be derived from existing tandemly repeated sequences, their evolution also involves frequent sequence change or rearrangement and in some cases a nonsatellite origin can be identified (Pasero *et al.* 1993; Rossi *et al.* 1993). It has been suggested that abundant families may have been selected for amplification because of their ability to specifically bind nuclear proteins, for example, during mitotic transitions (Csink and Henikoff 1998); on the other hand, a relatively consistent property of satellite sequences is the presence of bent DNA structures (Fitzgerald *et al.* 1994), and evolutionary success may simply reflect the ability to form compact inert heterochromatin. Only rarely have satellites been linked to specific activities such as en-

*Corresponding author:* Glyn Jenkins, Institute of Biological Sciences, Cledwyn Bldg., University of Wales, Aberystwyth, Penglais, Aberystwyth, Ceredigion, SY23 3DD, United Kingdom.
E-mail: gmj@aber.ac.uk

hanced recombination or segregation distortion (Moschetti *et al.* 1996), although heterochromatic domains display neocentromeric activity in both rye and maize (Yu *et al.* 1997).

Cultivated rye (*Secale cereale*) is a good model for studying satellite evolution, as it has particularly prominent heterochromatic domains located subtelomerically on most chromosome arms. The domains are made up principally of a small number of satellite families (Jones and Flavell 1982; Vershinin *et al.* 1995). The repeat unit of each family is relatively short (120–630 bp) and organized into relatively homogeneous long tandem arrays. The most ancient family (as gauged by its occurrence in other species) is most centromere proximal in the subtelomeric blocks; it also has the shortest monomer length and the simplest head-to-tail array organization (Vershinin *et al.* 1995). In contrast, the more distal satellites display considerable monomer length variation or unexpectedly complex head-to-head organization (Vershinin *et al.* 1995).

Rye is also notable for the frequent occurrence of a supernumerary, or B chromosome, which is found in many populations throughout its geographical range. A single rye B chromosome contains the equivalent of 10% of the host haploid DNA content, but no major genes have been located on it, and its presence appears to alter the host phenotype only via nonspecific nucleotypic effects (reviewed in Jones and Puertas 1993). It has been suggested on theoretical grounds that the addition of B chromosomes may improve nuclear organization in rye (Ostaschevsky 1996), although it is more generally believed to be "selfish." The B chromosome appears to be optimized for its own nuclear function, as variants that arise at a significant frequency in laboratory stocks are not found in the wild nor are they maintained in experimental populations. Much of the B chromosome, including the centromere, is closely related to the host A genome (Tsujimoto and Niwa 1992; Wilkes *et al.* 1995), but a subtelomeric domain on the long arm fails to show significant cross-hybridization to rye DNA in *in situ* hybridization. This is the most polymorphic domain in natural populations and appears at metaphase to be the physical counterpart of the heterochromatic subtelomeric domains found on the rye A chromosomes. Surprisingly, none of the A satellite families are found within the B chromosome heterochromatin. Instead, at least two novel families are found at a high copy number. Both are longer than typical satellites (1.1 and 3.9 kb). Here, we describe other atypical complexities in this domain. This report shows that the two sequence families have been assembled from fragments of a variety of sequence elements, and it proposes that they have been recently recruited to a heterochromatic role. We discuss their current organization in terms of a transitory step in the evolution of a more conventional satellite domain composed of homogeneous arrays.

## MATERIALS AND METHODS

**Plant material:** All somatic and meiotic materials were taken from a local experimental B population of *S. cereale* containing B chromosomes ($2n = 2x = 14 + B$).

**DNA sequencing and molecular methods:** Isolation of the E3900 clone has been described previously (Blunden *et al.* 1993). Sequencing of both strands was carried out using fluorescent labeling (ABI Thermosequenase) and E3900-specific primers. The rye genomic library was constructed by ligating partially digested *Sau*3A DNA (15–20 kb, gel purified) from a plant containing four B chromosomes to λEMBL3 arms (Stratagene, La Jolla, CA). The ligation mix was packaged using a Stratagene Gigagpack Gold kit and used to infect *Escherichia coli* strain SRB. Genomic DNA was made using a modification of the method of Raeder and Broda (1985).

Oligonucleotides and conditions used for PCR were as follows: *crw-R39* UTR and *gag* recovery, T3LX (TATCAAGTGGT ATCAGAGCCAGG) and AW39R (CCATCTCCTTGTARTAYT CSTCAAC) or T3LX and 3900-13(CTGATTGCTTAATCTAT CTTTAC), 96 × 35 sec, 7 cycles of 92 × 40 sec, 48 × 1 min, 68 × 4 min, 22 cycles of 92 × 40 sec, 48 × 10 sec, and 68 × 3 min. *Crw-R39*-related downstream sequences were as follows: AW37 (TATGTKCTKATHTGGTGGGAYCARAT) and BOTYR (GGCATGACAAGCCACTCATA), under conditions as described above.

Probes were made by single-primer PCR labeling of PCR fragments from the 3900 plasmid template: 15OUT (CAATAT TGTGAGTGTTTTGCGA) was used to label the 15OUT/600OUT (GTCTTTCTTCCAACTATCTTT) product (3900 sequence bases 2287–2779), and 800R (CGCACTAGCCTACGG ATAGA) was used to label 800R/15HR2 (GTTCTATACTAGA TCCAATAA; 915–1584).

**Fluorescence *in situ* hybridization:** Preparation and pretreatment of the cytological preparations and fluorescence *in situ* hybridization (FISH) were performed according to published procedures (Heslop-Harrison *et al.* 1991; Leitch and Heslop-Harrison 1992; Pan *et al.* 1992; Zhong *et al.* 1996). Briefly, probe DNA was labeled either with digoxigenin-11-dUTP or biotin-11-dUTP and was hybridized to pretreated chromosome preparations overnight at 37° in the presence *inter alia* of 50% deionized formamide in an Omnislide *in situ* hybridization system (Hybaid). Slides were washed stringently in 20% (v/v) formamide in 0.1× SSC at 42° before probe detection with FITC-conjugated antidigoxigenin antibodies or avidin-rhodamine, as appropriate. Amplification of the signals was effected either by FITC-conjugated secondary antibodies or by anti-avidin-biotin followed by a second round of avidin-rhodamine binding. The chromosomes were counterstained in DAPI and mounted in Vectashield. Fluorescent images were captured by a cooled CCD camera, assigned false color, and manipulated uniformly in Adobe Photoshop.

**Sequence analysis:** Database searches were performed with BLAST, and further sequence analysis was carried out with the Genetics Computer Group (Madison, WI) programs. Sequence alignments were displayed using GeneDoc (K. B. Nicholas, and H. B. Nicholas, distributed by the authors). Curvature predictions were made using the bend.it server (http://www2.icgeb.trieste.it/~dna/index.html) using the DNase I-based bendability parameters of Brukner *et al.* (1995) and the consensus bendability scale (Gabrielian and Pongor 1996).

## RESULTS

**Complex organization of the B chromosome subtelomeric domain:** Two repeat families (D1100 and E3900)

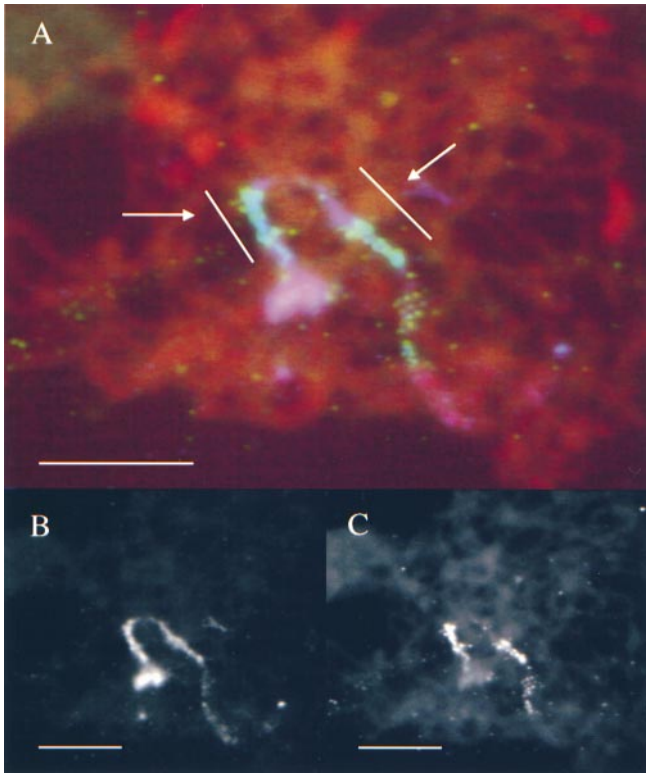


Figure 1.—Meiotic prophase chromosomes of rye, hybridized with E3900 and D1100 clones. Complex organization is seen, including an apparent large duplication (arrow). (A) Simultaneous display of DAPI-stained DNA (red), E3900 sequences (blue), and D1100 sequences (green). (B) Blue channel alone (E3900). (C) Green channel alone (D1100). Bar, 5 μm.

have previously been shown to be major components of the subtelomeric domain of rye B chromosomes (Sandery *et al.* 1990; Blunden *et al.* 1993). Southern analysis indicated that the unit size of each family is relatively constant (1.1 and 3.9 kb, respectively) and that most members of each are arranged in tandem arrays (Sandery *et al.* 1990; Blunden *et al.* 1993). The simplest mechanism with which to generate this pattern is unequal sister chromatid exchange (Smith 1976), which would eventually result in large homogeneous blocks of particular families, as seen for the A genome satellites (Vershinin *et al.* 1995). FISH of D1100 and E3900 probes to the extended chromosomes found at meiotic prophase demonstrates a more complex arrangement (Figure 1). Both sequences are distributed throughout the domain, but marked local variations occur, suggesting the presence of distinct subdomains, with E3900 being particularly highly concentrated at the most distal subdomain. A megabase-scale duplication of alternating D1100- and E3900-rich subdomains appears to have occurred (Figure 1, arrows). Finally, there are regions of the domain in which neither repeat is present. This is most striking in the proximal region of the domain, where a spotty distribution of D1100/E3900 hybridization appears to represent isolated islands of these sequences. It is difficult to reconcile this range of distribution patterns with simple expansion of tandem arrays through recombination at homologous sites, even if complex higher-order or hybrid repeats are considered.

**TABLE 1**

**Characterization of *Eco*RI restriction fragments from rye genomic clones**

| Clone | Band size (kb) | E3900 homology | D1100 homology | Hybridization specificity |
|---|---|---|---|---|
| λGD1100-1 | 3.2 | − | + | B specific |
| | 2.5 | − | + | A + B |
| | 2.1 | − | + | B specific |
| | 1.1[a] | − | + | B specific |
| λGE3900-2 | 3.9[a] | + | − | B specific |
| | 2.7 | + | − | B specific |
| | 2.1 | − | − | A + B |
| | 0.5 | + | − | B specific |
| λGE3900-4 | 6.5 | + | − | A + B |
| | 3.9[a] | + | − | B specific |
| | 3.2 | + | − | A + B |
| | 2.7 | + | − | B specific |
| λGE3900-5 | 2.7 | + | − | B specific |
| | 2.6 | − | + | B specific |
| λGE3900-6 | 2.6 | − | + | B specific |

[a] Correspondence of bands to the predicted size of canonical B-chromosome-specific families.

