Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 8761–8765, August 1996
Plant Biology

The physical and genomic organization of microsatellites in sugar beet

(*Beta vulgaris*/simple sequence repeats/genome evolution/repetitive DNA/*in situ* hybridization/chromosomes)

T. SCHMIDT AND J. S. HESLOP-HARRISON

Karyobiology Group, Department of Cell Biology, John Innes Centre, Colney Lane, Norwich, NR4 7UH, United Kingdom

Communicated by J. Heslop-Harrison, Herefordshire, United Kingdom, May 1, 1996 (received for review February 2, 1996)

ABSTRACT **Microsatellites, tandem arrays of short (2–5 bp) nucleotide motifs, are present in high numbers in most eukaryotic genomes. We have characterized the physical distribution of microsatellites on chromosomes of sugar beet (*Beta vulgaris* L.). Each microsatellite sequence shows a characteristic genomic distribution and motif-dependent dispersion, with site-specific amplification on one to seven pairs of centromeres or intercalary chromosomal regions and weaker, dispersed hybridization along chromosomes. Exclusion of some microsatellites from 18S–5.8S–25S rRNA gene sites, centromeres, and intercalary sites was observed. In-gel and *in situ* hybridization patterns are correlated, with highly repeated restriction fragments indicating major centromeric sites of microsatellite arrays. The results have implications for genome evolution and the suitability of particular microsatellite markers for genetic mapping and genome analysis.**

Runs of repetitions of short sequence motifs 2–5 bp long, described as microsatellites or simple sequence repeats, are probably ubiquitous elements of eukaryotic genomes (1–3). They provide highly informative and polymorphic markers for genetics (4–6) or plant, fungal, and animal fingerprinting (7, 8). Genetic mapping using microsatellites as markers involves amplification of repeat arrays using PCR with primers flanking the arrays. Primers are often chosen based on database searches or sequence data from cloned arrays (9–12), which give selective data about genomic distribution. However, little is known about the real chromosomal organization and physical localization of microsatellite motifs within plant genomes. The only microsatellite repeat mapped physically in plants, the polypurine motif $(GAA)_7$, has been correlated with the positions of C-bands in barley (13). Some physical mapping in fish and primates (8, 14) using *in situ* hybridization shows clustering of microsatellites on some chromosomes. Recent data in mouse show that mapped microsatellites, mostly $(CA)_n$ repeats mapped by PCR polymorphisms, are distributed among autosomes in proportion to chromosome length, while the X chromosome shows a clear deficit of the microsatellites (15). Within chromosomes, the frequency of both large and small clusters with respect to meiotic crossovers slightly exceeded expectation.

Sugar beet (*Beta vulgaris* L.; 2n = 2x = 18) is a valuable model species for investigating the large scale organization of the nuclear genome because (*i*) the genome is relatively small with 758 Mbp (16), (*ii*) fluorescent *in situ* hybridization can accurately locate sequences along the metaphase chromosomes and within interphase nuclei (17), (*iii*) major classes of the repetitive DNA have been characterized including both satellite and retrotransposon sequences (18–22), and (*iv*) microsatellites are known to be highly abundant (23).

Here, we aimed to characterize the genomic distribution of microsatellite sequences in sugar beet. We used seven simple sequences representing a range of nucleotide motifs, many used for genetic mapping in plants (9–11), and chosen so that the tetranucleotides were composites of the dinucleotides. We examined their chromosomal distribution patterns by *in situ* hybridization and their in-gel hybridization patterns to size-separated restriction enzyme digests of genomic DNA, allowing comparison of the physical and molecular organization of the sequences.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Plant Material, DNA Extraction, and in-Gel Hybridization. Genomic DNA of *B. vulgaris* (beet, cultivar "Rosamona") was extracted from young leaf tissue (19). For in-gel hybridization, genomic DNA was digested with *Alu*I, *Hae*III, *Hin*fI, and *Rsa*I and separated in 1% agarose gels. The same restriction enzyme digests were used for loading four gels. Gels were dried in a slab dryer, pretreated, and hybridized with oligonucleotides end-labeled with [γ-^{32}P]dATP as described (24). The oligonucleotides used were $(GGAT)_4$, $(GATA)_4$, $(GACA)_4$, $(CA)_8$, $(GA)_{12}$, $(TA)_{10}$, and $(CAC)_5$. Hybridization and stringent washes were carried out at the respective T_m −5°C according to Thein and Wallace (25) such that probes complementary to target sequences of tetramers [(GGAT), (GATA), (GACA)], pentamers (CAC), octamers [(CA), (GA)], and decamers (TA), respectively, remain stably hybridized.

