Nucleic Acids Research, 1998, Vol. 26, No. 3 761–767

Matrix attachment regions and structural colinearity in the genomes of two grass species

Zoya Avramova[1,*], Alexander Tikhonov[1], Mingsheng Chen[2] and Jeffrey L. Bennetzen[1,2]

[1]Department of Biological Sciences and [2]Purdue Genetics Program, Purdue University, West Lafayette, IN 47907, USA

Received October 14, 1997; Revised and Accepted December 4, 1997

ABSTRACT

In order to gain insights into the relationship between spatial organization of the genome and genome function we have initiated studies of the co-linear *Sh2/A1*-homologous regions of rice (30 kb) and sorghum (50 kb). We have identified the locations of matrix attachment regions (MARs) in these homologous chromosome segments, which could serve as anchors for individual structural units or loops. Despite the fact that the nucleotide sequences serving as MARs were not detectably conserved, the general organizational patterns of MARs relative to the neighboring genes were preserved. All identified genes were placed in individual loops that were of comparable size for homologous genes. Hence, gene composition, gene orientation, gene order and the placement of genes into structural units has been evolutionarily conserved in this region. Our analysis demonstrated that the occurrence of various 'MAR motifs' is not indicative of MAR location. However, most of the MARs discovered in the two genomic regions were found to co-localize with miniature inverted repeat transposable elements (MITEs), suggesting that MITEs preferentially insert near MARs and/or that they can serve as MARs.

INTRODUCTION

The higher level structural organization of the genome is believed to be important both for compaction of chromosomes in the nucleus and for regulating genome functions. According to the chromatin domain model (1–3) the genome is folded into structural domains (loops), the bases of which are attached to a proteinaceous nuclear skeleton (matrix). Such loops are believed to provide an additional 1000-fold compaction of the genome, necessary for its accommodation into the interphase nucleus. The DNA sequences (matrix attachment regions, or MARs) anchoring loops of heterogeneous size to this matrix are considered to be important structural elements of the genome. Their ability to affect gene expression has been shown (3–6). In plants MARs have been reported to play a role in reducing both position effects and homology-dependent gene silencing (7–12). Besides being structural elements, therefore, MARs are believed to bear functional information as well.

For obvious reasons researchers have focused mainly on the gene-containing fraction of the genome. Hence, most of the existing information on the structural organization of the genome describes individual genes and their immediate surroundings. Information on domain organization and chromosome folding at a supragenic level in animal systems is limited to only a few studies: a 320 kb continuum of the *Drosophila rosy–Ace* locus (13), an 800 kb region of *Drosophila* chromosome 1 (14,15), the 240 kb amplicon of the chinese hamster dihydrofolate reductase gene (16) and 200 kb around the mouse heavy chain IgH locus (17). The first study in plants devoted to the higher order structural organization of a large chromosomal continuum was focused on a 290 kb region around maize *Adh1* (18). The location of MARs along this chromosomal segment defined several loops of various sizes and a strong, although not perfect, correlation between MAR locations and the junctions of repetitive and low copy number DNA blocks. The distribution of the different classes of DNA within this continuum (19) with respect to the structural loops revealed that the long stretches of mixed classes of highly repetitive DNAs are segregated into topologically sequestered units (18). It was interesting, therefore, to study the possible loop folding of grass genomic regions void of highly repetitive DNA blocks in their intergenic space, as is the case for the *Sh2/A1*-homologous regions of sorghum and rice (20–22).

Earlier we showed that the *Adh1* gene was positioned in an individual loop (18). However, lack of information regarding the presence and location of other genes in the region did not allow us at that time to pursue a possible correlation between the structural organization of genes and their function. We have addressed this question here by examining the putative higher level structural organization of two large genomic continuums of known gene composition. This is the first attempt to compare the possible spatial organization of homologous genomic regions in two different species.

A characteristic feature of eukaryotic genomes is the enormous variation in genome size, which bears little relation to differences in organism complexity or to the number of genes that code for proteins (23). Much of this variation is due to differences in the amount of repetitive DNA. Recombinational mapping of different grass genomes has indicated extensive conservation of both gene content and gene map order (24,25), despite great variation in genome size

*To whom correspondence should be addressed: Tel: +1 765 494 9837; Fax: +1 765 496 1496; Email: zavramov@bilbo.bio.purdue.edu