The structural complexity revealed by cytogenetic methods was also seen in genomic clones. A partial library was constructed from the DNA of a rye plant containing four B chromosomes in the vector λEMBL3 and was screened with the E3900 clone. A total of 12 positive plaques were selected, from which four clones (designated λGE3900-2, -4, -5, and -6) were chosen for further characterization on the basis of their distinctive patterns in initial *Eco*RI digests. The library was rescreened with the D1100 clone and a number of positive plaques were recovered. Each gave the same pattern on *Eco*RI digestion, and a single clone (designated λGD1100-1) was chosen for further characterization. Some restriction fragments of these clones were of sizes predicted from previous Southern analysis of the B-chromosome-specific families. However, other fragments (*e.g.*, 6.5-kb band in λGE3900-4) do not correspond to any known size variants. The identity of these fragments was investigated by Southern hybridization with D1100 and E3900 clones; individual fragments were also recovered from gels and used as probes on genomic DNA extracted from rye plants with B chromosomes (+B DNA) or without B chromosomes (0B DNA). Table 1 summarizes the results.

There are three conclusions from this analysis. First, each clone contains more than one size class of *Eco*RI fragments derived from the D1100 or E3900 families,

```
2100rev:      ......CATCAATCACCAAAAAGGGGGAGATTGTAAGTGCATCTAGTGCCCC  46
hv2100:   599 TGTTGTCATCAATCACCAAAAAGGGGGAGATTGTAAGTGCATCTAGTGCCAC 650
prem2:   8107 TGTTGGCATCAATCACCAAAAAGGGGGAGATTGAAAGGGAATTAGGCTTACA 8158
ropie:   7665 TGTTGGCATAAATCACCAAAAAGGGGGAGATTGAAAGGGAAATGTGCCTTTG 7715
```

Figure 2.—pRAB2100 contains similarity to the polypurine tract region of the maize copia-like retrotransposons, Prem2 and Ropie. More extensive similarity is found with a sequence upstream of a barley proteinase gene (Z97022; polypurine tract homology only shown).

and none represents a simple monotonous array. Second, two of the clones identified as containing E3900 also contain sequences related to the D1100 family, indicating that the two B-chromosome-specific families are frequently contiguous. Third, several fragments contain sequences that are not B chromosome specific but show similarity to sequences from rye A chromosomes. Of the four bands showing this behavior, three also show hybridization to B-chromosome-specific bands diagnostic for the characterized families, and they presumably correspond to junctions between B-chromosome-specific elements and repetitive sequences common to A and B chromosomes. The fourth fragment (λGE3900-2, 2.1 kb) does not hybridize to either D1100 or E3900.

This last fragment was subcloned into pUC18 and designated as pRAB2100. Southern hybridization patterns of this clone with digested 0B and +B genomic DNAs indicate that it is part of a large medium-copy-number repeat family present on both A and B chromosomes (not shown). B-chromosome-specific fragments identified by the probe suggest that divergent members of the family have been amplified on the B chromosome. The probe also hybridizes to sequences in wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*). Sequencing of the ends of this clone indicated that it was homologous to a sequence found in the barley genome. A highly conserved motif found in both sequences shows identity to the polypurine binding site of the maize copia-like retrotransposons PREM2 and Ropie (Figure 2). The pRAB2100 sequence is therefore likely to be derived at least in part from the LTR of a related retrotransposon. FISH to rye metaphase chromosomes with pRAB2100 demonstrated that it is dispersed throughout both A and B genomes (not shown).

**Unusual and contrasting nucleotypic behavior by E3900- and D1100-rich domains:** The subtelomeric domain of the rye B chromosome has been described as heterochromatic as a consequence of its highly condensed appearance in mitotic and meiotic metaphase karyotypes (Figure 3A). However, a dispersed appearance of D1100-containing domains in some interphase cells has been described (Morais-Cecilio *et al.* 1997), and this is typically seen in the outbred B chromosome line used in the current work. This contrasts with the constitutively condensed appearance of the A subtelomeric heterochromatin, which is visible as chromocenters throughout interphase (Figure 3B). The contrast is particularly striking as the B-chromosome-specific domain appears to share the property of the A domains of localization to the nuclear periphery, so that the extended B chromosome domain occupies a disproportionate surface area and appears to force the A chromosome domains into clustered territories. We have found that the B-chromosome-specific domain shows further structural differentiation into the more terminal E3900-rich subdomain, which is decondensed even at very late and very early stages of interphase in both meiotic and mitotic tissue, and the more proximal D1100-rich subdomains, whose appearance varies from dispersed to thread-like, but is consistently more highly condensed than the E3900 region (Figure 3C). The dispersed E3900 subdomain is frequently associated with a discontinuous "satellite" at meiotic prophase, although we are not able to resolve a chromatin link between the two (Figure 3D).

**The E3900 clone contains a retrotransposon fragment:** The isolation of the E3900 clone has been described previously (Blunden *et al.* 1993). The cloned DNA is 3984 bp long (Figure 4). Its sequence contains no significant internal redundancy, other than a single direct tandem repeat of the sequence between bases 1189 and 1270, and so does not represent a higher-order arrangement of satellite DNA.

The only extensive region of similarity to previously characterized sequences lies between bases 2853 and 3384, which encodes a partial reading frame for the *gag* protein of an LTR retrotransposon, most closely related to a highly conserved Ty3/gypsy family, *crwydryn*, which has colonized the centromeres of the Poaceae (T. Langdon, C. Seago, M. Mende, M. Leggett, H. Thomas, J. W. Forster, H. Thomas, R. N. Jones and G. Jenkins, unpublished results; Presting *et al.* 1998; Figure 5; see below). This region is not contiguous in the E3900 clone, which was derived by *Eco*RI digestion of genomic DNA, and is therefore cut at the *Eco*RI site within the *gag* regions of adjacent tandem repeats. However, the integrity of this sequence in the genome was confirmed by PCR with oligonucleotides that span the *Eco*RI site (not shown). The open reading frame is continuous throughout its length, and the degree of conservation with *gag* genes from divergent species implies that the sequence has recently been derived from an active element, which we have designated as *crw-R39*.

The remainder of the E3900 clone appears to be unrelated to this *gag* sequence. We were unable to find even fragmentary homologies with other canonical retrotransposon protein components, which usually show levels of conservation far higher than those of *gag*. In

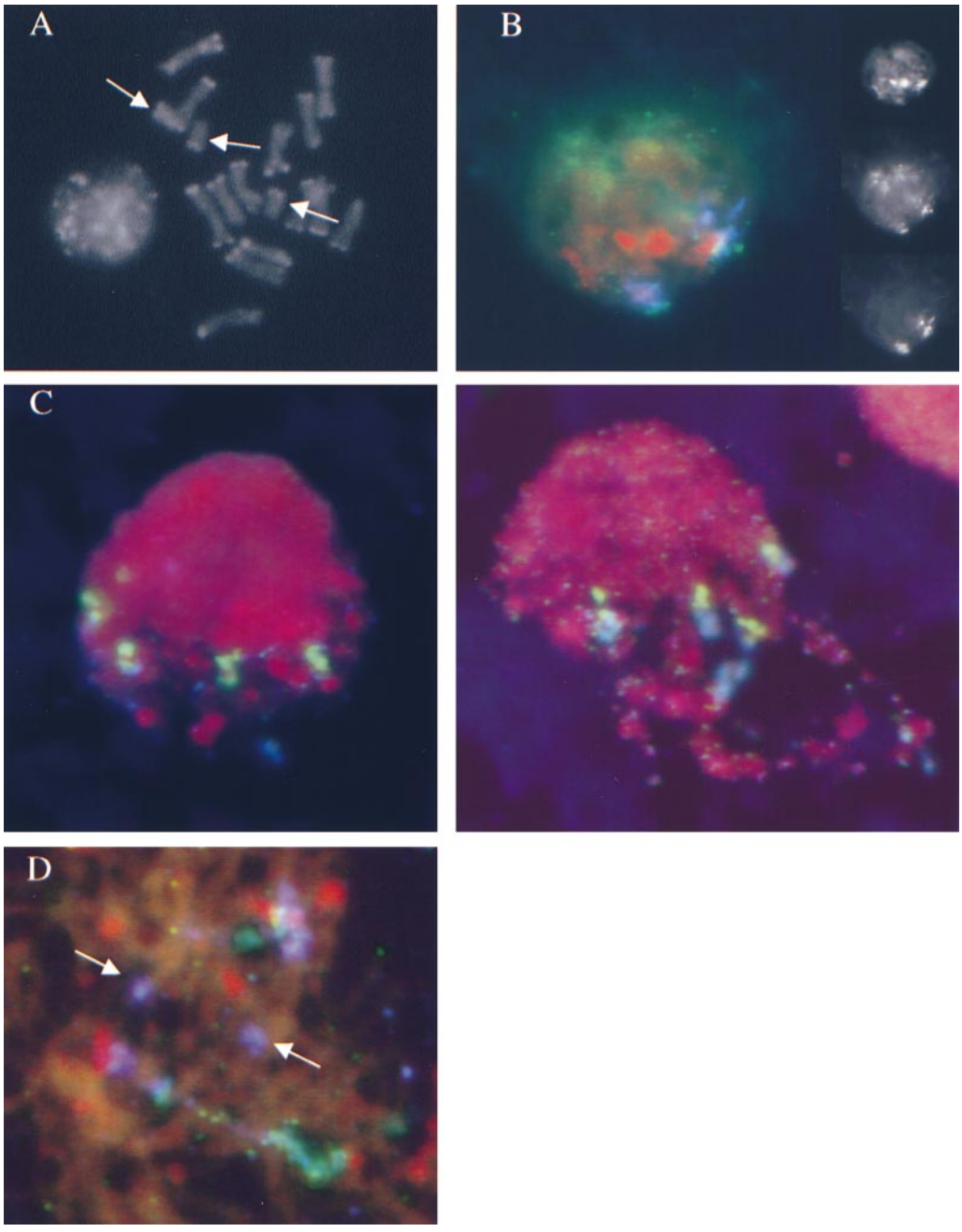


Figure 3.—The B chromosome subtelomeric domain has distinctive cytological properties when compared with equivalent regions of the A genome. (A) At somatic C metaphase, DAPI-stained subtelomeric domains of both A and B chromosomes are highly condensed. B chromosomes are identified by arrows. (B) In pollen, mother cells at interphase, DAPI-stained chromocenters (red, main picture and top right inset) representing condensed A genome heterochromatin contrast with dispersed centromeric (green, upper region main picture and mid-right inset), and B-chromosome-specific sequences (E3900 green, lower region main picture and mid-right inset; D1100 blue, main picture and lower right inset). (C) At somatic interphase, D1100-rich subdomains (green) are more compact and show a greater degree of organization than E3900-rich subdomains (blue). (D) Extended chromosomes at meiotic prophase (see also Figure 1) demonstrate that there are equivalent amounts of the D1100-rich (green) and E3900-rich (blue) subdomains, and they show that the latter subdomains are more dispersed, appearing as satellite blocks physically disassociated from the main chromosomes (arrows).

addition, there is neither similarity to the LTR sequences associated with related *gag* genes nor duplication of sequences flanking the *gag* reading frame, as would be expected for LTRs. Taken together with the absence of upstream tRNA-primer-binding sites, it is unlikely that the *gag* sequence retains any potential for mobility (see below and discussion). The E3900 family, therefore, does not appear to correspond to a scrambled, "slave," or otherwise degenerate retrotransposon despite the presence of this single fragment.

