Structural Domains and Matrix Attachment Regions along Colinear Chromosomal Segments of Maize and Sorghum

Alexander P. Tikhonov, Jeffrey L. Bennetzen, and Zoya V. Avramova
Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907

Although a gene’s location can greatly influence its expression, genome sequencing has shown that orthologous genes may exist in very different environments in the genomes of closely related species. Four genes in the maize alcohol dehydrogenase (adh1) region represent solitary genes dispersed among large repetitive blocks, whereas the orthologous genes in sorghum are located in a different setting surrounded by low-copy-number DNAs. A specific class of DNA sequences, matrix attachment regions (MARs), was found to be in comparable positions in the two species, often flanking individual genes. If these MARs define structural domains, then the orthologous genes in maize and sorghum should experience similar chromatin environments. In addition, MARs were divided into two groups, based on the competitive affinity of their association with the matrix. The “durable” MARs retained matrix associations at the highest concentrations of competitor DNA. Most of the durable MARs mapped outside genes, defining the borders of putative chromatin loops. The “unstable” MARs lost their association with the matrix under similar competitor conditions and mapped mainly within introns. These results suggest that MARs possess both domain-defining and regulatory roles. Miniature inverted repeat transposable elements (MITEs) often were found on the same fragments as the MARs. Our studies showed that many MITEs can bind to isolated nuclear matrices, suggesting that MITEs may function as MARs in vivo.

INTRODUCTION

Our recent results have illustrated that genes residing in the maize alcohol dehydrogenase (adh1) region display two distribution patterns: individual genes amid a sea of repetitive DNA and clustered genes occupying a space uninterrupted by highly repetitive DNAs (Tikhonov et al., 1999). Sorghum, a close relative of maize with an ~3.5-fold smaller genome (Laurie and Bennett, 1985), has orthologous genes located in an environment that lacks highly repeated DNAs (Tikhonov et al., 1999). Similarly, the maize shrunken2/a1 genes are located ~140 kb apart (Civardi et al., 1994), separated by blocks of repeats of retroelements (P. SanMiguel, unpublished observations), whereas no large repeats were found in the orthologous regions of sorghum and rice (Chen and Bennetzen, 1996; Chen et al., 1997). Given the established importance of the genomic context for the function of a gene, best illustrated by the phenomenon known as position-effect variegation, these findings point to an apparent paradox: orthologous genes in related species may show similar functions in very different chromosomal settings.

However, all results describing the effects of the surrounding DNA on a gene’s behavior have been obtained either in transgenic systems or with translocation events in which tested genes are placed in an environment that is different from their wild-type location. Therefore, the chromosomal context and its putative effect on a gene’s function in its natural environment have not been assessed. The existence of functional genes among blocks of retroelement repeats indicates that natural systems must have evolved mechanisms and strategies that would allow orthologous genes to function at their respective, evolutionarily established locations, despite the very different natures of the neighboring sequences.

A current model accounting for this apparent paradox suggests that genes and blocks of repetitive DNAs might exist in different, structurally separated nuclear compartments (Henikoff and Comai, 1998; Lamond and Earnshaw, 1998; Cockell and Gasser, 1999) and that mislocation alters expression. Each gene in the nucleus may have only one “address” at which it functions correctly, and during evolution, natural genes may have acquired “anchors” to stably position them in the spatial architecture of the nucleus (Flavell, 1994).

A specific class of DNA sequences, matrix attachment regions (MARs), may be involved in this anchoring function. MARs are operationally defined as DNA sequences that preferentially bind to the proteins of the nuclear matrix, potentially defining and delimiting individual structural units of chromatin (reviewed in Bode et al., 1996). One step in testing such a model therefore would be to screen for the distribution of MARs along large chromosomal continuums. MAR location would suggest a pattern for a putative folding of the
region into individual structural domains. Defining how biologically active units are positioned within these domains would be an obvious next step.