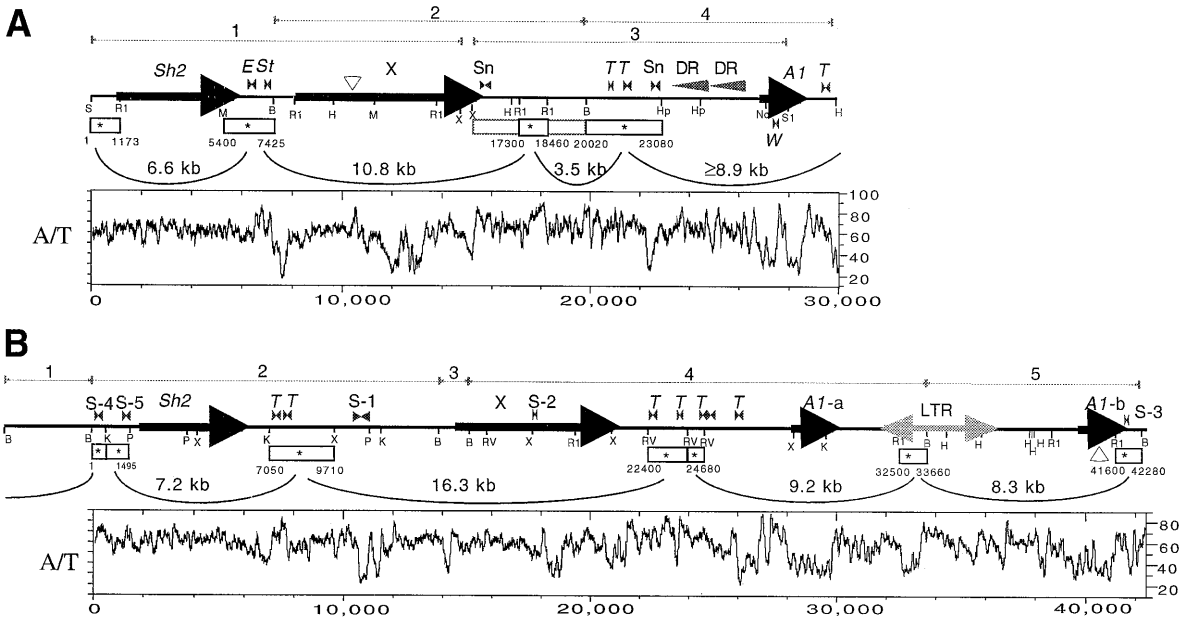

Figure 1. Maps of the *Sh2/A1*-homologous regions of rice and sorghum. (**A**) The black horizontal arrows indicate the location of the *Sh2*, X and *A1* homologs on the rice contig. The four overlapping clones covering 30 kb of the rice region are shown above. The sites for the restriction enzymes used in the screening are shown: X, *Xho*I; S, *Sac*I; RI, *Eco*RI; M, *Mlu*I; Hp, *Hpa*I; H, *Hin*dIII. The small black head-to-head triangles illustrate the MITEs identified in the region: **T**, *tourist*; **St**, *stowaway*; **W**, *wanderer*; **E**, *explorer*; **Sn**, *snabo* (20). The two arrowheads upstream of the *A1* homolog show the size and location of a pair of direct repeats, while the vertical empty triangle marks the position of simple sequence repeats (20). The restriction fragments binding to the matrix are shown as open boxes with a star under the bar. The numbers flanking the boxes correspond to the map positions of the restriction cuts generating the respective fragment. The location of the 1.1 kb MAR inside the larger 4.8 kb *Bam*HI–*Xho*I MAR, 3′ of gene X, is shown as a box inside a larger box. The putative loops and their lengths are shown at the lower level. The distribution patterns of A/T are shown in the boxed area on the same scale as the restriction map above. (**B**) Five adjacent clones covering 50 kb of the colinear *Sh2*/*A1* region of sorghum are shown on top. Small vertical bars show the *Bam*HI sites used for their subcloning. The restriction enzymes used in the analysis are: B, *Bam*HI; K, *Kpn*I; P, *Pac*I; X, *Xho*I; RV, *Eco*RV; RI, *Eco*RI; H, *Hin*dIII. **S**-1–**S**-5 are putative new MITEs defined through their ability to form snap-back structures and/or because they were found around other genes in the databases (22; A.Tikhonov, unpublished results). The solo LTR between the two duplicated *A1* homologs is shown as a double-headed light arrow. All other designations are as in (A).

and chromosome number (26). Recent studies have shown that interspecies gene content and order have often been preserved also at the 200–500 kb level (20–22,27–29).

With this background we have asked three questions. What will be the potential of large chromosomal regions of known sequence composition to fold into individual structural units? Given the micro-colinearity of grass genomes, will the folding of colinear regions into structural units follow a similar pattern? Will there be DNA sequence motifs that are common and/or conserved in plant MARs?

About 50 and 30 kb of sorghum and rice DNA respectively, containing *A1/Sh2*-homologous regions, were screened for the location and distribution of MARs as anchors for the bases of putative loops (domains). Several MAR-containing fragments (four in rice and seven in sorghum) were isolated and their sequences compared with each other and with reported characteristics of animal MARs. The results indicated significant preservation of structural organization but no detected conservation of sequences or motifs responsible for folding of this region.

MATERIALS AND METHODS

Materials

Subclones of rice and sorghum bacterial artificial chromosomes (BACs) containing *Sh2* and *A1* homologs have been described previously (20,21; Fig. 1). Restriction enzymes and T4 polynucleotide kinase were from New England Biolabs. Calf intestine alkaline phosphatase (Pharmacia) was used for dephosphorylation.

Nuclear and matrix preparations

Leaves from 3-week-old rice (variety Teqing) and sorghum (cultivar BT×623) seedlings were used for isolation of nuclei. Excised leaves were immediately frozen in liquid nitrogen and finely ground in a mortar. Nuclear isolation was according to the protocol established previously (30). Aliquots containing 3–5 A_{260} units of nuclei were dispersed in 50 mM Tris, 0.1 M NaCl, 5 mM $MgCl_2$ buffer, pH 7.2, in 70% glycerol and stored for several months at –80°C. Nuclear matrices were prepared by the high salt extraction procedure, as described (30).

MAR binding experiments

Each of the subclones was digested with a combination of restriction enzymes as shown in the legend to Figure 2. The fragments resulting from restriction were dephosphorylated and end-labeled with [γ-^{32}P]ATP (Amersham) in One-Phore All Buffer PLUS (Pharmacia). The *in vitro* binding method was used to screen the rice and sorghum clones for the presence of MARs, essentially as described previously (30). In pilot experiments the amount of nuclear matrices and competitor DNA were established with the

purpose of carrying out the binding assays under reproducible conditions and to eliminate weak and background associations. In a typical assay 100 µg/ml extensively sheared *Escherichia coli* DNA were mixed with nuclear matrix aliquots, corresponding to 0.5 A_{260} units of sorghum nuclei and 0. 35 A_{260} units of rice nuclei per binding sample. After a 10 min incubation the labeled restriction fragments were added and the binding reaction was usually carried out for several hours or overnight at room temperature with shaking. The separation of matrix-bound (B, bound) from unbound DNA was accomplished by centrifugation for 2 min in an Eppendorff centrifuge. The sedimented fraction was treated with proteinase K in TE buffer containing 1% SDS for 3–4 h at 37°C and loaded in 1% agarose gels, next to a sample of total (T) input DNA. The amount of input DNA loaded on the gel was 50% of that used in the binding reaction, except the input DNA shown with rice clone 1, which represents 25% of the labeled probe used in the assay. It was necessary to load a lower amount of the input DNA in the gel in order to achieve a better resolution of the closely fractionating restriction fragments. Identification of bound fragments was facilitated by the presence of detailed restriction maps of the two regions (21).