**The E3900 *gag* fragment has been generated by chromosomal rearrangements:** The downstream limit of *gag*-related sequences is defined by a 73-bp insert, flanked by a 13-bp duplication (12/13 base identity). The inserted sequence is strikingly similar to a coding region of the 2-oxoglutarate/malate translocator genes of rice and millet (Taniguchi and Sugiyama 1996; Figure 6A), which suggests that the equivalent region of the rye translocator homologue may have been recruited as "filler" DNA during repair at this site. This suggestion is supported by the 7/8-bp match between the 3′ part of the duplication and the relevant 5′ region of the rice translocator, which would allow invasion of the translocator sequences by the broken *gag* end. The 13-bp duplication is expected to be derived from the *gag* gene as a further filler sequence, following the pattern found at the junctions of spontaneous deletions within the waxy (*Wx*) gene of maize (Wessler *et al.* 1990), presum-

```
CTATCCTGTCCGGTGTCCCTTTTGGTTCAACTGAGGCGCCGGAGCGCCGTCTCGAACTCGGGCAGCTCCGCGTACTAGCT     80
ACCCTAGCCCGTCGTCCTATTTCCGTGTAACCATTAACAGTGTTTCGCCGACCCAACCCAATGGAGGGGCGGGGGGATTT    160
GCACGCTCCCATCTTGCTTTCCAAGGCGCTTGGTGCAAACCTTTGCTGGGAGTCTCGGATCCTTTTCATTCGCCGAGCTA    240
CTGGCTTTTCGGGAGGGGGTAATCTTTGCTCAGCTCCGAGGTTACTCACATGTGATAGGCAATTCCAGATCGATTGCATC    320
GCCTATCTTTGGAGAATTATCGGACATTGCACCACACTTCGTTTTCTTTTTCGGTTCGACATGTCGGTAGGAACTTGAAT    400
GTGCCGGCCCATCTCTCTGCTCAGCGTGCCTGTACGCTCGATGGGATGGATACGTTGGCTAGTTgttagccctgatttcct    480
acgcactagcctacggatagactgtaaccggctactgtcattctgaataaagtcctaaattcgctagaaaaaaaaaacat    560
acatgcatgcactgcgcctcagtctacaattTAGTTAGTTTGCTTAGTTAGTTTGTATTTCTCATTAGAGTTATAGTTAAC    640
GAGACCGCCCTAGCTGCCATGATACAGCATGCATGCATCAACTTTACTCGATCCATCGACTAGACCGCGCAAGTTTAATT    720
TCCAACTTAGCGGGCCCAGGGTTCCTCATCGACCATGGTCCGACCGATGACATGGGGTTGAGCGTATCATTGCCGTGAGT    800
TGAGACGTGTCCCGTGGTTGCCGACCGGTGGTGGAGAGGCGATGACCGCCAAAGAGAGAAATATGCGACGACAGATGAAC    880
TAGATCGATAAAAAGCTCTCTAGCTAGAGGGGATTCCTCCTATCATGTGACCGACACGATGTAGTCAACATGTCATGGTG    960
TCCGAGATTGTCCCGGACAAGTAGACTGGTTGTAGCGACGAGGTCCATATGATTGGTTGTACACGGTGAGAGACATGTTG   1040
TTTTGTCGGCAAGGATCCCTTTCTCCTGGCCACGAGTTGTTTCTTTCTACGTGTTACAATGAATGTGGTTTCATGTGTTA   1120
GTGGGAAAGGGGACTAGGCAATACAGAGAAAATAATACCGCAAAAAAAGATGTCATGTCCAAATGAACATGTTTACATCT   1200
AGGGGTGTTATTGGGTCGATCCGAGCATAACCACCATAAGGAGAAATTTAGGAAATCCAACAAGCCAATGTTTAGATCTG   1280
GGGGATTGGGTCGATCCGACCATAACCACCGTAAACATGAATACAGGAAATCCGCCAAGCACATGGGAGAAATGAAATGA   1360
AATCCCAACCAGGTGCACATAAAAATCATAGAACAAATTTGGGATTGTTACCACGCTGAGAAGGCGACCAATGCACGTCA   1440
ACATTAGATGGATTTTTGTGACATGGTACAAGATTGTTGTGGACAAGCAAAAAGAGGTGGGAGTGACGAGGCTCTCGCGA   1520
TCGGTCAGTCAACAACCGCGAAACTTATTGGATCTAGTATAGAACGGCTGGGATGAGACGAACAGGAGCCTCGATCCGTT   1600
GCATGTCCTACTTGCTAAAGCCCAAGACAATCAAACTATTATCGCCGGTCGCTCCCGATTTTGTAGTTGGAATAATAATG   1680
ATGTGTGCACCAATGCTGCAGGTTTACTTGGAGCGACTCCCAAGTGTTGAACAGAAAACCAAGCACGACCATTTGTTTTG   1760
AGCGCAAGTATACCAGAGGCTTCACATTATAATCTTCCGACTATGTTACGGCTTGGATCTTTCCGCTTTGAGGTTTCGCC   1840
GATGGAGCCGAAGAAGGACACGGTGAAGCCAAACTTCAGAACCCCTCCGGCCTGCACAAATTTTGGTAGTATTTCATGTC   1920
CGGTGAGGTAGTTTCAAGCGTGCAAAGCAAATAAGCCCTTCCTTTGTTTTGAATTAGTCACGTGTGATCAACGTGGCCGG   2000
ATGATTAATTAGGTTCCTAATTAATGAATACTAGAATAAAGGGTCACGTACAAGTTGTGCTAGTTAGCTTGTTTTTTTTT   2080
TCTCTCTCTCCGAGATGTAAAAGGAGAGGATCTCGGTGTTAGTTTGAGACTTTGAGTTGGTTAGTAGTCTAGGCCCTTAC   2160
GGCAGGTGTGACTAGGATATTAAAGGCGTCGTCCTGGGTCTTGTACGGGGATTAGTTTTGGAAGATTGATTTCTTATCGA   2240
GAGTTACCAGTTGGGGAACGCCAATATTGTGAGTGTTTTGCGATCCTTGATCGGGGATTGTTTCTGCTACGCTGGTGAGA   2320
TTAACCTTGGGTTAATTTCGATGCAAGTCTACTCCTTTGGTTTCATGCCATCTCCTTGATCCAAAGGCCTTTTCTTCTTC   2400
TACGTTTGCTATCTATTTTTTATCCGACCGTCGTTGCCGCCGTTCGTTGTTGTTTTGCAACTACCTATGTTGCACTTTCA   2480
GCTTGGTTTGTCTTCCCAGAGAGGAAACTTCACGTGAGGGATCACCACCACCGTCTCTGACTCTCTGTAAAGCTCAGCAT   2560
ACTACAATCCTTCCATGATACCAAAGGACGCACATGCAGATGCATGCGTGGCTCTTCTCCATGGAGCTTTTCACCTGCCC   2640
ATATGGCACGTAGGGAGAAGAAAGAGAAAGAGAACCGTGTGTGATCTGAATGTTAACATTCATACGAGTATATATATACA   2760
AGTAAAAAAGATAGTTGGAAGAAAGACCCCAATCGGACGTATGGTAACAACTTTCTCTATGCGACCTCTCCTCTCTGACA   2800
CGTAACTCAAGGAGAGGTGTTTTTGCCCTCAAAAATCAAATGGTTTTTGATCACatgatcgccgtttggctgcacaacctc   2880
atgttggagatgATGCTTTTTGATCACGGCGTTGACCAGGCAGCGGGGTACGCTCCACGACCACGACACTATGAGCCTCGT   2960
CCTGAAACTTCCGTTCGTGGCGGAGTGCATGACCGAGTTGGTCAAGCTGTGCGTGTGCCTTTGGATGATGGAATTGGGCG   3040
CATAAGAATTTCAATCCCTTCGTTCTCCGGCAAGTGCGAACCAGAAGGCTATTTGGAGTGGGAGATGCGTGTTGATCAAA   3120
TTTTTGATGCCCATCATTATAGCGAGGAGAAGAAAGTTCAACTTGCTGGAATTGAGTTCACCGGTTACGCGCTAATTTGG   3200
TGGAATCAAATTTGTCGTTCAAGACATCGGCCGACGACTTGGCGAGGTATGAAAGAATTCATGAGGCGACGTTTTGTTCC   3280
TGAGCACTATAAAAGAGATGTGTACAATAAGTTGCAGCGGCTCTCTCAAGGTAATATGAGTGTTGATGAATATTACAAGG   3360
AGATGGAATTGCTTATGATACGTATTCATGCCACGTtggtcgaggactttgcaactacggtcaccaagaagatgtttgctaat   3440
gaggaagttggttccttttactgtttCGTATGGTTCCACTAAAGATAGATTAAGCAATCAGCTGGGCAACGACTGGAGTCA   3520
GCCAGCAGCAAAAACACGTAGAGAAATGGCATCTAGCATACACCATCACCTCCACATACACGAAGGTTGATTGTATTGTT   3600
TGATGGTGATGGTGATGCGTACGTGTGCACCTACCAAACCAAACTGCTCCCCCACCTCagctCTAGCAAGAAACTCCTCC   3680
CCCCCCTGGTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTGGTGCGCCAGGTGATGGATCGAGACAGCGGTCTGCTT   3760
CGGTAACGAAAAAAGGAAGGAAGGAAGGAAGGAAGCCCGCCTGAGCGACGTCGCGGACTGCTGCATGCATGCAGGTGACG   3840
AGGCTCAACTCTCCGACCCAAACTtccctcaaaaaaaaaataaagaaaagaaaaACGTCTCGACCCAAACTGAAACTGCCGT   3920
GTATGTGGAGACAGTTGCAGAATCTTATCATGTTTTTTATTTCCTGTATATACAACTAGTAAAA                    3984
```

Figure 4.—Sequence of the E3900 clone (GenBank accession no. AF222021). The *gag* reading frame is indicated in boldface type, duplicated sequences at SDSA rearrangement junctions are highlighted, and filler sequences are given in lowercase. Imperfect tandem direct repeats at 591–603/604–615 and 1188–1265/1267–1340 are underlined, and short microsatellite sequences associated with rearrangements are in italics (including, at 1251–1255 and 1326–1330, the presumed ancestral motifs of the GAAT array at 1348–1363).

ably generated by a slip mispairing mechanism during DNA synthesis (see discussion). In agreement with this model, the sequences downstream of the second duplication appear to be unrelated to either *gag* or translocator genes. The absence of further translocator-derived sequence is particularly clear, as the peptide motif after the presumed insert is well conserved even between millet and bovine homologues (Taniguchi and Sugiyama 1996).