In an earlier study, we reported that the long stretches of mixed classes of repetitive DNAs in the maize adh1 region were segregated into topologically sequestered units (Avramova et al., 1995). The lack of DNA sequence data, however, precluded more detailed comparisons between the possible structural organization of the region and gene-genome function. Subsequent studies of the region (Avramova et al., 1996; SanMiguel et al., 1996; Tikhonov et al., 1999) provided a molecular basis for renewing our pursuit of a possible relationship between gene structure and function. Although some limitations of our analysis and screening approach, at that time, have been acknowledged and discussed (Avramova et al., 1995), most of them have now been overcome, as described below.

Recently, we were able to demonstrate that microcolinear- ity of gene composition between maize and rice was paralleled by genomic folding into colinear structural units. More importantly, the pattern of potential structural organization was followed by similar placement of the genes within the putative loops, suggesting that the positioning of the genes in their respective structural domains might be conserved in evolution (Avramova et al., 1998). However, the colinear shrunken2/a1-homologous regions of both rice and sorghum lack large amounts of repetitive DNA and hence are in gene-rich environments (one gene per 8 to 10 kb). The colinear maize and sorghum adh regions differ greatly in the amount of intervening repetitive DNAs and resulting gene richness (one gene per 25 kb in maize versus one gene per 6 kb in sorghum; Tikhonov et al., 1999).

Here, we studied a possible structural folding of large colinear segments carrying orthologous genes in different genomic environments by using the distribution of MARs along the regions as genomic elements that may define structural units. The placement of candidate genes within the putative structural domains supports a model in which plant genes have evolved within structural domains separated from neighboring nongenic or genic sequences.

RESULTS

Comprehensive Analysis of MARs in the Maize adh1 Region

The screening for MARs was conducted by using an approach that tests the capacity of isolated nuclear matrices to interact with end-labeled tested DNA fragments in the presence of excess unlabeled competitor DNA. In our previous study, due to the absence of sequence information, MARs were positioned on the basis of the available map created by a set of three restriction endonucleases (Springer et al., 1994). As a result, MARs could have been destroyed by the digestion (Avramova et al., 1995) or missed because of too short or too long fragment lengths (Izaurralde et al., 1988; Mielke et al., 1990). Recent data on the sequence composition of this region (Tikhonov et al., 1999) allowed MAR screening after digestions with sets of carefully chosen restriction endonucleases (see Methods). This approach, however, generated a very large number of fragments, and we have chosen to reexamine only the low-copy-number regions of the maize contig under new binding conditions.

Multiple overlapping fragments from the low-copy-number regions were generated and studied. The shortest fragments still displaying binding activities were defined as MARs. This approach made it possible to identify individual attachment points within larger matrix binding regions. Each identified MAR was characterized under several different competitive concentrations and on several overlapping fragments. These multiple reexaminations allowed independent confirmation of the presence of a MAR and its characteristics. Moreover, these experiments also permitted a more precise definition of MAR location and complexity. Knowing that the largest low-copy-number domain from the maize adh1 contig was segregated in a putative loop of ~45 kb (Avramova et al., 1995), we wished to test the possible structural organization within this domain, particularly with regard to several candidate genes (Tikhonov et al., 1999). A goal of this study was to compare the possible chromatin organization of orthologous genes in colinear genomic regions of two species differing in the amount and nature of intervening DNAs. Hence, we have focused on the possible structural organization of the putative gene-containing regions.

In our earlier study, MAR screening was conducted under stringent binding conditions (Avramova et al., 1995). High concentrations of competitor DNA and nonsaturating quantities of matrices in the assays allowed only fragments with the highest matrix affinity to be identified. Here, binding assays were performed with saturating amounts of matrix proteins. These have been established empirically by testing increasing amounts of isolated matrices, until a point was reached at which all of a labeled MAR fragment was matrix bound. The concentration of matrix proteins at which all of the labeled MAR fragments were recovered from the pellet, with none being found in the supernatant, was defined as saturating. It corresponded, approximately, to the residual protein obtained after high-salt extraction of 1 A260 OD of starting nuclei (see Methods). A higher concentration of matrix proteins provides more binding sites in the system, allowing for the discovery of weaker but specific MAR-containing fragments (Mielke et al., 1990).