MAR sequence analysis

A total of 30 035 bp (GenBank accession no. U70541) and 42 447 bp (GenBank accession no. AF010283) of the colinear *Sh2/A1*-homologous regions of rice and sorghum respectively have been sequenced (20,22). After location of the MARs, therefore, we were able to immediately analyze their sequence composition. The GCG sequence analysis software package, Version 8. 0 (Genetics Computer Group Inc., University Research Park, Madison, WI), was used. The distributions of A/T and various 'boxes' were estimated by the Window program with default setup of window size at 100 and shift increment at 3.

RESULTS

Identification and mapping of MARs in the *Sh2/A1*-homologous region of rice

Screening of a BAC library containing rice DNA with a maize *Sh2* probe has led to the isolation of a clone containing a 130 kb insert (21). Detailed molecular analysis of this region indicated that a homolog of the maize *A1* gene was present downstream of the *Sh2* homolog. The order of the two genes, as well as the direction of their transcription, was the same as in maize (31). A major difference, however, was that the two genes were 19 kb apart in rice and 140 kb apart in maize. Subsequently, ~30 kb of the region covering the two genes in rice was completely sequenced and a third gene, gene X (encoding a putative transcription factor), was discovered between the *Sh2* and *A1* homologs (20,21; Fig. 1A). Numerous elements with structures corresponding to miniature inverted repeat transposable elements (MITEs) (32), a simple sequence repeat and a direct tandem duplication of 1432 bp were also identified (20).

Four overlapping clones, covering 30 kb of the rice region encompassing the *Sh2* and *A1* homologs, were individually screened for the presence and location of MARs. The fragments shown in the right hand lanes represent DNA preferentially retained by the matrix and are shown as boxes with a star (Fig. 1A).

Clone 1, containing a *Sac*I–*Xho*I insert covering 15 kb at the 5′-end of the region, was digested with a combination of

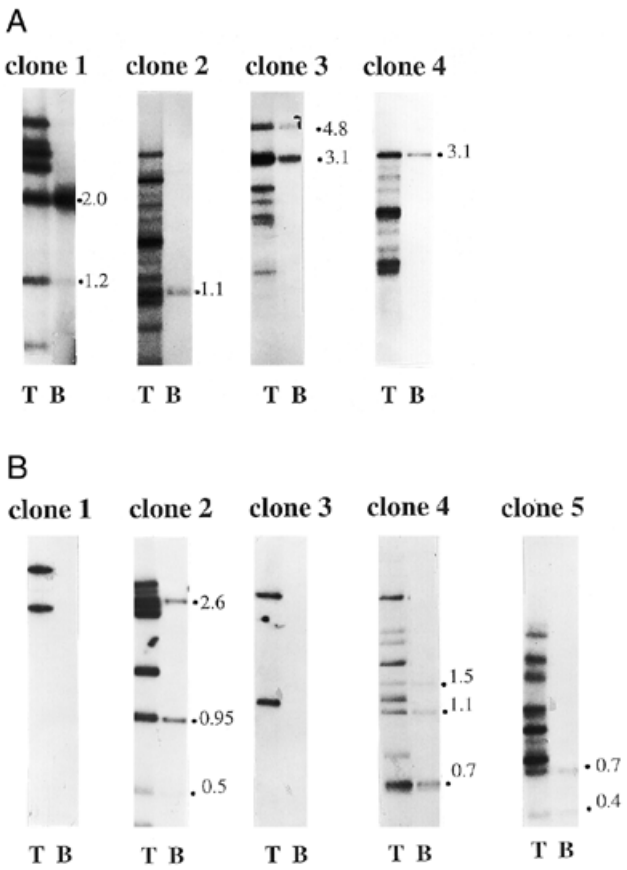

Figure 2. Screening for MAR activity in fragments from the *Sh2/A1*-homologous regions of rice and sorghum. (**A**) Individual overlapping clones, covering the rice genomic region as shown in Figure 1A, were digested with the following restriction enzymes: clone 1 with *Sac*I, *Eco*RI, *Mlu*I and *Bam*HI; clone 2 with *Bam*HI, *Eco*RI, *Mlu*I and *Hin*dIII; clone 3 with *Xho*I, *Bam*HI, *Hpa*I and *Sac*I; clone 4 with *Bam*HI, *Hpa*I, *Eco*RI and *Hin*dIII. The fragments generated by the respective enzyme treatments were labeled and are shown in the left hand lane of each panel (T, total input DNA). The adjacent lane shows the fragments bound preferentially to the matrix (B, bound DNA). In all but one T lane the amount of loaded labeled DNA is 50% of that present initially in the binding. For clone 1 it is 25%. The sizes of the bound fragments are shown by the numbers to their right and their locations are shown by the starred boxes under the central bar in Figure 1A. (**B**) Individual clones covering the sorghum region, as shown in Figure 1B, were digested with the following restriction enzymes: clone 1 with *Bam*HI; clone 2 with *Bam*HI, *Kpn*I, *Pac*I and *Xho*I; clone 3 with *Bam*HI; clone 4 with *Eco*RV, *Eco*RI, *Xho*I and *Bam*HI; clone 5 with *Bam*HI, *Hin*dIII and *Eco*RI. Lane indications are as in Figure 2A above. The location of the fragments binding to the nuclear matrix are shown by the starred boxes under the central bar in Figure 1B.