Fluorescent *in Situ* Hybridization. Metaphase and prometaphase chromosomes of beet (cultivar "Rosamona") were prepared from primary root meristems and pretreated for *in situ* hybridization (17). The same oligonucleotides as above were 3′ end-labeled with biotin-16-dUTP (detected with streptavidin-Cy3) or digoxigenin-11-dUTP (detected with anti-digoxigenin–fluorescein isothiocyanate) by terminal transferase according to the instructions of the manufacturer (Boehringer Mannheim). Chromosomes were denatured at 70°C for 8 min in 30 μl of hybridization solution containing 1–2 pmol of probe. After hybridization overnight, essentially following ref. 26, slides were washed, with the highest stringency wash being at T_m −5°C. Probe detection, counterstaining, chromosome examination, image acquisition and processing was as described (22). Results are based on the analysis of three to six metaphases for each probe.

RESULTS

Metaphase chromosomes (stained with 4′,6-diamidino-2-phenylindole; sub-terminal 18S–5.8S–25S rRNA genes are indicated with an arrow on the satellited chromosomes) showed strong and characteristic patterns of *in situ* hybridization (signal corresponding to Cy3 or fluorescein isothiocyanate fluorescence, respectively) with six of the seven microsatellites (Fig. 1). All seven microsatellites showed different in-gel hybridization patterns (Fig. 2).

Strong hybridization of $(GGAT)_4$ bands was detected on the proximal region of the short arm of two chromosome pairs, one of which carries the distal 18S–5.8S–25S rRNA gene locus (Fig.