Matrices isolated from the nuclei of sorghum seedlings were used. Matrix proteins are evolutionarily conserved such that they recognize and bind MARs across species in test assays conducted in vitro (Cockerill and Garrard, 1986b; Avramova and Bennetzen, 1993). Here, matrix preparations from sorghum seedlings were found to bind maize MAR-containing fragments with the same specificity shown earlier for matrices isolated from maize leaves and roots.
Also, a MAR flanking a tobacco root-specific gene, RB7 (Hall et al., 1991), bound with very high specificity to the matrix isolated from nuclei of sorghum seedlings (Figure 1C), supporting the idea that matrices from one species can be used to discover MARs in the genome of another.

Finally, three different concentrations of competitor DNA were tested. The results allowed us to evaluate both the relative binding strength of individual MARs and to follow the profile of their binding to the matrix under competition (Figures 2 and 3). Two classes of MARs have been tentatively discriminated: “durable” MARs and “unstable” MARs.

Two Classes of MARs

Under the binding conditions described, many fragments from the tested clones bound specifically to the nuclear matrices. This created ambiguities for defining possible loops. We looked for criteria that would allow us to distinguish between different types of matrix binding DNAs. Usually, MARs are arbitrarily defined as strong and weak, based on their different affinities for the nuclear matrix in the presence of an excess of nonspecific DNA.

Following this criterion, we quantified the amount of remaining matrix-bound DNA after competition with various concentrations of nonspecific DNA. On the basis of the density-profile peaks of a labeled fragment recovered from the pellet (MAR-containing fragment), its amount was calculated as a percentage of the density of each fragment present in the input fraction, taken as 100% (Figure 2; see Methods). Restriction fragments recovered from the pellet were quantitatively analyzed for their matrix binding affinity in the presence of 100 to 400 μg/mL competitor DNA. These correspond to 5000- to 20,000-fold excesses of Escherichia coli DNA over the substrate genomic DNA. Approximately 270 DNA fragments were analyzed and 36 MARs were identified along the two contigs.
After plotting the amount of recovered matrix-bound fraction over the concentration of competitor DNA, we were able to follow the changes in binding of the different MARs. Analyzing the profiles of the curves, we observed that MARs could be divided into two classes. A common feature of the first class, as illustrated by the representative profiles of several MARs (Figure 3A), is that these MARs continue to bind the matrix in the presence of the highest tested concentrations of competitor DNA. Some of these MARs, if considered at only one competitor DNA concentration (even the lowest, i.e., a 5000-fold excess), could have been classified as weak and possibly dismissed from further analysis. However, these MARs still could bind the matrix (15 to 30% recovered from the pellet) under higher (up to 20,000-fold) excesses of competitor DNA (Figure 3A, curve 5). This class represents the durable MARs, and 16 MAR fragments were found to belong to this category. All but one of them mapped in the space outside genes.

The MARs in the second class exhibit different profiles. The matrix binding curves drop steeply under increasing concentrations of competitor DNA (Figure 3B). Although some of the MARs in this class could qualify as strong at lower excesses of E. coli DNA (e.g., Figure 3B, curves 1 to 3), their binding to the matrix was blocked by higher concentrations of competitor DNA. These less competitive, unstable MARs mapped to regions inside genes and, predominantly, within putative introns (Tikhonov et al., 1999).

A notable exception is the dissociation curve of a MAR located internal to maize gene 334B7.4; this MAR displays a durable profile (Figure 3C). It is discussed in more detail later in the article.