restriction enzymes and was tested for MAR activity. In the first panel of Figure 2A the total (T) input fragments before binding and those bound (B) to isolated rice nuclear matrices in the presence of competitor DNA are shown. A strong binding was observed for the 2.0 kb fragment generated with *Mlu*I and *Bam*HI. This region immediately flanks the rice *Sh2* homolog at its 3′-end. A weaker binding is displayed by a 1.2 kb fragment located 5′ of the gene. These two neighboring attachment sites delineate a putative loop of ~6.6 kb containing the *Sh2* homolog. It should be pointed out that different functions have been suggested for strong and weak MARs (2,15) and that, in at least one case, involvement of a strong MAR in attenuating transgene silencing

has been reported (9). While this is certainly an interesting issue, it is beyond the scope of this work. Hence, weak and strong MARs are defined arbitrarily in this study, based solely on the apparent differences in band density of a bound fragment relative to the input probe.

An adjacent clone (clone 3), covering ~13 kb downstream, contained two large MAR fragments (Fig. 2A). The ~4.8 kb *Xho*I–*Bam*HI matrix binding fragment covered a region 3′ of gene X, enclosing it in an individual loop (Fig. 1A). This 4.8 kb MAR was further mapped. After digestion of the overlapping ~13 kb *Bam*HI clone (clone 2) and testing the matrix binding capacity of the resulting smaller fragments only a 1.1 kb *Eco*RI fragment bound to the matrix (Fig. 2A). The adjacent 3.1 kb *Bam*HI/*Hpa*I MAR is seen in both clone 3 and in the overlapping clone 4.

Mapping MARs in the sorghum *Sh2/A1*-homologous region

When a sorghum BAC library was screened with a maize *Sh2* probe (21) a positive clone containing 80 kb of DNA was selected and characterized in detail. As in the case of rice, the presence, order and direction of transcription of the *Sh2* and *A1* homologs were the same as in maize and, as in rice, the distance between the two genes was ~19 kb. In addition, a direct duplication of *A1* was discovered ~10 kb downstream (Fig. 1B). A gene homologous to gene X was discovered between the *Sh2* and *A1* homologs in sorghum as well, further supporting the genetic colinearity of the region (21,22). Several MITEs belonging to different classes of mobile elements, a solo long terminal repeat (LTR) of a retroelement and a simple sequence repeat were identified in intergenic locations (22; Fig. 1B).

Five adjacent clones, covering ~50 kb of the region, were screened for MARs. Matrix complexes prepared from sorghum leaf nuclei were used in the binding assay. The results of these experiments are shown in Figure 2B. Clone 1, containing a 3.5 kb sorghum insert and located at the most 5′-end of the region studied, did not show any matrix binding capacity. In the adjacent 15 kb region (clone 2) one weak and two prominent binding sites were revealed: on the 2.6 kb *Kpn*I–*Xho*I intergenic fragment separating the *Sh2* and X homologs and on the 0.95 kb *Kpn*I–*Pac*I fragment 5′ of the *Sh2* homolog. A weakly bound 0.5 kb band corresponds to the *Bam*HI–*Kpn*I fragment located immediately upstream of the 0.95 kb fragment. It is possible that these adjacent binding fragments are part of the same anchorage site. Clone 3, covering a 1.2 kb region between two large clones, did not reveal any potential attachment site. Clone 4 contained an insert with ~19 kb of the region. Two genes were located in clone 4: the putative transcription factor gene, gene X and an *A1* homolog (A1-a). Two MARs were identified on the 1.5 kb *Eco*RV fragment at map positions 22400–23970 and on the adjacent 0.5 kb region, at positions 23970–24680. These two attachment points may act in concert at the 3′-end of a putative loop enclosing the gene X homolog. A third attachment point, located some 9 kb downstream, was identified in a 1.1 kb *Eco*RI–*Bam*HI fragment. It mapped to a region occupied by the solo LTR and closed a putative loop containing the A1-a gene.

In the 3′-end of the region a duplicated *A1* homolog (A1-b) was located. The fifth clone tested for matrix binding contained an 8.5 kb insert encompassing A1-b. After digestion with *Bam*HI, *Hin*dIII and *Eco*RI, a 0.7 kb *Eco*RI–*Bam*HI fragment bound to the matrix (last panel in Fig. 2B). It enclosed A1-b in a separate putative loop of 8.3 kb. A weak binding was also observed for the 0.4 kb *Eco*RI fragment located immediately upstream of the 0.7 kb MAR, suggesting that these two fragments also represent a single attachment point.

Sequence characteristics of the rice and sorghum MARs

Hybridization and sequence analysis of the two colinear genomic regions indicated that the sequence homology between the two species was limited to the regions occupied by genes (20–22). The regions containing MARs did not show conservation of their primary sequence (20, 22). It was unexpected, therefore, that the general placement pattern of the MARs, with respect to the neighboring genes, was so remarkably well preserved.

Since MARs were initially identified (33–35), the nature of the DNA sequences responsible for MAR activity has not been fully defined. Comparison of the sequences of numerous MAR-containing DNA fragments has indicated that they are A/T rich. This is usually displayed as motifs containing various combinations of A and T residues: A boxes, T boxes, base unpairing sequences (BURs), consensus elements for topoisomerase II, etc. (34–38). Therefore, we compared the composition of these plant MARs with regard to criteria established earlier for MARs. The ability to analyze several MARs belonging to two different plant species and the availability of the primary sequence of these large genomic regions permitted a detailed evaluation of MAR composition.