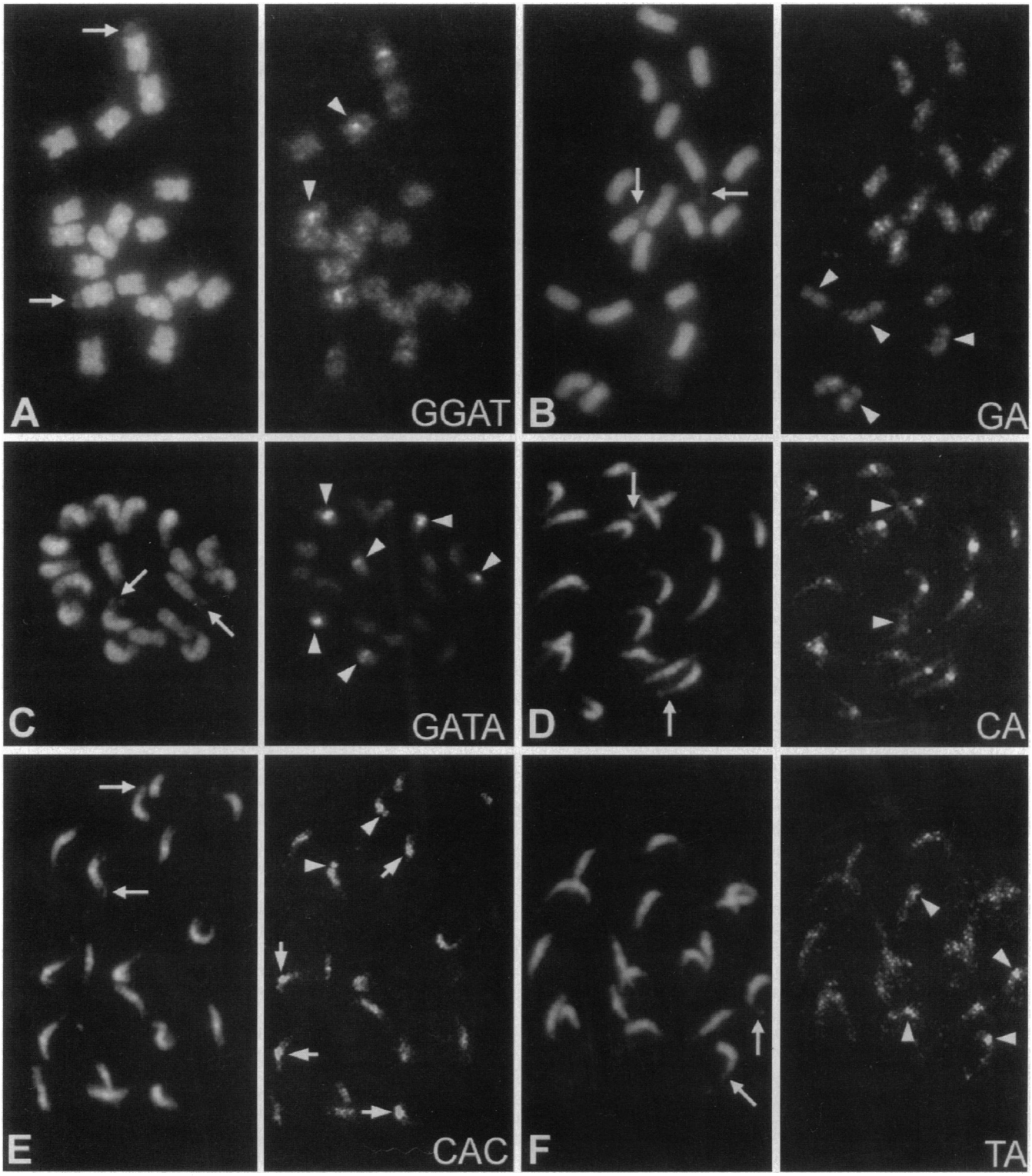

FIG. 1. Micrographs showing the distribution of microsatellites on metaphase (*A* and *B*; ×2000) and prometaphase chromosomes (*C–F*; ×1500) of *B. vulgaris*. Left panels show chromosomes after staining with 4′,6-diamidino-2-phenylindole; the subterminal 18S–5.8S–25S rRNA gene cluster is indicated by arrows on the satellite chromosome pair. Right panels show the same chromosome preparation after fluorescent *in situ* hybridization with oligonucleotides complementary to microsatellites. (*A*) $(GGAT)_4$ repeats show dispersed hybridization and four major hybridization sites. One pair of strong signals is proximal on the satellited arm, while the other is proximal on another short chromosome arm (arrowheads). (*B*) Dispersed *in situ* hybridization of $(GA)_{12}$ is seen, with some intercalary amplified sites and gaps at many centromeres (e.g., arrowheads) and 18S–5.8S–25S rRNA genes. (*C*) Six major centromeric sites (arrowheads) are detected on the late prometaphase after *in situ* hybridization with $(GATA)_4$. (*D*) Amplification of $(CA)_8$ is detected near the centromere of all prometaphase chromosomes but very weak on one pair of chromosomes (arrowheads). (*E*) *In situ* hybridization signal of $(CAC)_5$ is detected in the proximal regions of all prometaphase chromosomes. Hybridization reveals a negative band (arrowheads) on the long arms of the satellited chromosome pair, while short arrows point to six major centromeric sites. (*F*) Dispersed signal is seen after fluorescence *in situ* hybridization of $(TA)_{10}$ on all prometaphase chromosomes, with intercalary amplification sites on the satellited chromosomes and another chromosome pair (arrowheads).