Orthologous adh Genes of Maize and Sorghum and Their Structural Domains

The maize adh1 gene is an example of a lone gene amid large blocks of highly repetitive elements. Previous studies have localized a MAR 5’ to the gene (Avramova and Bennetzen, 1993; Paul and Ferl, 1993) and two MARs at its 3’ end (Avramova et al., 1995). The precise location of the weaker MAR closer to the 3’ end, however, has remained unclear. Thus, the size of the putative adh1-containing loop could vary from 3 to 9 kb, depending on the position of the MAR inside a 7-kb restriction fragment. Also, a small MAR-containing fragment (~500 bp) from the region was not positioned on the previously available map (Avramova et al., 1995).

Here, the 5’ and 3’ regions flanking maize adh1 have been examined in detail after being subcloned and digested with a combination of restriction endonucleases. The recovery of the tested fragments from the pellet (MARs) and the supernatant (nonbound DNA) is shown (Figure 1B). The gradual increase in the percentage of the fragment in the supernatant, as a result of increasing concentration of competitor, illustrates the different affinities for the matrix proteins displayed by the various MAR-containing and non-MAR fragments.

Analysis of the matrix-bound fragments indicated that they were adjacent on the contig, probably constituting multiple attachment points. A model for the structure of a putative (~4 kb) adh1 domain, bordered by the two flanking MARs, is shown (Figure 4). Thus, both MARs surrounding the adh1 gene appear to be composite and durable (Figure 3A, curves 1 and 5) and to be tightly bordering the long terminal repeats of the flanking retroelements (CinfuI-2 and MilI). These complex MARs are positioned to precisely separate the gene from the neighboring 70-kb (5’) and 14-kb (3’) blocks of highly repetitive DNAs.

It was of interest to determine the structure of the region around the orthologous sorghum adh gene because the regions occupied by MARs in maize did not exhibit obvious sequence homology with sorghum and because there are no large repetitive DNA blocks around the sorghum adh gene (Tikhonov et al., 1999). A 13-kb region containing the...
sorghum adh gene was subcloned and digested with two different sets of restriction endonucleases, and the overlapping fragments were tested for binding activity. The results from the binding of two different sets of fragments are shown (Figure 1A). Both 5' and 3' MARs flanking the sorghum adh1 homolog were found to be durable (data not shown).

Analysis of the matrix-bound fragments suggested a potential structural organization of the sorghum adh gene-containing domain (Figure 4). The similarity in the sizes of the putative loops, defined by the positioning of the matrix binding regions, is quite remarkable, given the absence of MAR sequence homology. In the different chromosomal contexts in which the orthologous genes exist in the two species, the preservation of the structural organization of the gene may be relevant to its expression.

Putative Genes and Putative Domains along the Maize Contig

The possible chromatin-loop arrangements of the gene candidates recently identified on the maize contig (Tikhonov et al., 1999) were examined. Four genes, 334B7.2, 334B7.3, 334B7.4, and 334B7.9, are individual loci, surrounded by blocks of retroelement DNAs. Screening for MARs has identified MARs around the genes, suggesting structural organization similar to the 334B7.3 (adh1) gene. Putative gene 334B7.2, discovered at one end of the contig, is separated from its nearest neighbor, adh1, by ~70 kb of retroelements (Figure 5). A MAR marks one end of its putative structural loop, whereas the other end was not defined because it was outside the contig. Downstream of the adh1-containing loop, after an ~14-kb block of highly repeated retroelements, another solitary putative gene, 334B7.4, was discovered amid repetitive DNAs (Tikhonov et al., 1999). In this region, both a newly found and a previously discovered MAR (fragment 47 in Avramova et al., 1995) mapped inside the gene (Figure 5). One of the internal MARs, overlapping putative exon sequences of the gene, displayed a matrix-affinity profile relating it to the durable, possibly loop-defining MARs (Figure 3C). The curve of the other was somewhat intermediate (see Discussion). The closest MARs located outside the 334B7.4 gene sequence were, on the one side, the MAR from the 3' end region of adh1 and, on the other, two MARs on adjacent fragments (see below). These would define a putative loop of ~60 kb. This predicted loop is unusual because it is composed of various highly repetitive retroelement DNAs.