When the rice and sorghum MARs were examined for the presence of A or T residues a general tendency for enrichment in A/T was observed: all MAR-containing fragments were 70–80% A/T. Comparing the A/T profile of the whole region, however, showed a high level of A/T in several locations across the entire region in both species (Fig. 1). Evidently, AT-richness *per se* could not be a reliable criterion for MAR prediction. All three *A1* homologs and both gene X homologs displayed a low A/T content. In contrast, the *Sh2* homologs were rather high in A/T (Fig. 1).

In the search for a characteristic MAR sequence several motifs have been reported as elements clustered in MAR regions: the ‘A box’ (AATAAAYAAA), the ‘T box’ (TTWTWTTWTT), ‘BURs’ (AATATATT/AATATT) and topoisomerase II consensus binding sites from *Drosophila* (GTNWAYATTNATNNG) or mouse (RNYNNCNNGYNGKTNYNY) (reviewed in 38).

Both DNA strands of each of the rice and sorghum MARs were tested for the presence of such motifs. As shown in Figures 3 and 4, many of these motifs were found in the MAR-containing fragments, suggesting similar overall sequence composition for MARs independent of their species of origin. A notable exception, though, is the absence of a *Drosophila* topoisomerase II consensus motif from the MARs, as well as from the entire tested regions. A similar lack of this consensus motif has just been reported for the MARs located in the plastocyanin gene region of *Arabidopsis thaliana* (39). A motif, believed to be a specific marker for MARs in *Arabidopsis*, has been deduced (39). However, comparison of this consensus with the sequences from the sorghum and rice regions failed to uncover preferential appearance of this motif in the MARs (not shown).

Recently a 25 nt recognition sequence for SATB1 has been found as a key MAR binding motif for matrix proteins (40). SATB1 is a novel type of DNA binding protein that recognizes a specific sequence context in which one DNA strand exclusively consists of mixed A, T and C nucleotides (A/T/C) and lacks G.

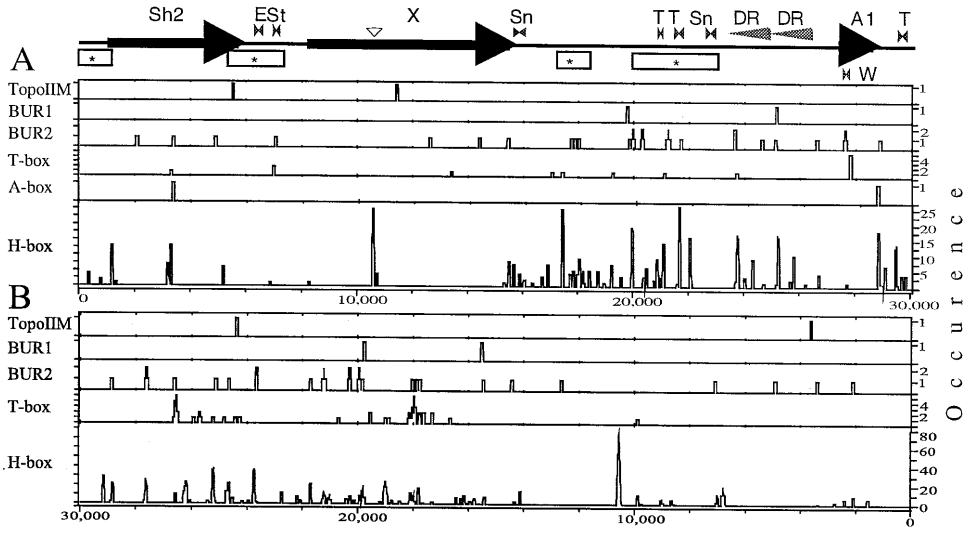

Figure 3. Distribution of 'MAR-specific' motifs in the rice genomic region homologous to *Sh2/A1*. The upper bar represents 30 kb of DNA sequence containing the three gene homologs, MITEs, direct repeats and MARs, as described in Figure 1A. In the panels below the computational analysis for the occurrence and distribution of various sequence motifs are shown for the sense DNA strand (**A**) and for the opposite DNA strand (**B**). TopoIIM (RNYNNCNNGYNGKTNYNY) is the mouse topoisomerase II consensus; BUR 1 (AATATATTT) and BUR 2 (AATATT) are the base unpairing consensus motifs. Also shown are T box (TTWTWTTWTT), A box (AATAAAYAAA) and H box (an uninterrupted stretch of A/T/C) motifs. The size of the bars correspond to the occurrence of clustered motifs found at close locations. Note that the vertical scales vary for the different motifs. No sequence matching the *Drosophila* topoisomerase II consensus (GTNWAYATTNATNNG) was found.

Clustered A/T/C sequences found in MARs have a strong tendency to unwind by base unpairing (41) and it is believed that this unwinding property confers high affinity binding to the nuclear matrix (37).

We screened the two colinear regions of sorghum and rice for the distribution of A/T/C. A stretch of 25 nt of uninterrupted A/T/C is called an 'H box'. Once again, specific concentration of the motifs in MAR regions were not revealed, although some preferential clustering of H boxes in the intragenic space, including the MARs, was observed (Figs 3 and 4).

From the results presented so far it may be concluded that none of these sequence motifs could be used as a reliable probe for predicting a MAR function. However, we observed that in eight of 11 cases the MAR-containing restriction fragments also contained one or more MITEs (Fig. 1A and B). This observation raises the intriguing question of whether it is the MITEs that carry MAR activity or whether MITEs tend to integrate close to MARs, or both.