Evidence for a similar process is seen at the upstream limit of the *gag* fragment (Figure 6B). The conserved reading frame extends 5′ from the region shown in

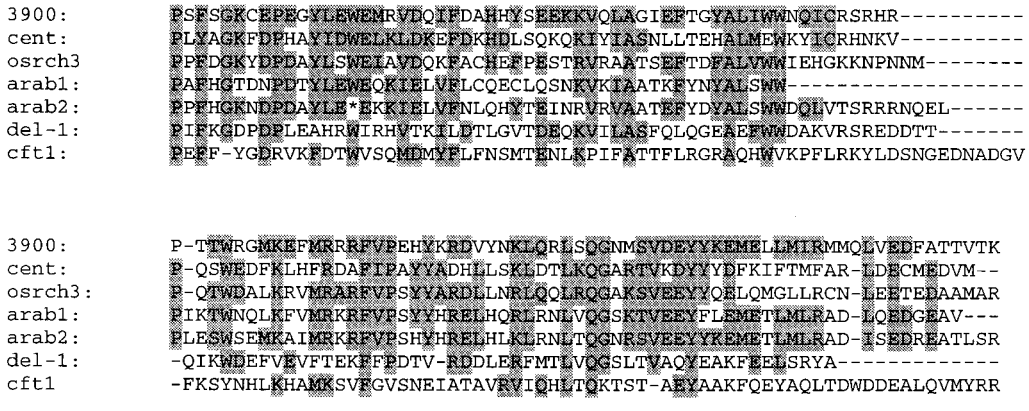


Figure 5.—Homologies of the E3900 ORF. Two centromeric elements, centC (AF078917) and osrch3 (AF0 58905), were described in maize (Ananiev *et al.* 1998b) and rice (Dong *et al.* 1998), respectively. Ty3/gypsy *gag* proteins were derived from two Arabidopsis database entries (arab1, AC006267; arab2, AF128395), and from the del-1 (X13886) and Cft-1 (Z11866) retrotransposons of Lilium and Cladosporium, respectively.

Figure 5B to a stop codon at bases 2850–2852, contained within 14 bp, which duplicates the sequence lying 39 bp downstream. There is no in-frame methionine between this position and the conserved peptide motifs. To confirm that this was the junction of a rearrangement and to clarify the origin of the *gag* sequence, we attempted to recover intact fragments of the original element. The *crwydryn* family, in common with other LTR retrotransposons, uses a $tRNA_{met}$ primer-binding site. We therefore carried out PCR with both 0B and +B rye DNA using a 5′ oligonucleotide based on the *crwydryn* priming site and a degenerate 3′ oligonucleotide based on an E3900/*crwydryn* consensus peptide motif. PCR product of ~900 bp was recovered from both template DNAs; however, two to four times more product was obtained from the +B DNA (not shown). Direct sequencing of the products indicated that the different yields reflected the presence of E3900 *gag* sequences in the +B reaction only. Products were cloned and sequenced; 4 +B clones all contained *gag* sequences identical to E3900, while all 14 0B clones characterized were derived from two related but divergent families. One of the +B clones (designated *crw-R39-1*) was sequenced in its entirety and aligned with E3900. Identity with E3900 is found downstream of the presumed junction (base 2853), but ends abruptly upstream of this point (Figure 6B). An open reading frame (ORF) in this clone extends a further 46 aa; a potential initiator methionine is found 16 aa from the beginning of this ORF. Translation initiation at this site would be consistent with the relative position of the conserved motifs in other *gag* genes shown in Figure 6. We therefore conclude that *crw-R39-1* represents the ancestral sequence of the E3900 *gag* fragment and that the sequence in E3900 has been generated by rearrangements involving slipped mispairing during DNA synthesis, which

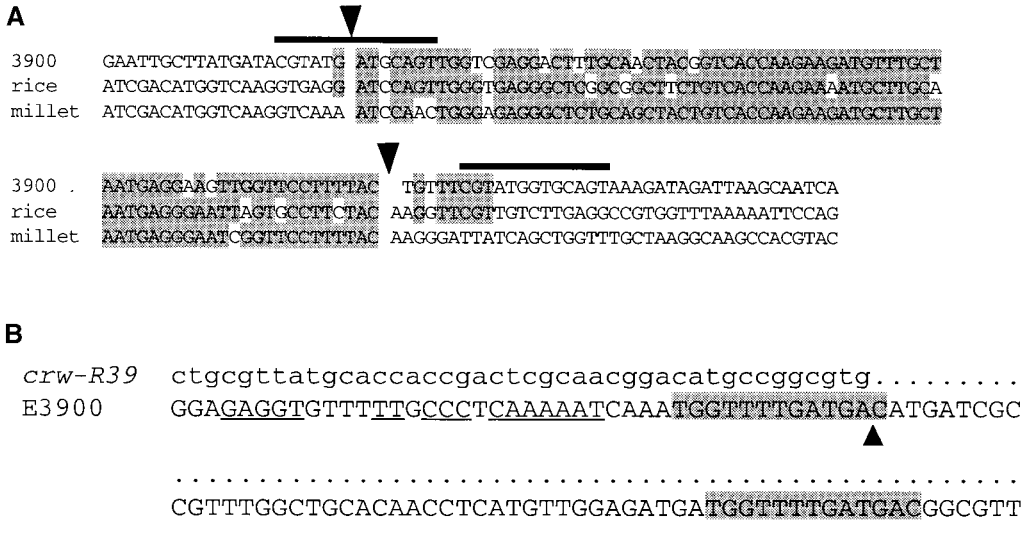


Figure 6.—Sequence analysis of the 3900 *gag* reading frame. (A) A 73-bp insertion in the 3900 *gag* reading frame is highly similar to coding regions of 2-oxoglutarate/malate translocator genes from rice (AU032846) and millet (D45073) (shaded). The insert occurs three bases downstream of a tripeptide motif conserved in the 3900 and *crw-1* reading frames (MIR, ATGATACGT), and it is flanked by a 13-bp duplication (overlined). Part of the duplication is similar to the 5′ region of the transporter sequence. (B) A 14-bp duplication is found at the 5′ junction of the 3900 *gag* gene. The duplication (boxed) is derived from a region 39 bp downstream of the *gag* breakpoint, which is identified by comparison with an intact *gag* sequence. The sequence of the original *gag* gene preceding the breakpoint is shown in lowercase, and bases 2811–2930 of the E3900 sequence are shown below. The breakpoint is indicated by a bar. Underlined bases are identical to those of a second potential instability motif found downstream of the E3900 *gag* sequence. The *crw-R39-1* GenBank accession no. is AF223161.

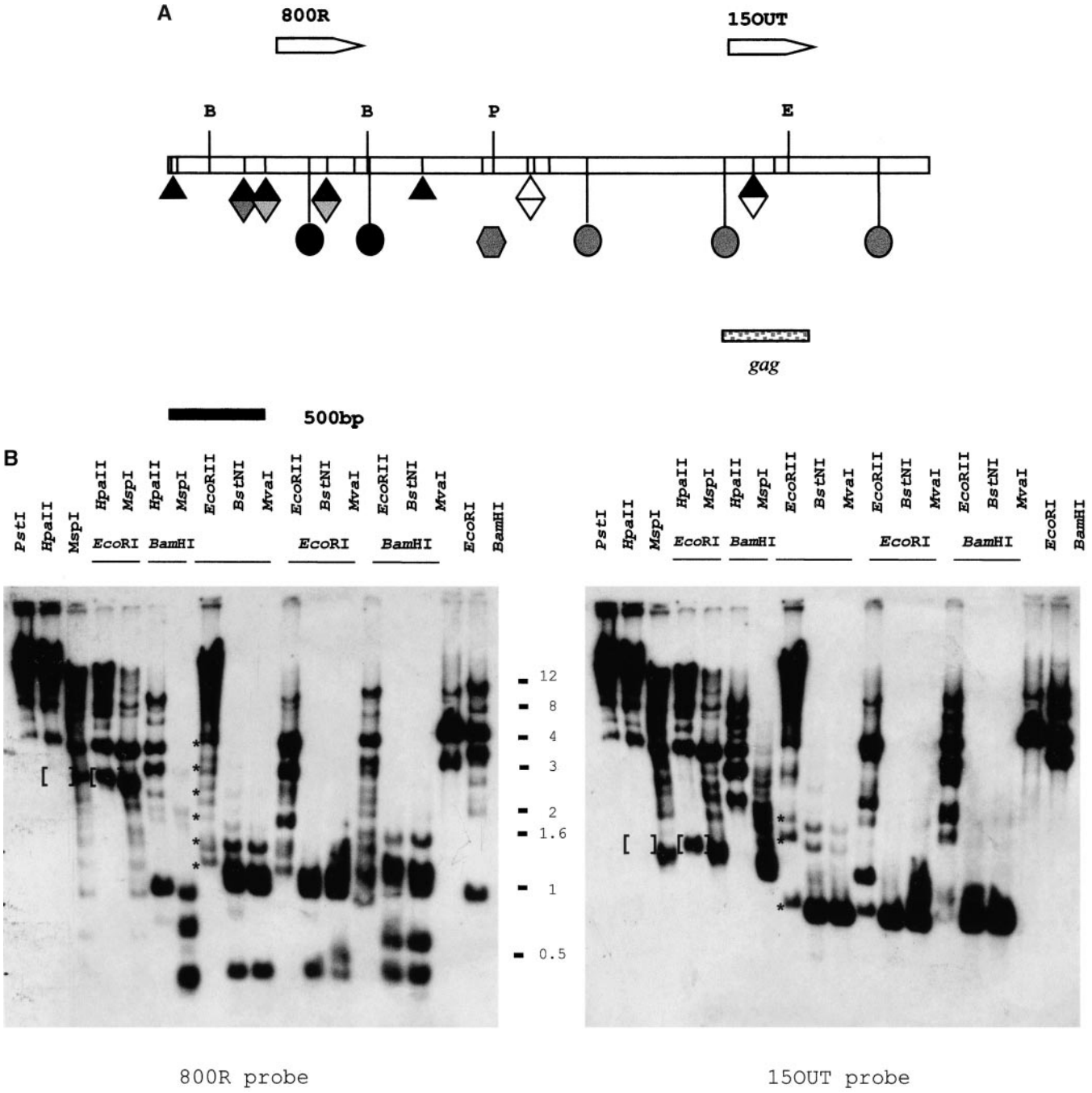


Figure 7.—Methylation of restriction sites in the E3900 family. (A) Location of restriction sites and probe templates in the E3900 clone [vertical lines contained within the box represent *Hpa*II/*Msp*I; vertical lines extending below box indicate *Eco*RII; and vertical lines extending above box indicate *Bam*HI (labeled B), *Pst*I (P), or *Eco*RI (E)]. The position of the *gag* open reading frame is indicated (stippled box). Also indicated is the position of the probes used for methylation characterization. Methylation status is indicated below the box, with shading representing the approximate extent of modification (black, complete methylation; white, no methylation). *Hpa*II/*Msp*I methylation indicated by lozenges—top, *Hpa*II; bottom, *Msp*I. Circles, *Eco*RII; hexagon, *Pst*I. (B) Analysis of methylation in leaf tissue. Specificity of the methylation pattern is indicated by the *Eco*RI/*Hpa*II fragments (brackets), and the relative abundance of fully digested *Eco*RII fragments was detected by probe 15OUT, but not by 800R (asterisks).

have resulted in the deletion of the remainder of the *crw-R39* retrotransposon.