Figure 3. Competition Profiles for Various MARs.

(A) Matrix associations of durable MAR fragments under three different concentrations of E. coli competitor DNA: 100, 200, 300, or 400 µg/mL. Representative graphs are shown for MARs from both maize and sorghum: curve 1, a 1.2-kb MAR fragment 5' of the maize adh1 gene; curve 2, a 245-bp MAR fragment between the sorghum 110K5.10 and 110K5.11 genes; curve 3, a 2.7-kb MAR fragment between the maize 334B7.7 and 334B7.8 genes; curve 4, a 160-bp MAR between the sorghum 110K5.8 and 110K5.9 genes; curve 5, a 2.5-kb MAR 3' of the adh1 gene; and curve 6v, a control vector fragment.

(B) Matrix associations of intronic MARs under similar binding conditions. Curves 1 and 2 are 2.3- and 2.2-kb MAR fragments, respectively, located in introns of the sorghum 110K5.4 gene; curve 3, a 440-bp MAR fragment inside the sorghum 110K5.11 gene; curve 4, a 2.3-kb MAR fragment in the last intron of the maize 334B7.7 gene; and curve 5v, a control vector fragment.

(C) Matrix affinity of a MAR located inside the maize 334B7.4 gene. Curve 1, a 1.4-kb MAR fragment found inside a putative exon sequence (this fragment also contains a MITE); and curve 2v, a control vector fragment.
retrotransposons and a gene. The maize 334B7.4 gene had been considered, hitherto, as an example of a plant gene “broken” by the insertion of three large blocks of highly repetitive elements, scattering the genomic sequence over 42 kb of genomic space (Tikhonov et al., 1999). Here, we show that 334B7.4 may have an unusual domain organization as well.

The last gene from the maize contig surrounded by repetitive DNAs is putative gene 334B7.9. However, its apparent loop also includes remnants of two members of retroelement families, Cinful-1 and Kake-1 (Figure 5). The possible loop-defining MARs of this (~10 kb) domain have been established previously (Avramova et al., 1995) and have not been restested in this study.

The remaining four genes, 334B7.5, 334B7.6, 334B7.7, and 334B7.8, are organized in a cluster, uninterrupted by repetitive retroelement DNA (Tikhonov et al., 1999). Multiple overlapping fragments from the region were generated and tested for matrix binding. The pattern of MAR distribution was rather complex, and the durable MARs were considered domain defining.

The first gene of the cluster, 334B7.5, is positioned in a putative 5-kb loop. The upstream binding region seems to contain at least two attachment points that, together, separate the low-copy-number region from the highly repetitive DNA block upstream. Each of these two attachment fragments behaved as weak MARs, but, nonetheless, they bound the matrix under the stringent assay conditions of the earlier study (Avramova et al., 1995) and displayed durable MAR profiles under the highest competition conditions of this assay (not shown).

The MAR closing the loop containing the 334B7.5 gene and defining the beginning of the adjacent loop was found positioned very near the 3’ end of the neighboring 334B7.6 gene. The other end was located ~20 kb downstream, where a durable MAR was found between the 334B7.7 and 334B7.8 genes (Figure 3A, curve 3). An unusual feature of this putative structural loop is that it encompasses two genes, because no MAR was found in the space separating the 334B7.6 and 334B7.7 loci. These two putative genes, apparently transcribed in opposite directions, were found as an unusual example of closely positioned maize genes that are not gene duplication products (Tikhonov et al., 1999). They are separated by only 200 bp of DNA and apparently share the same structural domain. Gene 334B7.7 contains several unusually large introns and two internal MARs (Figure 3B, curve 4). The last gene from the cluster, 334B7.8, is positioned in a putative loop defined by two durable MARs. An unstable MAR was found inside a putative intron (data not shown).