DISCUSSION

In maize *Sh2* and *A1* map ~140 kb apart (31). Molecular analysis of the comparable loci in two other grasses, rice and sorghum, has indicated that gene arrangement and composition are conserved in these regions for these species (20–22). The close physical and recombinational linkage of these two genes (31,42) makes this region particularly informative for studies of intergenic chromosomal organization. The complete sequence information available for the regions and the identification of individual elements and genes (20–22) make it an excellent model for studies of a possible relationships between genome structural organization and function. An unexpected feature of this region in both species was the absence of retroelements in intergenic spaces, in contrast to their abundance in maize (43). The only exception was a solo LTR present in sorghum. This retroelement segment was found to also carry the MAR that could segregate the duplicated *A1* homologs into individual loops. Earlier we identified MARs in regions of repetitive DNA around maize *Adh1* (18) and a few of them were shown to be carried by retroelements (43). It should be noted that not all members of the same retrotransposon family displayed matrix binding activity. Recently MAR activity has been found inside another retroelement, part of the transformation booster sequence (TBS) from *Petunia* (44). The authors suggested a role for this MAR in increasing the transformational and/or recombinational activity of TBS-containing plasmids.

Mapping the MARs along the chromosome continuums in the two species allowed us to uncover commonalities in the predicted organization of the two genomes. First, all genes present in the region were placed in individual loops, defined by neighboring MARs. MARs identified in two adjacent restriction fragments were considered as parts of the same anchoring site. Each of the duplicated sorghum *A1* homologs was found in a separate loop, separated by the MAR located in the solo LTR. Analysis of data in the literature seemed to suggest that placement of genes into individual relatively small (3–10 kb) loops is a common pattern in plant genomes. Thus all four genes in the 17.1 kb of the soybean lectin locus are segregated into separate domains (7) the tomato heat shock cognate 80 gene (45) and the maize proton H^+ ATPase gene (Avramova *et al.*, unpublished results) are placed in individual putative loops. The β-phaseolin gene has been found in a 3 kb loop, the smallest reported so far (46). A recent study of 16 kb in the *A.thaliana* plastocyanin region, containing seven different genes, provided insight into the loop organization of a small plant genome (39). In this case each putative structural loop contained two neighboring genes. The possible significance of this type of gene arrangement within a loop remains to be studied.

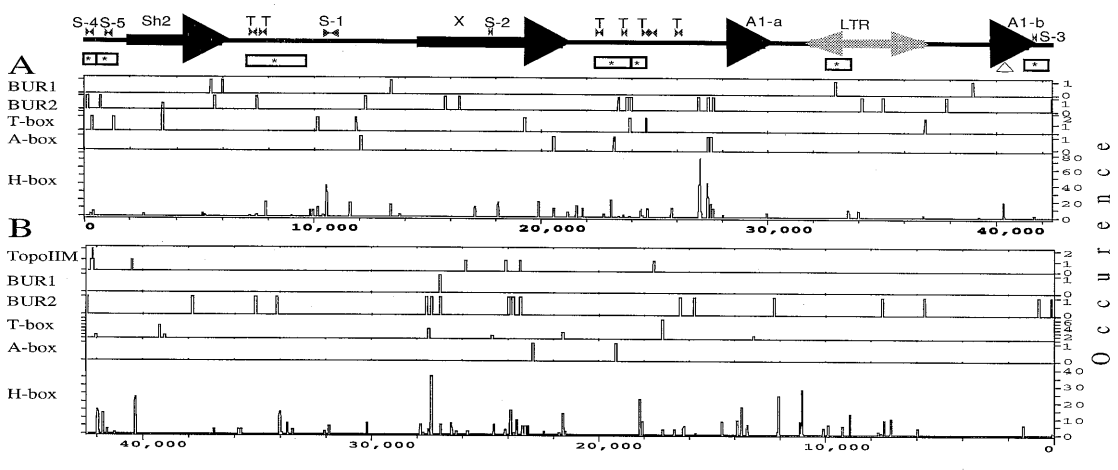

Figure 4. Distribution of 'MAR-specific' motifs in the sorghum region homologous to maize *Sh2/A1*. The upper bar represents 42 kb of DNA sequence containing the four gene homologs, MITEs, a solo LTR and MARs, as described under Figure 1B. In the panels below computational analysis for the occurrence and distribution of various sequence motifs are shown for the sense DNA strand (**A**) and for the opposite DNA strand (**B**). Designations are the same as those described in Figure 3.

Second, although the sizes of the proposed loops vary, homologous gene domains are comparable in sorghum and rice. The size of the loop containing the rice *A1* homolog has not been determined, because its 3′-end was beyond the sequence available on our clone. A common feature of all three *A1* homologs, however, is that their promoters are placed relatively far from the base of the loop, with various repetitive elements present between their 5′ MARs and the respective transcription start sites. The *Sh2* homologs of both rice and sorghum are placed in smaller predicted loops, while the genes for the putative transcription factor appear to occupy the largest structural domains in the regions.

Third, hybridization and subsequent sequencing of the two colinear genomic regions indicated that the sequence homology between the two species was strictly limited to the regions occupied by the genes (20–22). The regions containing the MARs did not show sequence conservation, aside from the fact that all were A/T rich and possessed some common DNA motifs. Since the sequence heterogeneity outside the genes was an established fact, it was not expected that the placement pattern of the MARs, with respect to the neighboring genes, would be so well preserved in the two species. This fact suggests that it is not only the gene order on the chromosome but also the placement of the genes into structural units that is evolutionarily conserved. This structural conservation suggests a relationship between the location of genes in chromatin domains and their capacity to function. However, the nature of this relationship and the ways in which the structural domains might affect gene regulation are still an enigma.