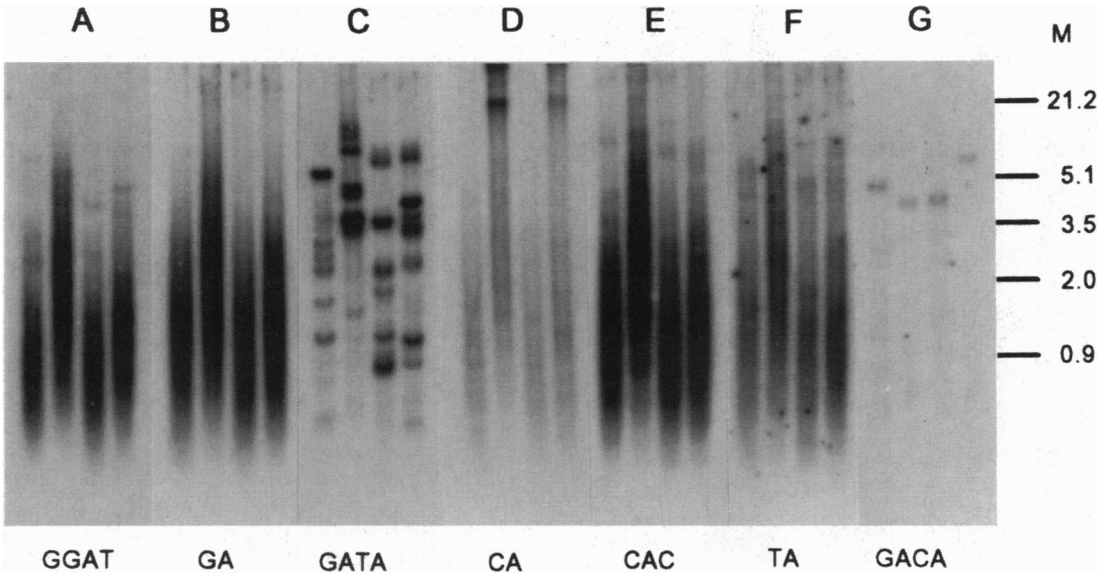

FIG. 2. (*A–G*) Genomic distribution of microsatellites (motif given under each panel) in *B. vulgaris* shown by in-gel hybridization to genomic DNA digested with *Alu*I, *Hae*III, *Hin*fI, and *Rsa*I (left to right in each panel). Size markers (M) are given in kb.

1*A* arrows). All chromosomes showed dispersed hybridization signals extending to the physical ends, with the sequence reduced at the centromeres of many chromosomes and excluded from the rRNA gene locus and adjacent sequences. In-gel hybridization showed a smear of hybridization, with repetitive fragments between 4.0 and 7.5 kb in all four restriction enzyme digests (Fig. 2*A*).

All chromosomes showed dispersed signal after *in situ* hybridization of $(GA)_{12}$ with presence at low density at 18S–5.8S–25S rRNA gene sites (arrows). The sequence was sharply excluded from many centromeres, and no major centromeric hybridization sites were found (Fig. 1*B*). In-gel hybridization showed a continuous smear with no amplified fragments (Fig. 2*B*).

In situ hybridization of $(GATA)_4$ revealed six centromeric and adjacent regions showing amplification of (GATA) repeats (Fig. 1*C*). Weak hybridization was found along all chromosome arms. Regions with a low density of (GATA) arrays were detectable as gaps on three chromosome pairs. The in-gel hybridization showed multiple repetitive fragments over a wide range of molecular weight in genomic DNA cleaved with *Alu*I, *Hae*III, *Hin*fI, and *Rsa*I (Fig. 2*C*).

The microsatellite motif $(CA)_8$ was strongly amplified at 16 centromeric regions (Fig. 1*D*), while one chromosome pair showed less amplification. Interphase hybridization showed that most, but not all, 4′,6-diamidino-2-phenylindole-positive heterochromatin clusters co-located with the (CA) microsatellite (data not shown). In-gel hybridization showed numerous fragments, including high molecular weight fragments. *Hae*III (Fig. 2*D* center left) and *Rsa*I (Fig. 2*D* right) digested DNA showed hybridization to fragments larger than 21.2 kb, indicating the absence of the restriction sites and the existence of large arrays of simple structure.