**Sequential evolution of the E3900 *gag* fragment:** We also attempted to recover downstream sequences of the *crw-R39* retrotransposon to better characterize the 3′ junction present in E3900. PCR using degenerate oligonucleotides based on conserved peptide motifs in the E3900 *gag* gene and the *crwydryn* reverse transcriptase generated equal amounts of a 1.9-kb product from equivalent quantities of both 0B and +B templates. Di-

rect sequencing indicated that the product of both templates was indeed very similar and that it was related but not identical to that in E3900. The homology was sufficiently close that, on the basis of the presence of a conserved tripeptide motif (Figure 6A), we were able to confirm the position of the downstream breakpoint as lying within 3 bases of the start of the translocator-like sequence. No similarities were found downstream of this point, indicating that, as with the 5′ junction, the rearrangement had deleted associated retrotransposon sequences.

One possible explanation for the failure to recover *crw-R39* downstream sequences is that only rearranged fragments are present at significant levels. We carried out PCR with the *crwydryn* PBS oligonucleotide described above and an E3900-specific oligonucleotide whose target lies downstream of the 3′ *gag* rearrangement; a 1-kb product was readily amplified from +B but not 0B rye genomic DNA. It therefore appears that intact *crw-R39* elements are rare, even on the B chromosome, but that after the 3′ truncation event, a B-chromosome-specific amplification of the rearranged sequences has occurred. Following the second 5′ rearrangement that removed upstream *crw-R39* sequences, further round(s) of amplification have led to the current situation where the majority of *crw-R39 gag* sequences detectable by Southern hybridization are part of the E3900 repeat. Given that most functional components of the *crw-R39* element are removed by the first rearrangement, it is likely that the *gag* sequences have played a passive role throughout the amplification process and are not themselves responsible for their instability.

**The E3900 family displays unusual methylation patterns:** We used the isoschizomers *Hpa*II/*Msp*I and *Eco*RII/*Bst*NI/*Mva*I to examine the methylation patterns of E3900 in genomic DNA extracted from leaf tissue. Surprisingly, we found that while most *Hpa*II sites screened were modified, as expected for plant repetitive DNA, the site(s) at 1920 and/or 1997 appear(s) to be specifically protected (Figure 7). The *Msp*I digest indicates that there is also relative undermethylation at site 3067. The situation at *Eco*RII sites is more complex, but again, there appears to be differential methylation with a relative lack of modification of the *gag* region and sequences immediately upstream of it. This is demonstrated, for example, by low levels of completely digested fragments detected by the 15OUT probe; only partially digested fragments are seen with probe 800R. The *Pst*I site in the central region (position 1700) is also cleaved in a significant proportion of family members.

**The D1100 family contains a rearranged MITE element:** The published sequence of the D1100 family (Houben *et al.* 1996) contains sequences related to Tnr1, a miniature inverted-repeat transposable element (MITE) originally identified in rice (Tenzen *et al.* 1994). While the origin and activities of MITEs are still unclear, they are common in cereal genomes and it has been found that they show a marked, and possibly functional, association with matrix attachment sites (MARs; Avramova *et al.* 1998). D1100 contains only the core and one arm of the MITE, as determined by comparison with intact elements found in a number of database entries for barley and wheat genes (Figure 8A), the most notable of which are found clustered at a site downstream of the Ost I gene within a 60-kb contiguous barley sequence (Panstruga *et al.* 1998). Three Tnr-like elements are found within ~600 bp at this position, whereas only two more individual elements occur elsewhere in the sequence, one within 2.5 kb of the cluster, immediately downstream of the Ost I gene, and the second some 8 kb away, immediately downstream of the Mlo gene. Three other unrelated inverted repeats are found within the sequence (at 7.6, 15.6, and 28.9 kb); only one shows similarities to other current database entries (with barley limit dextrinase, $P = 7.4e - 10$). The Tnr-like sequence within D1100 therefore appears to be derived from one of the most abundant MITE families in temperate cereals, whose distribution is likely to mirror that of the MAR-linked elements identified in tropical species.

**A potential instability motif is present in E3900 and D1100:** The Tnr-like sequence in D1100 is truncated and appears to have been rearranged by the same slipped mispairing and synthesis mechanism seen in E3900. A 12-bp duplication, separated by 15 bp of conserved Tnr-like sequence, is found at the limit of the D1100 Tnr homology (Figure 8B). Strikingly, a similar motif is found immediately upstream of the 5′ junctions of both the E3900 *gag* fragment and the D1100 Tnr element (Figure 8), although there is no significant sequence similarity elsewhere between the families. The motif is repeated some 400 bp upstream of the Tnr junction in D1100 and at the 3′ end of E3900. A related sequence is present in the maize 180-bp satellite that forms tandem arrays within knob heterochromatin. The second E3900 motif and the 180-bp motif are both partially duplicated (Figure 9); the maize duplication is seen as an occasional 202-bp variant (Ananiev *et al.* 1998a).

## DISCUSSION

**Generation of novel sequences in the B chromosome-specific domain:** The E3900 and D1100 families, which are amplified specifically within the B chromosome subtelomeric domain in the rye genome, are made up of fragments of a number of previously unrelated elements. Those fragment junctions that can be unambiguously identified indicate that a common process has been responsible for their rearrangement, suggesting that the assembly of each family may have occurred within the B chromosome domain itself as a consequence of the domain's inherent instability, rather than


Figure 8.—The D1100 sequence contains a Tnr1-like MITE, which has been truncated by an SDSA repair-mediated event. (A) Similarity of the D1100 sequence to other MITEs. The ends of the elements are particularly well conserved, as shown by similarities to equivalent regions in rice Tnr1B and an element in the *Vicia faba* chitinase gene. Two wheat and five barley elements are shown, including the three clustered elements found near the Mlo gene. (B) The upstream junction of the D1100 MITE. Tnr-like sequences are shaded. A 12-bp sequence conserved in the MITE is duplicated at the rearrangement junction and is partially duplicated downstream (boxes). Sequences duplicated exactly upstream in D1100 are underlined; the duplication may contain an instability motif (Figure 9).

through any intrinsic properties of the ancestral components. This suggestion is supported by a number of observations. First, although both E3900 and D1100 contain fragments of mobile elements, key regions of each element required in conventional mechanisms of mobility have been lost in the most highly amplified family members. Second, the complexity of the genomic clones argues against any simple clonal expansion of the families as the result of the acquisition of novel emergent properties (for example, array expansion as a result of the generation of recombinogenic ends by transposase nicking; Kipling and Warburton 1997). Finally, rearrangements on a megabase scale are also seen within the B-chromosome domain (Figure 1). In total, these observations indicate that some process acts indiscriminately on the B chromosome-specific domain, creating rearrangements with a tendency for duplication rather than deletion.

We propose, therefore, that the E3900 and D1100 families have originated from part of the rye A genome that has undergone a progressive series of amplification and simplification steps as a consequence of its position as the most centromere-distal region in the ancestral B chromosome fragment. The observed concentration gradient of D1100 and, even more strikingly, E3900 (Figure 1) indicates that even within the B-chromosome-specific domain, selective amplification occurs chiefly, if not entirely, at the most distal point. A possible basis for this pattern is suggested by the consistent observation of apparently discontinuous blocks of E3900 at this distal tip during meiotic prophase (Figure 3). While the chromatin linking these blocks may be present although

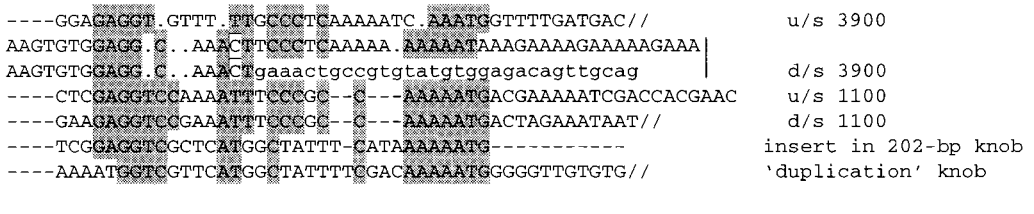


Figure 9.—A potential instability motif is found adjacent to rearrangement junctions. The motif (shaded) is found upstream of rearrangements in E3900 (labeled u/s 3900) and D1100 (d/s 1100) and upstream of preferential insertion sites of retrotransposons in maize knob DNA (duplication knob); junctions are indicated by //. Partial duplications of the motif occur at a second site in E3900 (d/s 3900—the sequence given is contiguous; the second duplication is aligned underneath the first and is shown in capital letters) and in the 202-bp variant of maize knob DNA (insert). A second motif is also present in D1100; it is not known if the flanking sequences are rearranged.

not detectable, it is clear that it must be at best highly stretched and prone to breakage. A number of mechanisms could result in the DNA flanking a breakpoint in this region being used as a template in its repair (for example, by SDSA using a related region of a sister chromatid as a template). If increases in size are favored while the position of the potential fragile site is maintained at a fixed distance from the chromosome end, rounds of sampling will result in ever-decreasing complexity of distal sequences. It is significant that equivalent regions of the rye A chromosomes are also unusual in the quantity of amplified repetitive DNA that they contain, which is far in excess of that seen in related species, such as wheat, and that rearrangements of these satellites are more frequent in distal regions (Vershinin *et al.* 1995). Spontaneous breakage in the heterochromatin of A chromosomes has been reported previously (Gustafson *et al.* 1983), giving credence to the notion that this class of DNA on B chromosomes may be relatively fragile. Therefore, both instability and the drive to expand terminal regions may be general features of the rye genome rather than novel properties of the B chromosome.