Despite many previous studies, there is no definitive answer regarding features that make a DNA sequence perform as a MAR. A generally accepted criterion is a high A/T content for MARs. The MAR fragments isolated from rice and sorghum were 70–85% A/T, but high A/T composition was also found at numerous locations outside the MARs. This observation supports the conclusion that an abundance of A and T residues is not sufficient for MAR function (reviewed in 38).

MARs have been found to be enriched in various A/T-containing motifs or 'boxes' and, for a long time, the presence of the *Drosophila* topoisomerase II consensus motif has been considered an indispensable MAR feature (34,35,47). In contrast, no *Drosophila* topoisomerase II motif was found in the MARs of rice, sorghum or *A.thaliana* (39). Later different recognition motifs were discovered for topoisomerase II of vertebrate origin (48) and this raised a question about the meaning of the *Drosophila* topoisomerase II consensus motifs in the MARs of vertebrates. Motifs similar to the mouse consensus were subsequently found in many animal MARs (reviewed in 38) and a few mouse 'Topo II boxes' were identified in the colinear *Sh2/A1* regions. Some co-localized with MARs. The possible significance, if any, of this fact is not clear at present and it will be interesting to map the locations of a plant topoisomerase II binding sequence when it becomes available. As a result of our analysis of the distribution of various sequence motifs in both isolated MARs and along adjacent genomic regions it became evident that the occurrence of previously identified 'MAR motifs' is not indicative of a MAR location in this region of the rice or sorghum genomes.

Currently it is believed that DNA structure (perhaps a narrow minor groove, a tendency to form bent DNA, a tendency for single-strandedness and a tendency to form looped structures) is responsible for the matrix binding activity of a region (36–38,49–52). The data indicating that MARs may be enriched in inverted repeats (reviewed in 38) are of particular interest. Longer palindromes, 144 bp or bigger, are believed to convert into cruciform structures under torsional stress. HMG1 protein has been shown to specifically bind cruciform DNA (53) and HMG1 has been found to specifically bind a MAR in a nuclear matrix preparation (54). The small DNA elements (MITEs), abundantly present in plant genomes around various genes, often contain inverted repeats and may form cruciform structures (reviewed in 32). Most of the MARs discovered in the two genomic regions described here were found to co-localize with MITEs. This observation raises the question of whether MITEs preferentially insert near MARs and/or whether these elements can

serve as MARs. As is also true of the MARs associated with some members of a few retroelement families, it does not seem likely that mobile DNAs would be key determinants of chromosome structure. However, once present at a new genomic location a mobile DNA might be selected for new local functions, like gene regulation, recombinational initiation (44) or MAR activity. Further studies will investigate the evolution, function and detailed structure of the MARs and MITEs in various grass genomes.

ACKNOWLEDGEMENTS

The authors are indebted to Phillip SanMiguel for providing us with the sorghum clones and their restriction maps. The work was supported by research grants from the USDA/NRICGP (no. 94-37300-0299 to J.L.B. and no. 93-37300-8769 to Z.A.).