Three chromosome pairs had amplification of (CAC) microsatellite repeats near the centromere (Fig. 1*C*). Chromosomes showed characteristic dispersed signals, and hybridization was excluded from the terminal regions. The absence or reduced frequency of (CAC) repeats on a chromosome segment on the opposite arm to the 18S–5.8S–25S rRNA genes gave a conspicuous gap. In-gel hybridization showed a smear in all four size-separated restriction enzyme digests with multiple fragments larger than 5 kb (Fig. 2*E*).

Strong hybridization of $(TA)_{10}$ was detected at an intercalary position on an arm of the chromosome pair with the rRNA genes (Fig. 1*F*). Some chromosome pairs showed reduced signal density giving negative bands at the centromeres, and all of the chromosome arms showed weak dispersed signal. In-gel hybridization of the oligonucleotide probe showed some amplified fragments on a background smear after extended exposure (Fig. 2*F*).

In-gel hybridization with $(GACA)_4$ after extended exposure time revealed a single fragment (Fig. 2*G*). No hybridization signal was detected *in situ*.

The diagram in Fig. 3 summarizes the distribution of the microsatellites studied on the satellite chromosome. This chromosome, which was always identified unequivocally, showed the most discrete distribution pattern of sequences complementary to simple DNA motifs.

DISCUSSION

Taken together, the complementary *in situ* (Fig. 1 *A–F*) and in-gel hybridization results (Fig. 2*A–G*) show that the genomic organization of different microsatellite sequences varies widely, with implications for amplification and dispersion mechanisms and hence evolution and their utility for mapping. The hybridization pattern depended on the sequence of the oligomer used, although a typical picture of microsatellite distribution in the sugar beet genome emerges: hybridization with major amplification sites near the centromeres of one to seven chromosome pairs, weaker dispersed hybridization along chromosome arms out to the telomeres, and exclusion from particular centromeric (e.g., Fig. 1*B*), intercalary chromosome regions (e.g., Fig. 1*E*), and 18S–5.8S–25S rRNA gene sites (e.g., Fig. 1*A*). Microsatellite motifs were chosen to have 75% homology to at least one other array [e.g. $(GACA)_n$, $(GATA)_n$ and $(GA)_{2n}$, where n is an integer], and the contrasting patterns along chromosomes show that the hybridization conditions used are able to discriminate each motif.

The presence, absence and strength of repetitive microsatellite fragments detected by in-gel hybridization tended to correlate with the hybridization pattern on chromosomes; amplification at centromeres was accompanied by major repetitive fragments in gels (Figs. 1 and 2). In tomato, genetic mapping shows both $(GATA)_4$ and $(GACA)_4$ cluster in the same chromosomal regions, likely to be the centromeric regions (27, 28). $(GATA)_4$ is a frequently used probe for the detection of multilocus microsatellites and is informative for plant genome analysis (7, 23). Genetic mapping using polymorphic PCR products from primers flanking short microsat-

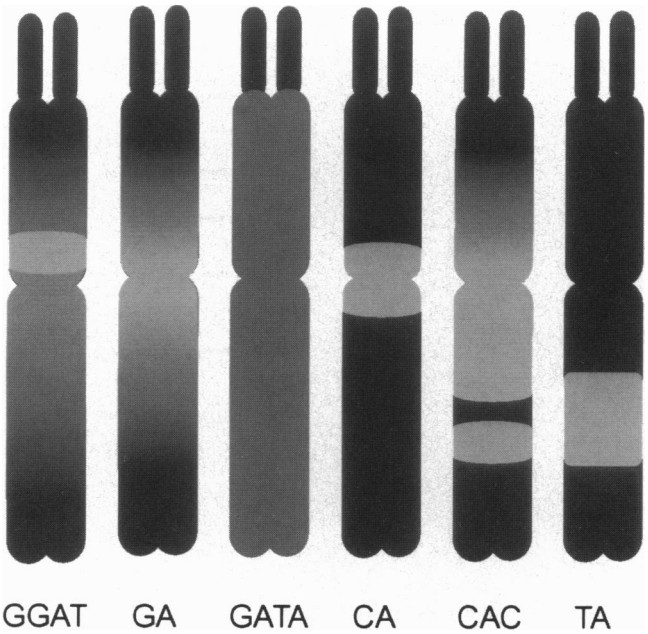

FIG. 3. Schematic representation of microsatellite amplification sites on the sugar beet chromosome carrying the subterminal 18S–5.8S–25S rRNA genes (top).