**A single mechanism for B-chromosome-specific domain rearrangements?** The junctions between the component fragments of the B-chromosome-specific families are unusually consistent in containing duplications of flanking sequences. Those examples described above are clearly defined by comparison with conserved sequences but similar arrangements involving uncharacterized elements are also likely to be present; for example a duplication of bases 3641–3658 is present 12 bp downstream and immediately flanking a series of simple sequence repeats [$(TTTTC)_4$; Figure 4; also see below]. Duplication patterns of this type have been found to occur in the majority of spontaneous deletions in the maize *Wx* gene (Wessler *et al.* 1990) and during the creation of maize *Ds* elements (Rubin and Levy 1997). These cases are believed to be examples of a specialized version of SDSA repair, a process in which the 3′ ends of a donor template invade a recipient and extend within synthesis "bubbles" rather than generate large stable heteroduplexes of newly synthesized strands (Nassif *et al.* 1994). The relatively short length of contact between donor and recipient facilitates dissociation of the molecules and repair of the break by end joining of the donor 3′ ends, while leaving the recipient molecule unchanged. If SDSA uses a homologous template, the original donor sequence may be reconstituted; however, nonhomologous templates have been found to be used frequently in the repair of double-strand breaks in plants (Gorbonuva and Levy 1997; Salomon and Puchta 1998). Most of the filler DNA incorporated during SDSA repair of these induced breaks is derived from sequences that are not physically linked, however, and does not give rise to the duplication patterns described above. Wessler *et al.* (1990) suggested that duplications may arise during lagging-strand DNA replication when discontinuous DNA synthesis provides both free ends and long stretches of single-stranded DNA in close proximity. Rearrangements occurring at such a well-defined stage may also be consistently resolved in a particular manner as a consequence of the presence of specialized enzymes, RNA primers, and so on. A simple explanation for the E3900 and D1100 rearrangements may then be that DNA replication is in some way impaired in this domain, leading to frequent lesions that are repaired in this specific manner. This explanation is not incompatible with the stretching/breakage model described above, if lesions initiated during DNA replication are not resolved until prophase.

**Preferred sites for rearrangements in amplified sequences:** Although the lesions repaired by SDSA may be created by random domain-wide processes, there appear to be sequences that are preferentially involved in rearrangements. The most prominent of these is the instability motif seen adjacent to junctions in both E3900 and D1100, which is similar to a sequence found near the hot spot for retrotransposon integration identified in the maize knob satellite (Ananiev *et al.* 1998a; Figure 9). Almost all characterized knob insertions have occurred at either of two sites that, like the B-chromosome family junctions, lie approximately one helical turn downstream of the conserved motif. It was suggested that the hot spot may reflect the physical structure of this region (Ananiev *et al.* 1998a). In support of this suggestion, A tracts, such as those found in the motif, have previously been found to adopt unusual conformations (Young *et al.* 1995; Mollegaard *et al.* 1997), and the preferential integration sites of a variety of retroelements are bent or kinked (Muller and Varmus 1994; Pruss *et al.* 1994; Tatout *et al.* 1998). Analysis of B-chromosome-specific sequences using the BEND algorithm, as implemented by the bend.it server (Goodsell and Dickerson 1994; Munteanu *et al.* 1998), further strengthen this proposal. Curvature maxima are predicted at instability motifs in both E3900 (*gag* junction) and D1100 (upstream duplication; Figure 10). It therefore appears possible that distortion of the DNA helix in the vicinity of these motifs is sufficiently great that nicking or melting may occur, providing substrates both for SDSA and integrase reactions.

**Sequence rearrangements in chromatin structure:** Strikingly, the E3900 maximum takes the form of a double peak, separated by 54 bp, which resembles the nucleosome-positioning structure seen in a number of satellite families (Fitzgerald *et al.* 1994), raising the possibility that the motifs help determine the families' conformation as well as susceptibility to rearrangement. Linker regions that determine both nucleosome phasing and sensitivity to chromosome breakage have been described previously (Lanzer 1994). The E3900 double peaks are separated by an SDSA junction, so that the putative nucleosome positioning region has been cre-

ated by the rearrangement. Other potentially curved regions in E3900 are also associated with rearrangements, most notably two regions in the 5′ half of the sequence. In both cases, two events appear to have occurred at a single motif, again suggesting preferential use of particular motifs or structures. At ~600 bp, the motif TAGTTAGTTKGT has been duplicated in tandem; a subsequent SDSA event appears to have created a further duplication of the core sequence TAGTTGGCTAGTT, flanking a filler sequence of 128 bp that contains a potentially bent region (Figure 4). Further downstream, the motif GGCAAGCA beginning at 1363 bp appears to represent the end point of two partial duplication events, one creating two approximate direct repeats of 79 and 74 bp and the second creating a short GAAAT microsatellite. Immediately downstream of the motif is a potentially bent region.

In all three of these cases, unusual structures appear to have been created or enhanced at preexisting unstable rearrangement sites. The predicted structures tend to have a regular spatial organization, with a periodicity of multiples of 207 bp (Figure 10). The most plausible explanation for these events is that nucleosome positioning, typical of heterochromatic repeats, is evolving on previously euchromatic DNA, which is unlikely to have contained strong packaging signals (Lowary and Widom 1997). In a number of cases rearrangements have optimized the periodicity. Thus, the direct repeats at 1188–1341 may have created an 828-bp spacing (4*n*) from the previous ~750 bp, while the complex rearrangements downstream of the *gag* fragment may have resulted in more gradual optimization, leading to identical 4*n* spacing. The D1100 family has also adopted the same periodicity; the monomer size, 1043 bp, represents a potential 5*n* unit (1035 bp). The 207-bp unit length is longer than that described for nucleosomes of rye A genome satellites or bulk euchromatin (Vershinin and Heslop-Harrison 1998), although a similar arrangement of 210-bp units anchored by nucleosomes positioned ~1 kb apart has been described recently in the chicken genome (Liu and Stein 1997). The discrepancy in unit size may reflect an unusual chromatin composition or conformation, and it emphasizes that the spacing seen has been generated during E3900 evolution rather than imported as a property of component fragments. Presumably the acquisition and reinforcement of a specific chromatin structure has favored the amplification of the current B-chromosome-specific families.

**Methylation and chromatin structure:** A particularly intriguing aspect of the curvature maxima is the contribution of CpG dinucleotides, which may underlie the unusual hypomethylation pattern seen. The predicted structure near 1908 bp is partly determined by the CpG dinucleotide contained within the *Hpa*II site at 1920 bp, which appears to be unmethylated (Figure 7). Although there is no current model for reliably predicting the effect of methylation on DNA structure, an approximation can be made by replacing the CpG motif with TpA. This single change decreased the predicted curvature of the region from ~12° per helical turn to ~10.5°. The specificity of this effect is demonstrated in Figure 10C, where all CpG dinucleotides have been replaced by TpA. Most peaks show little change either in position or magnitude, despite replacements of almost 10% of the E3900 sequence. Remarkably, there is very strong enhancement of the predicted peak adjacent to 1908 bp, which again can be shown to be due almost entirely to a single dinucleotide replacement. Replacement at both peaks increases the probability of a 5*n* (1035-bp) structure while reducing the probability of an out-of-phase structure at 2.5*n*. Another strong single replacement effect is seen at ~3540 bp, where methylation increases the probability of a 2*n* structure. Although these simulations must be treated with caution, it is tempting to speculate that the 1920-bp *Hpa*II site is protected from methylation as a consequence of structural constraints, either because the bases are not accessible to methylases or because the site acts as a switch between two chromatin conformations and its methylation is not compatible with the function of the B chromosome domain in leaf tissue. This possibility is discussed further below.

**Potential significance of amplified sequences:** Structural considerations appear to have been important in the more recent stages of evolution of the B-chromosome-specific families, but are unlikely to have been significant in the initial stages of amplification, when individual families would not have been sufficiently concentrated to have had large-scale effects on chromatin conformation. Although the large number of fragments apparently involved in E3900 evolution suggest that many random events have occurred, there are distinctive elements in both E3900 and D1100 that may have provided adaptive advantages during their initial amplification.

The most interesting fragment present in the B-chromosome-specific families is the E3900 sequence derived from a Ty3-gypsy retrotransposon. Initially amplified as an uninterrupted but chimeric *gag* reading frame, probably attached to its original LTR promoter, this sequence had the potential to express a peptide analogous to the mouse Fv1 product, a *gag*-like protein that confers resistance to leukemia retrovirus by interfering with a stage of infection that follows the virus' entry into the cell (Best *et al.* 1996). Similar immunity conferred by the ancestral E3900 *gag* peptide against endogenous retrotransposons, which constitute a major fraction of cereal genomes, could have reduced host genetic load inflicted by aggressive retrotransposon families, such as BARE-1. A less likely explanation is that the *gag* chimeric protein has acquired a novel host function, such as a role in chromosome activity analogous to the helicase-like genes carried by repetitive units in fungal subterminal domains (Sanchez-Alonso and Guzman 1998; Yamada *et al.* 1998). The most obvious role for such a