REFERENCES

1 Gross,D. and Garrard,W. (1987) *Trends Biochem. Sci.*, **12**, 293–297.
2 Gasser,S.M. (1988) In Kahl,G. (ed.), *Architecture of Eukaryotic Genes*. VCH Verlagsgeselschaft, Wienheim, Germany, pp. 461–471.
3 Bonifer,C., Hecht,A., Saueressig,H., Winter,D. and Sippel,A.E. (1991) *J. Cell Biochem.*, **47**, 99–108.
4 Stief,A., Winter,D., Straetling,W. and Sippel,A. (1989) *Nature*, **341**, 343–345.
5 Phi-Van,L., von Kries,J.P., Ostretag,W. and Straetling,W.H. (1990) *Mol. Cell. Biol.*, **10**, 2302–2307.
6 McKnight,R.A., Shamey,A., Sankaran,L., Wall,R.G. and Hennighausen,I. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 6943–6947.
7 Breyne,P., van Montague,M., Depicker,A. and Gheysen,G. (1992) *Plant Cell*, **4**, 463–471.
8 Allen,G.C., Hall,G.E., Childs,L.C., Weissinger,A.K., Spiker,S. and Thompson,W.F. (1993) *Plant Cell*, **5**, 603–613.
9 Allen,G.C., Hall,G.E., Michalowski,S., Newman,W., Spiker,S., Weissinger,A.K. and Thompson,W.F. (1996) *Plant Cell*, **8**, 899–913.
10 Mlynarova,L., Jansen,R.C., Conner,A.J., Stiekema,W.J. and Nap,J.-P. (1995) *Plant Cell*, **7**, 599–609.
11 Mlynarova,L., Loonen.A., Heldens,J., Jansen,R.C., Keizer,P., Stiekema,W.J. and Nap,J.-P. (1994) *Plant Cell*, **6**, 417–426.
12 Thompson,W.F., Allen,G.C., Hall,G.,Jr and Spiker,S. (1996) In Gustafson,J.P. and Flavell,R.B. (eds), *Genomes of Plants and Animals*. Plenum Press, New York, NY, pp. 243–269.
13 Mirkovitch,I., Spiere,P. and Laemmli,U.K (1986) *J. Mol. Biol.*, **190**, 255–258.
14 Surdej,P., Got,C., Rosset,R. and Miassod,R. (1990) *Nucleic Acids Res.*, **18**, 3713–3722.
15 Brun,C., Dang,Q. and Miassod,R. (1990) *Mol. Cell. Biol.*, **10**, 5455–5463.
16 Dijkwell,P.A. and Hamlin,J.L. (1988) *Mol. Cell. Biol.*, **8**, 5398–5409.
17 Cockerill,P.N. (1990) *Nucleic Acids Res.*, **18**, 2643–2648.
18 Avramova,Z., SanMiguel,P., Georgieva,E. and Bennetzen,J. (1995) *Plant Cell*.,**7**, 1667–1680.
19 Springer,P., Edwards,K.S. and Bennetzen,J.L., (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 863–867.
20 Chen,M. and Bennetzen,J.L. (1996) *Plant Mol. Biol.*, **32**, 999–1001.
21 Chen,M., SanMiguel,P., de Olivieira,A., Woo,S.-S., Zhang,H., Wing,R. and Bennetzen,J. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 3431–3435.
22 Chen,M., SanMiguel,P. and Bennetzen,J.L. (1997) *Genetics*, in press.
23 Orgel,L.E. and Crick,F.H. (1980) *Nature*, **284**, 604–607.
24 Hulbert,S., Richter,T.E., Axtell,J.D. and Bennetzen,J.L (1990) *Proc. Natl. Acad. Sci. USA*, **47**, 99–108.
25 Ahn,S., Anderson,J.A., Sorrells,M.E. and Tanksley,S.D. (1993) *Mol. Gen. Genet.*, **241**, 483–490.
26 Arumuganathan,E. and Earle,E.D. (1991) *Plant Mol. Biol. Rep.*, **9**, 208–218.
27 Dunford,R.P., Kurata,N., Laurie,D.A., Money,T.A. Minobe,Y. and Moore,G. (1995) *Nucleic Acids Res.*, **23**, 2724–2728.
28 Kilian,A., Kudrna,D.A., Kleinhofs,A., Yano,M., Kurata,N., Steffenson,B. and Sasaki,T. (1995) *Nucleic Acids Res.*, **23**, 2729–2733.
29 Avramova,Z., Tikhonov,A., SanMiguel,P., Jin,Y.-K., Liu,C., Woo,S.-S., Wing,R. and Bennetzen,J. (1996) *Plant J.*, **10**, 1163–1168.
30 Avramova,Z. and Bennetzen,J.L. (1993) *Plant Mol. Biol.*, **22**, 1135–1143.
31 Civardi,L., Xia,Y., Edwards,K.J., Schnable,P.S. and Nikolau,B.J. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 8268–8272.
32 Wessler,S.R., Bureau,T.E. and White,S.E. (1995) *Curr. Opin. Genet. Dev.*, **5**, 814–821.
33 Mirkovitch,I., Mirault,M.-E. and Laemmli,U.K (1984) *Cell*, **39**, 223–232.
34 Cockerill,P. and Garrard,W.T. (1986) *Cell*, **44**, 273–282.
35 Gasser,S.M. and Laemmli,U.K. (1986) *EMBO J.*, **5**, 511–517.
36 Kaes,E., Izaurralde,E. and Laemmli,U.K. (1989) *J. Mol. Biol.*, **210**, 587–599.
37 Bode,J., Kohwi,Y., Dickinson,L., Joh,T., Klehr,D., Mielke,C. and Kohwi-Shigematsu,T. (1992) *Science*, **255**, 195–197.
38 Boulikas,T. (1995) *Int. Rev. Cytol.*, **162A**, 279–388.
39 van Drunen,C.M., Oosterling,R.W., Keultjes,G.G.M., Weisbeek,P.J., van Driel,R. and Smeekens,J.C.M. (1997) *Nucleic Acids Res.*, **25**, 3904–3911.
40 Wang,B., Dickinson,L., Koivunen,E., Ruoslahti,E. and Kohwi-Shigematsu,T. (1995) *J. Biol. Chem.*, **270**, 23239–23242.
41 Kohwi-Shigematsu,T. and Kohwi,Y. (1990) *Biochemistry*, **29**, 9551–9560.
42 Xu,X., Hsia,A.-P., Zhang,L., Nikolau,B.J. and Schnable,P.S. (1995) *Plant Cell*, **7**, 2151–2161.
43 SanMiguel,P., Tikhonov,A., Jin,Y.-K., Motchoulskaya,N., Zakharov,D., Berhan,A., Springer,P., Edwards,K., Lee,M., Avramova,Z. and Bennetzen,J. (1996) *Science*, **274**, 765–768.
44 Galliano,H., Muller,A.E., Lucht,J.M. and Meyer,P. (1995) *Mol. Gen. Genet.*, **247**, 614–622.
45 Chinn,A. and Comai,L. (1996) *Plant Mol. Biol.*, **32** 959–968.
46 van der Geest,A.H.M., Hall,G.E., Spiker,S. and Hall,T.C. (1994) *Plant J.*, **6**, 413–423.
47 Jarovaya,O., Hancock,R., Lagarkova,M., Miassod,R. and Razin,S. (1996) *Mol. Cell. Biol.*, **16**, 302–308.
48 Spitzner,J.R. and Muller,M.T. (1988) *Nucleic Acids Res.*, **16**, 5533–5556.
49 Homberger,H.P. (1989) *Chromosoma*, **98**, 99–104.
50 von Kries,J.P., Phi-Van,L., Diekman,S. and Straetling,W. (1990) *Nucleic Acids Res.*, **18**, 3881–3885.
51 Mielke,C., Kohwi,Y., Kohwi-Shigematsu,T. and Bode,J. (1990) *Biochemistry*, **29**, 7475–7485.
52 Kay,V. and Bode,J. (1994) *Biochemistry*, **33**, 367–374.
53 Bianchi,M.E., Beltrame,M. and Paonessa,G. (1989) *Science*, **243**, 1053–1059.
54 Ivanchenko,M. and Avramova,Z. (1992) *J. Cell. Biochem.*, **50**, 190–200.