ellite arrays is likely to detect some loci represented by individual fluorescent sites making up the weak *in situ* hybridization signal along chromosome arms (Fig. 1), supported by evidence that microsatellite markers give a relatively uniform coverage of the genome (9–11). Mapped microsatellite loci are often compounds of more than one simple sequence motif; the dispersed *in situ* hybridization observed is consistent with the co-localization of several motifs. We suggest that the major sites detected by *in situ* hybridization represent microsatellite arrays that are often too long to amplify using flanking primers and hence are not mapped genetically by this method. The hybridization of all 49 possible mono-, di-, tri-, and tetramer microsatellites to *Arabidopsis thaliana* genomic digests has been examined (29). Although 40 microsatellites showed no hybridization, nine showed similar hybridization to the in-gel patterns reported here, mostly with distinct fragments and possibly explained by the low amount of nuclear DNA in *A. thaliana*.

The only microsatellite previously localized in plants, $(GAA)_7$, shows multiple sites around the centromeres of most barley chromosomes, and the pattern is similar to that obtained by C-banding except that the terminal and nucleolar constriction bands were missing (13); no diffuse hybridization was reported. In beet, C-bands are mainly restricted to the centromere regions (30), frequently sites of strong microsatellite hybridization. It has been suggested elsewhere that renaturation properties of microsatellite sequences could relate to C-banding (31), as was also found with the $(CAC)_n$ microsatellite in human (8), most likely to be caused by distorted base pair ratio and helical conformation of microsatellite stretches.

In some mammals, $(GACA)_n$ sequences have been associated with the 18S–5.8S–25S rRNA gene loci (8), contrasting with sugar beet where all the microsatellites examined are largely excluded from these loci. In fish and mammals, *in situ* hybridization results showed that some simple repetitive DNA sequences are located in different chromosomal regions (8) that are constrained considerably during evolution, often on the sex chromosomes. Although there are remarkable differences in the abundance of microsatellite sequence motifs between plants and vertebrates (32), strong amplification of microsatellites at specific chromosomal locations may be a general feature of eukaryotic genomes.

The microsatellite sequence (CA) has the strongest amplification at the centromeres of all sequences analyzed here and also shows hybridization to limiting mobility DNA fragments in the *Hae*III and *Rsa*I digests. *Hae*III is sensitive to cytosine 5-methylation, and higher molecular weight fragments are detected with most microsatellites (Fig. 2). *Rsa*I, with no known cytosine methylation sensitivity, shows limiting mobility fragments (>21.2 kb) only after hybridization with $(CA)_8$ (Fig. 2*D*). This implies that the large arrays of (CA) occur in genomic regions differing from all the other microsatellites used, correlating with its presence and amplification at seven pairs of centromeres. During sequencing of tandemly repeated DNA from sugar beet, we have found a sequence of 327 bp, which includes the microsatellite $(AC)_8$ (18). The tandemly repeated sequence is located around all centromeres in blocks more than 80 kb long, and homology in the microsatellite domain is probably responsible for some of the hybridization signal (Fig. 1*E*). Because of the genomic distribution and high copy number, we suggest the motif (CA) is unlikely to be suitable for use as a marker in genomic mapping of sugar beet, in contrast to mammals (15). Broun and Tanksley (28) examined clones hybridizing to $(GA)_n$, $(GT)_n$, and $(ATT)_n$, finding all analyzed clones included perfect repeats of 5–75 copies of the motif. We expect that the *in situ* pattern we observe includes hybridization to such low copy sequences, to microsatellites included within larger, tandemly repeated motifs, and hybridization to very large, perfect or degenerate arrays of microsatellites. The remarkable differences between the motifs (GA) and (CA) and the composite microsatellite (GACA) revealed by in-gel and *in situ* hybridization patterns provide evidence for the specific amplification of each microsatellite at particular chromosomal locations and for the resolution and reliability of our approach. Similarly, the microsatellites (GA), (TA), and (GATA) show marked hybridization differences.