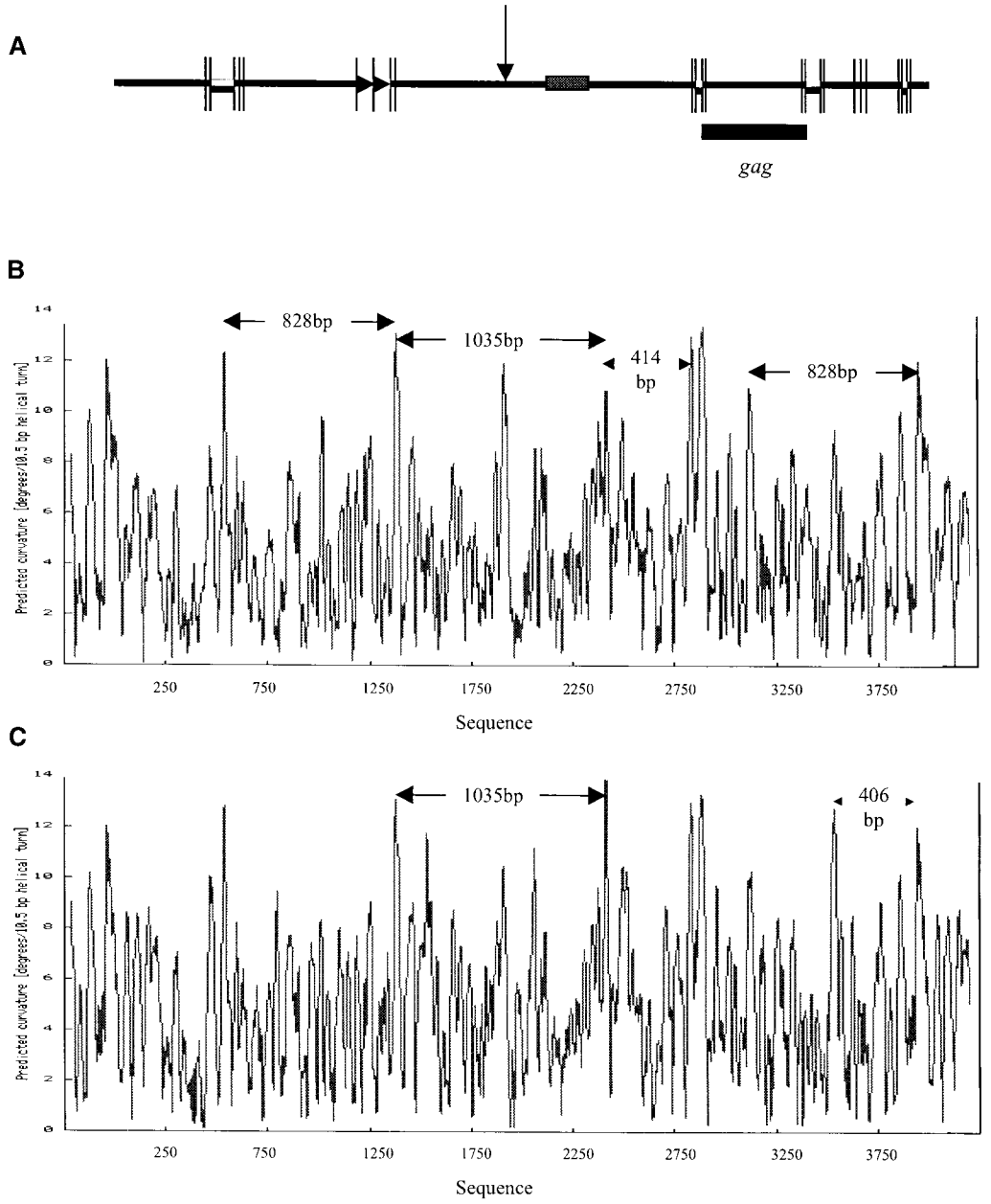


Figure 10.—Structural features correlate with E3900 rearrangement sites. (A) Location of rearrangement sites in E3900. Start and end points of duplicated sequences are indicated by vertical lines, and filler sequences are offset. The large tandem repeats are indicated by horizontal arrowheads. A region having TT dinucleotides at ~10-bp spacing, which may favor nucleosome binding, is indicated by a stippled box. The unmethylated HpaII site is indicated by an arrow, and the Ty3 *gag* reading frame is also indicated. (B) bend.it curvature prediction for the E3900 sequence. A total of 250 bp of tandemly arranged E3900 flanking sequences are included at each end of the sequence. Nucleotide scale corresponds to the diagram in A, and numbering begins at position 1 in Figure 4. Spacings corresponding to multiples of 207 bp are indicated (414 bp = 2*n*, 828 bp = 4*n*, and 1035 bp = 5*n*). (C) Same as B, but CpG dinucleotides in E3900 have been replaced throughout by TpA.

protein would be involvement in the nondisjunction mechanism that drives the increase in B chromosome numbers over generations and that requires a *trans*-acting factor from the B-chromosome-specific domain (Lima-de-Faria 1962). The probable origin of the *gag* gene in a centromere-specific retrotransposon element lends some support to this possibility, as nondisjunction occurs at paracentromeric regions that also may be targets for the ancestral *crwydryn* retrotransposition; an interaction by a *gag* motif could be the basis for localization of both processes.

The D1100 family, in contrast to E3900, is likely to have played a role in chromosome organization since its origin. The elements are smaller than E3900 and show no evidence of a genic origin or of such high levels of rearrangement. Instead, a MITE fragment is present, which appears to have been inserted into an AT-rich sequence, an organization typical of cereal MARs (Avramova *et al.* 1998). MARs are particularly abundant in heterochromatin (Strausbaugh and Williams 1996; Craig *et al.* 1997) and the D1100 sequence, whether singly or in arrays, may still act in this capacity, balancing the effect of the E3900 core in the domain overall and leading to the greater degree of condensation shown specifically by D1100-rich subdomains. The presence of a single arm of a mobile element is also reminiscent of the array of *Bari* transposons in Drosophila (Caizzi *et al.* 1993) and of the ancestral transposon model proposed for human α-satellite arrays (Kipling and Warburton 1997). In both cases, although normal transposition is blocked because of damage or divergence of one transposon end, residual transposase activity may

generate recombinogenic nicks that drive array expansion or homogenization. Similar behavior by the D1100 element could have provided it with an advantage in its initial colonization of the B-chromosome-specific domain. It is also possible that head-to-head rearrangements of D1100 units could reconstitute transposable units. Megatransposons, created by head-to-head arrangements of repetitive arrays, have been proposed to account for the instability of maize knob domains (Ananiev *et al.* 1998a); the apparent duplication involving the D1100-rich subdomains (Figure 1) may have occurred by this means.

**The B-chromosome-specific domain represents a transient stage in satellite evolution:** Satellite families have been studied extensively in both animals and plants, and the evolution of tandemly repeated sequences after their organization into arrays has been well documented. The origin of most satellites, however, is unknown. The data reported here are unusual in describing the creation of satellite repeats *de novo* from complex euchromatic sequences. It is still not clear why the evolution of new families on the B chromosome should be favored over the apparently simpler processes of capture and amplification or modification of existing repeats from the A chromosomes, particularly as translocation of A genome domains to the B chromosome has been observed (Wilkes *et al.* 1995). We assume that there is a requirement that the B chromosome subtelomeric domain remain distinct, possibly to ensure correct segregation or to avoid suppression of the postmeiotic drive mechanism.

It is of some interest to consider if the contemporary B-chromosome-specific domain represents a stable evolutionary stage. The unusual discontinuous appearance of the most terminal subdomain at meiotic prophase and the variable degree of condensation of the domain in vegetative cells indicate that the region does not behave consistently. An obvious explanation is that this reflects the conflicting properties of the two major sequence families, D1100 and E3900, with conformation effects similar to those seen in the occurrence of position effect variegation (PEV; Wallrath and Elgin 1995). For a standard A chromosome, it may be expected that predictable behavior would be favored, and that domain(s) containing a single family would be selected rapidly. However, it may be to the long-term advantage of the B chromosome to maintain this instability. In rye, the B chromosome possesses a postmeiotic drive mechanism based on nondisjunction, which leads to preferential segregation of the chromatids of B chromosomes to the generative nuclei (Hasegawa 1934). The resulting increase in copy number in progeny is balanced by some loss of B chromosomes that fail to pair during meiosis. The standard B chromosome has been remarkably successful in maintaining its presence in rye populations across their geographical and environmental range despite the lack of any strong phenotypic benefit associated with it, and the balance between B chromosome loss and amplification might be expected to be capable of adjustment in response to different conditions to prevent either extinction or unacceptably high numbers of B chromosomes in the host. In agreement with this expectation, large variations in transmission frequency have been described between and within populations (Romera *et al.* 1991). Surprisingly, nondisjunction occurs at a constant high frequency and the variation in transmission rates appears to be generated by differences in pairing efficiency (Ortiz *et al.* 1996). This variation has been recently inferred to be under genetic control by the B chromosome itself (Puertas *et al.* 1998); we suggest that the instability of the B chromosome subterminal domain is likely to play a key role in maintaining this variation.

By analogy with the demonstrated role of heterochromatic domains in Drosophila (Dernburg *et al.* 1996), the subtelomeric domain may enhance B chromosome pairing potential when it adopts a heterochromatic conformation. B chromosomes carrying a complex arrangement of D1100 and E3900 families, and hence susceptible to PEV, may then give rise to copies that adopt either a condensed or extended conformation at some stage critical for chromosome pairing, in turn giving rise to a range of B chromosome copy numbers in progeny. Small changes in the relative amounts of the families, generated by chromosomal rearrangements, ectopic recombination, or gene conversion, may shift the balance between condensed and extended forms, creating novel transmission frequencies that may allow adaptation to particular conditions. However, major changes are likely to result in a chromosome capable of adopting only a single conformation that will rapidly condemn either itself or its host to extinction. The sensitivity of the curvature predictions to the position of CpG dinucleotides suggests that the balance may be maintained even at the epigenetic level, via methylation, consistent with the demonstrated sensitivity to methylation inhibitors of B chromosome segregation in vegetative tissue (Neves *et al.* 1992) and providing a possible link with the maternal imprinting effect on B chromosome transmission (Puertas *et al.* 1990). The constraints imposed by such a sensitive system are unlikely to be met by a simple homogeneous array, which would be prone both to large stochastic changes and to control by the host via specific components. If this model is correct, then the B chromosome domain represents a new and sophisticated example of heterochromatin function.

This work was supported by Biotechnology and Biological Sciences Research Council grant PO1643 to G.J., R.N.J., and J.W.F.

## LITERATURE CITED

Ananiev, E. V., R. L. Phillips and H. W. Rines, 1998a Complex structure of knob DNA on maize chromosome 9: retrotransposon invasion into heterochromatin. Genetics **149:** 2025–2037.

Ananiev, E. V., H. W. Rines and R. L. Phillips, 1998b Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc. Natl. Acad. Sci. USA **95:** 13071–13078.

Avramova, Z., A. Tikhonov, M. Chen and J. L. Bennetzen, 1998 Matrix attachment regions and structural colinearity in the genomes of two grass species. Nucleic Acids Res. **26:** 761–767.

Bennetzen, J. L., and E. A. Kellog, 1997 Do plants have a one-way ticket to genomic obesity? Plant Cell **9:** 1509–1514.

Bennetzen, J. L., P. Sanmiguel, M. S. Chen, A. Tikhonov, M. Francki *et al.*, 1998 Grass genomes. Proc. Natl. Acad. Sci. USA **95:** 1975–1978.

Best, B., P. le Tissier, G. Towers and J. P. Stoye, 1996 Positional cloning of the mouse retrovirus restriction gene *Fv1*. Nature **382:** 826–829.

Blunden, R., T. J. Wilkes, J. W. Forster, M. M. Jimenez, M. J. Sandery *et al.*, 1993 Identification of the E3900 family, a second family of rye B chromosome specific repeated sequences. Genome **36:** 706–712.

Brukner, I. R., R. Sanchez, D. Suck and S. Pongor, 1995 Trinucleotide models for DNA bending propensity: comparison of models based on DnaseI digestion and nucleosome packaging data. J. Biomol. Struct. Dyn. **13:** 309–317.

Caizzi, R., C. Caggese and S. Pimpinelli, 1993 BarI-1, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. Genetics **133:** 335–345.

Craig, J., S. Boyle, P. Perry and W. Bickmore, 1997 Scaffold attachments within the human genome. J. Cell Sci. **110:** 2673–2682.

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

Dernburg, A. F., J. W. Sedat and R. S. Hawley, 1996 Direct evidence of a role for heterochromatin in meiotic chromosome segregation. Cell **86:** 135–146.

Dong, F. G., J. T. Miller, S. A. Jackson, G. L. Wang, P. C. Ronald *et al.*, 1998 Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc. Natl. Acad. Sci. USA **95:** 8135–8140.

Fitzgerald, D. J., G. L. Dryden, E. C. Bronson, J. S. Williams and J. N. Anderson, 1994 Conserved patterns of bending in satellite and nucleosome positioning DNA. J. Biol. Chem. **269:** 21303–21314.

Gabrielian, A., and S. Pongor, 1996 Correlation of intrinsic DNA curvature with DNA property periodicity. FEBS Lett. **393:** 65–68.

Goodsell, D. S., and R. E. Dickerson, 1994 Bending and curvature calculations in B-DNA. Nucleic Acids Res. **22:** 5497–5503.

Gorbonuva, V., and A. A. Levy, 1997 Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Res. **25:** 4650–4657.

Grebenstein, B., O. Grebenstein, W. Sauer and V. Hemleben, 1995 Characterization of a highly repeated DNA component of perennial oats (*Helictotrichon, Poaceae*) with sequence similarity to a A-genome-specific satellite DNA of rice (*Oryza*) Theor. Appl. Genet. **90:** 1101–1105.

Gustafson, J. P., A. J. Lukaszewski and M. D. Bennett, 1983 Somatic deletion and redistribution of telomeric heterochromatin in the genus *Secale* and *Triticale*. Chromosoma **88:** 293–298.

Hasegawa, N., 1934 A cytological study on 8-chromosome rye. Cytologia **6:** 68–77.

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

Houben, A., R. G. Kynast, U. Heim, H. Hermann, R. N. Jones *et al.*, 1996 Molecular cytogenetic characterisation of the terminal heterochromatic segment of the B-chromosome of rye (*Secale cereale*). Chromosoma **105:** 97–103.

Jones, J. D. G., and R. B. Flavell, 1982 The structure, amount and chromosomal localization of defined repeated DNA sequences in species of the genus *Secale*. Chromosoma **86:** 613–641.

Jones, R. N., and M. J. Puertas, 1993 The B-chromosomes of rye, pp. 81–112 in *Frontiers in Plant Science Research*, edited by K. K. Dhir and T. S. Sareen. Bhagwati Enterprises, Delhi.

Karp, A., P. G. Owen, S. H. Steele, P. J. Bebeli and P. J. Kaltsikes, 1992 Variation in telomeric heterochromatin in somaclones of rye. Genome **35:** 590–593.

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

Lanzer, M., S. P. Wertheimer, D. Debruin and J. V. Ravetch, 1994 Chromatin structure determines the sites of chromosome breakages in *Plasmodium falciparum*. Nucleic Acids Res. **22:** 3099–3103.

Leitch, I. J., and J. S. Heslop-Harrison, 1992 Detection of digoxygenin labelled DNA probes hybridized to plant chromosomes *in situ*, pp. 18–25 in *Methods in Molecular Biology: Protocols for Nucleic Acid Analysis by Non-Radioactive Techniques*, edited by P. G. Isaac. Humana Press, Clifton, NJ.

Lima-de-Faria, A., 1962 Genetic interaction in rye expressed at the chromosome phenotype. Genetics **47:** 1455–1462.

Liu, K., and A. Stein, 1997 DNA sequence encodes information for nucleosome array formation. J. Mol. Biol. **270:** 559–573.

Lowary, P. T., and J. Widom, 1997 Nucleosome packaging and nucleosome positioning of genomic DNA. Proc. Natl. Acad. Sci. USA **94:** 1183–1188.

Mollegaard, N. E., C. Bailly, M. J. Waring and P. E. Nielsen, 1997 Effects of diaminopurine and inosine substitutions on A-tract induced DNA curvature. Importance of the 3′-A-tract junction. Nucleic Acids Res. **25:** 3497–3502.

Morais-Cecilio, L., M. Delgado, R. N. Jones and W. Viegas, 1997 Interphase arrangement of rye B chromosomes in rye and wheat. Chrom. Res. **5:** 177–181.

Moschetti, R., R. Caizzi and S. Pimpinelli, 1996 Segregation distortion in *Drosophila melanogaster*: genomic organization of Responder sequences. Genetics **144:** 1665–1671.

Muller, H. P., and H. E. Varmus, 1994 DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. **13:** 4704–4714.

Munteanu, M. G., K. Vlahovicek, S. Parthasaraty, I. Simon and S. Pongor, 1998 Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. Trends Biochem. Sci. **23:** 341–346.

Nagaki, K., H. Tsujimoto and T. Sasakuma, 1998 Dynamics of tandem repetitive Afa-family sequences in triticeae, wheat-related species. J. Mol. Evol. **47:** 183–189.

Nassif, N., J. Penney, S. Pal, W. R. Engels and G. B. Gloor, 1994 Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol. Cell. Biol. **14:** 1613–1625.

Neves, N., A. Barao, A. Castilho, M. Silva, L. Morais *et al.*, 1992 Influence of DNA methylation on rye B-chromosome nondisjunction. Genome **35:** 650–652.

Ortiz, M., M. J. Puertas, M. M. Jimenez, F. Romera and R. N. Jones, 1996 B-chromosomes in inbred lines of rye (*Secale cereale* L.) 2. Effects on metaphase I and first pollen mitosis. Genetica **97:** 65–72.

Ostashevsky, J. Y., 1996 Centromeric locations in karyotypes: a rule derived from the theory of branched polymers. J. Theor. Biol. **181:** 293–298.

Pan, W. H., A. Houben and R. Schlegel, 1992 Highly effective cell synchronization in plant roots by hydroxyurea and amiprophos-methyl or colchicine. Genome **36:** 387–390.

Panstruga, R., R. Buschges, P. Riffanelli and P. Schulze-Lefert, 1998 A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Res. **26:** 1056–1062.

Pasero, P., N. Sjakste, C. Blettry, C. Got and M. Marilley, 1993 Long-range organization and sequence-directed curvature of *Xenopus laevis* satellite 1 DNA. Nucleic Acids Res. **21:** 4703–4710.

Presting, G. G., L. Malysheva, J. Fuchs and I. Z. Schubert, 1998 A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. **16:** 721–728.

Pruss, D., F. D. Bushman and A. P. Wolffe, 1994 Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl. Acad. Sci. USA **91:** 5913–5917.

Puertas, M. J., M. M. Jimenez, F. Romera, J. M. Vega and M. Diez, 1990 Maternal imprinting effect on B-chromosome transmission in rye. Heredity **64:** 197–204.

Puertas, M. J., M. Gonzalez-Sanchez, S. Manzanero, F. Romera and M. M. Jimenez, 1998 Genetic control of the rate of transmission of rye B chromosomes. IV. Localization of the genes controlling B transmission rate. Heredity **80:** 209–213.

Raeder, U., and P. Broda, 1985 Rapid preparation of DNA from filamentous fungi. Lett. Appl. Microbiol. **1:** 17–20.

Romera, F., M. M. Jimenez and M. J. Puertas, 1991 Factors control-

ling the dynamics of the B-chromosome polymorphism in Korean rye. Heredity **67:** 189–195.

Rossi, M. S., C. G. Pesce, O. A. Reig, A. R. Kornblihtt and J. Zorzopulos, 1993 Retroviral-like features in the monomer of the major satellite DNA from the South American rodents of the genus *Ctenomys*. DNA Seq. **3:** 379–381.

Rubin, E., and A. A. Levy, 1997 Abortive gap repair: underlying mechanism for Ds element formation. Mol. Cell. Biol. **17:** 6294–6302.

Salomon, S., and H. Puchta, 1998 Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J. **17:** 6086–6095.

Salser, W., S. Bowen, D. Browne, F. El-Adli, N. Federoff *et al.*, 1976 Investigation of the organization of mammalian chromosomes at the DNA sequence level. Fed. Proc. **35:** 23–35.

Sanchez-Alonso, P., and P. Guzman, 1998 Organization of chromosome ends in *Ustilago maydis*. RecQ-like helicase motifs at telomeric regions. Genetics **148:** 1043–1054.

Sandery, M. J., J. W. Forster, R. Blunden and R. N. Jones, 1990 Identification of a family of repeated sequences on the rye B chromosome. Genome **33:** 908–913.

SanMiguel, P., A. Tikhonov, Y. K. Jin, N. Motchoulskaia, D. Zakharov *et al.*, 1996 Nested retrotransposons in the intergenic regions of the maize genome. Science **274:** 765–768.

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

Strausbaugh, L. D., and S. M. Williams, 1996 High density of an SAR-associated motif differentiates heterochromatin from euchromatin. J. Theor. Biol. **183:** 159–167.

Taniguchi, M., and T. Sugiyama, 1996 Isolation, characterisation and expression of cDNA clones encoding a mitochondrial malate translocator from *Panicum miliaceum* L. Plant Mol. Biol. **30:** 51–64.

Tatout, C., L. Lavie and J. M. Deragon, 1998 Similar target site selection occurs in integration of plant and mammalian retroposons. J. Mol. Evol. **47:** 463–470.

Tenzen, T., Y. Matsuda, H. Ohtsubo and E. Ohtsubo, 1994 Transposition of Tnr1 in rice genomes to 5′-putapy-3′ sites, duplicating the Ta sequence. Mol. Gen. Genet. **245:** 441–448.

Tsujimoto, H., and K. Niwa, 1992 DNA-structure of the B-chromosome of rye revealed by in situ hybridization using repetitive sequences. Jap. J. Genet. **67:** 233–241.

Vershinin, A. V., and J. S. Heslop-Harrison, 1998 Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. Plant Mol. Biol. **36:** 149–161.

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

Vershinin, A. V., E. G. Alkhimova and J. S. Heslop-Harrison, 1996 Molecular diversification of tandemly organized DNA sequences and heterochromatic chromosome regions in some Triticeae species. Chrom. Res. **4:** 517–525.

Wallrath, L. L., and S. C. Elgin, 1995 Position effect variegation in *Drosophila* is associated with an altered chromatin structure. Genes Dev. **9:** 1263–1277.

Wessler, S., A. Tarpley, M. Purugganan, M. Spell and R. Okagaki, 1990 Filler DNA is associated with spontaneous deletions in maize. Proc. Natl. Acad. Sci. USA **87:** 8731–8735.

Wilkes, T. M., M. G. Francki, P. Langridge, A. Karp, R. N. Jones *et al.*, 1995 Analysis of rye B-chromosome structure using fluorescence *in-situ* hybridization (FISH). Chrom. Res. **3:** 466–472.

Yamada, M., N. Hayatsu, A. Matsuura and F. Ishikawa, 1998 Y′-Help1, a DNA helicase encoded by the yeast subtelomeric Y′ element, is induced in survivors defective for telomerase. J. Biol. Chem. **273:** 33360–33366.

Young, M. A., J. Srinivasan, I. Goljer, S. Kumar, D. L. Beveridge and P. H. Bolton, 1995 Structure determination and analysis of local bending in an A-tract DNA duplex: comparison of results from crystallography, nuclear magnetic resonance, and molecular dynamics simulation on d(CGCAAAAATGCG). Methods Enzymol. **261:** 121–144.

Yu, H. G., E. N. Hiatt, A. Chan, M. Sweeney and R. K. Dawe, 1997 Neocentromere-mediated chromosome movement in maize. J. Cell Biol. **139:** 831–840.

Zhong, X-B., P. F. Fransz, J. Wennekes van Eden, P. Zabel, A. van Kammen *et al.*, 1996 High resolution mapping on pachytene chromosomes and extended DNA fibres by fluorescence *in situ* hybridisation. Plant Mol. Biol. Rep. **14:** 232–242.

Communicating editor: J. A. Birchler