Microsatellites are ubiquitous in most organisms, and there is increasing interest in learning about large scale organization of genomes and their evolution (33). *In vitro* experiments strongly suggest that slippage replication is the main mechanism responsible for the formation and expansion of microsatellite stretches (34), but as a secondary mechanism, reamplification of repetitive DNA sequences including microsatellite arrays may occur as apparently happens with retrotransposons (22) or tandem repeats (35). Slippage may be sequence context-dependent, leading to the large number of alleles found at each locus in, for example, soybean (12). The biological function, if any, of most microsatellites remains unclear, although some have been associated with genetic diseases in human (36, 37), and they have been considered as a cryptic source for genetic variability, hot spots for recombination, or repetitive elements affecting chromatin structure (34, 38, 39). Within eukaryotes, coding sequences are closely similar and often syntenic across wide taxonomic groupings (40–42), but repetitive DNA is a dynamic component of the genome showing major differences in sequence, copy number, and chromosomal distribution between related species. Why some microsatellite sequences amplify and others do not and why some locate at particular centromeres are questions of importance for studies of genome organization. We suggest that knowledge of physical microsatellite distribution is likely to be valuable for choosing microsatellites suitable for use as genetic markers in plants. Our results demonstrate that microsatellite sequences, representing a substantial fraction of the genome, show chromosome-specific amplification in plants. Each microsatellite motif examined here has a characteristic and unique in-gel and *in situ* hybridization pattern, indicating that each amplifies and distributes independently.

We thank A. Brandes, G. Bryan, I. Nanda, M. L. Roose, M. Schmid, T. Schwarzacher, and A. Vershinin for discussion and critical reading

of the manuscript. T.S. is supported by European Union Fellowship ERB4001GT931647.

1. Tautz, D. & Renz, M. (1984) *Nucleic Acids Res.* **12,** 4127–4138.
2. Nowak, R. (1994) *Science* **263,** 608–609.
3. Valdes, A. M., Slatkin, M. & Freimer, N. B. (1993) *Genetics* **133,** 737–749.
4. Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G. & Lathrop, M. (1992) *Nature (London)* **359,** 794–778.
5. Wu, K.-S. & Tanksley, SD (1993) *Mol. Gen. Genet.* **241,** 225–235.
6. Akkaya, M. S., Bhagwat, A. A. & Cregan, P. B. (1992) *Genetics* **132,** 1131–1139.
7. Weising, K., Nybom, H., Wolff, K. & Meyer, W. (1995) *DNA Fingerprinting in Plants and Fungi* (CRC Press, Boca Raton, FL).
8. Nanda, I., Zischler, H., Epplen, C., Guttenbach, M. & Schmid, M. (1991) *Electrophoresis* **12,** 193–203.
9. Becker, J. & Heun, M. (1995) *Plant. Mol. Biol.* **27,** 835–845.
10. Röder, M., Plaschke, J., König, S.-U., Börner, A., Sorrells, M. E., Tanksley, S. D. & Ganal, M. W. (1995) *Mol. Gen. Genet.* **246,** 327–333.
11. Bell, C. J. & Ecker, J. R. (1994) *Genomics* **19,** 137–144.
12. Rafalski, J. A. & Tingey, S. V. (1993) *Trends Genet.* **9,** 275–280.
13. Pedersen, C. & Linde-Laursen, I. (1994) *Chromosome Res.* **2,** 65–71.
14. Nanda, I., Feichtinger, W., Schmid, M., Schröder, J. H., Zischler, H. & Epplen, J. T. (1990) *J. Mol. Evol.* **30,** 456–462.
15. Dietrich, W. F., Miller, J., Steen, R., Merchant, M. A., Damron-Boles, D., *et al.* (1996) *Nature (London)* **180,** 149–152.
16. Arumuganathan, K. & Earle, E. D. (1991) *Plant Mol. Biol. Rep.* **9,** 208–218.
17. Schmidt, T., Schwarzacher, T. & Heslop-Harrison, J. S. (1994) *Theor. Appl. Genet.* **88,** 629–636.
18. Schmidt, T. & Metzlaff, M. (1991) *Gene* **101,** 247–250.
19. Schmidt, T., Jung, C. & Metzlaff, M. (1991) *Theor. Appl. Genet.* **82,** 793–797.
20. Jung, K., Pillen, K., Frese, L., Fähr, S. & Melchinger, A. E. (1993) *Theor. Appl. Genet.* **86,** 449–457.
21. Schmidt, T. & Heslop-Harrison, J. S. (1993) *Genome* **36,** 1074–1079.
22. Schmidt, T., Kubis, S. & Heslop-Harrison, J. S. (1995) *Chromosome Res.* **3,** 335–345.
23. Schmidt, T., Boblenz, K., Metzlaff, M., Kaemmer, D., Weising, K. & Kahl, G. (1993) *Theor. Appl. Genet.* **85,** 653–657.
24. Ali, S., Müller, C. S. & Epplen, J. T. (1986) *Hum. Genet.* **47,** 239–243.
25. Thein, S. L. & Wallace, R. B. (1986) in *Human Genetic Diseases—A Practical Approach*, ed. Davies, K. E. (IRL Press, Oxford), pp. 33–50.
26. Zischler, H., Nanda, I., Schäfer, R., Schmid, M. & Epplen, J. T. (1989) *Hum. Genet.* **82,** 227–233.
27. Arens, P., Odinot, P., van Heusden, A. W., Lindhout, P. & Vosman, B. (1995) *Genome* **38,** 84–90.
28. Broun, P., Tanksley, S. D. (1996) *Mol. Gen. Genet.* **250,** 39–49.
29. Depeiges, A., Goubely, C., Lenoir, A., Cocherel, S., Picard, G., Raynal, M., Grellet, F. & Delseny, M. (1995) *Theor. Appl. Genet* **91,** 160–168.
30. De Jong, J. H. & Oud, J. L. (1979) *Kulturpflanze* **51,** 125–133.
31. Traut, W. (1991) *Chromosomen: Klassische und molekulare Cytogenetik* (Springer-Verlag, Berlin, Heidelberg, New York 1991), pp. 160–174.
32. Lagercrantz, U., Ellegren, H. & Andersson, L. (1993) *Nucleic Acids Res.* **21,** 1111–1115.
33. Charlesworth, B., Sniegowski, P. & Stephan, W. (1994) *Nature (London)* **371,** 215–220.
34. Schlötterer, C. & Tautz, D. (1992) *Nucleic Acids Res.* **20,** 211–215.
35. Ohtsubo, H. & Ohtsubo, E. (1994) *Mol. Gen. Genet.* **245,** 449–455.
36. Richards, R. I. & Sutherland, G. R. (1992) *Trends Genet.* **8,** 249–254.
37. Morell, V. (1993) *Science* **260,** 1422–1423.
38. Tautz, D., Trick, M. & Dover, G. A. (1986) *Nature (London)* **322,** 652–656.
39. Löwenhaupt, K., Rich, A. & Pardue, M. L. (1989) *Mol. Cell. Biol.* **9,** 1173–1182.
40. Paterson, A. H., Lin, Y.-R., Li, Z., Schertz, K. F., Doebley, J. F., Pinson, S. R. M., Liu, S.-C., Stansel, J. W. & Irvine, J. E. (1995) *Science* **269,** 1714–1718.
41. Shields, R. (1993) *Nature (London)* **365,** 297–298.
42. Moore, G., Devos, K. M., Wang, Z. & Gale, M. (1995) *Curr. Biol.* **5,** 737–739.