### Putative Folding of the Colinear Region of Sorghum

As a result of the absence of retrotransposons from the homologous region of sorghum, the genes are more densely arranged, with an average density of one gene per 6 kb. An unexpected observation was that three putative genes, in the otherwise colinear sorghum region, were missing in maize (Tikhonov et al., 1999). The differences in the genomic structures, at a sequence level, prompted us to compare the possible domain-folding patterns of the two regions. The distribution of the MARs along the 78-kb sorghum contig is linear with the 225-kb maize contig was examined. Each of seven subclones was digested with combinations of restriction nucleases (see Methods), and overlapping fragments were tested for binding capacity under the same conditions as described for maize (Figure 6).

The distribution of the intergenic MARs along the sorghum contig outlines a pattern, placing most of the genes in apparently well-defined putative loops. Notable exceptions are two genes, 110K5.9 and 110K5.10, that appear to share the same loop. The ends of this putative structural domain are defined by durable MARs positioned between 110K5.8 and 110K5.9 (Figure 3A, curve 4) and between 110K5.10 and 110K5.11 at the other end (Figure 3A, curve 2). The region separating putative gene 110K5.9 from 110K5.10 displayed no matrix binding potential. Putative intronic MARs were identified in both 110K5.10 (not shown) and 110K5.11 (Figure 3B, curve 3).
MARs and Miniature Inverted Repeat Transposable Elements

Based on the molecular analysis of two different orthologous chromosome segments, the shrunken2/a1 regions of rice and sorghum and the adh1 regions of maize and sorghum, ~80 miniature inverted repeat transposable elements (MITEs; reviewed in Wessler et al., 1995) were identified along the four genomic continuums (Chen and Bennetzen, 1996; Chen et al., 1997; Tikhonov et al., 1999). Screening the regions for MARs revealed ~40 restriction fragments with matrix affinity that also contained a MITE, raising the question of whether MITEs carry the MAR function and/or whether MITEs have inserted preferentially around MARs.

To study this question, we tested the capacity of isolated MITEs to bind to the matrix. Approximately 20 elements from the four genomic regions were examined under the binding conditions described in this study. The MITEs, in all tested cases, displayed matrix binding activity (Figure 7).

In approximately a dozen cases, the presence of convenient restriction sites allowed the isolation of MITEs as separate recognizable fragments from the surrounding genomic sequences. The binding profiles of a few examples are shown in Figures 7A to 7E.

Alternative digestions with restriction nucleases of the region containing two nested Tourist elements 3' to the sorghum 110K5.3 gene (Figure 6) released either the whole set (Figure 7A) or an individual element (Figure 7B; for detailed nucleotide information on the structure and location of the elements and the position of the restriction sites, see the legend to Figure 7 as well as GenBank accession number AF124945 and Tikhonov et al. [1999]). The matrix binding tests showed that both nested and individual Tourist elements can bind the matrix (Figures 7A and 7B). In accordance, an individual Tourist near the 5' end of the 110K5.5 gene from the same contig (Figure 6) binds also to the matrix (Figure 7C, marked by the arrow). The higher molecular weight bands belong to adjacent genomic sequences not discussed here.

The matrix binding of two MITEs of maize origin are shown as well: the MITE from the putative exon of maize gene 334B7.4 (Figure 7D) and the Tourist-Zm1 element (Bureau and Wessler, 1992; Figure 7E). The latter element is very short, ~130 bp, which may be a reason for its apparent weaker binding. It is possible that it is a component of a larger binding region at its genomic location.

Conveniently located restriction sites, however, often would leave flanking genomic sequences or eliminate sequences from the tested elements, which could impact the MAR activity of the tested fragments. To avoid this possibility, we examined several MITEs after PCR amplification of their sequences. The 894-bp fragment containing three complex MITEs, between maize genes 334B7.7 and 334B7.8 (Figure 5), has been PCR amplified using primers tightly flanking the whole set. The primers and their genomic positions are shown in Methods and provided under GenBank accession number AF123525. The matrix binding capacity of the whole fragment is shown in Figure 7G (arrow 1).

The potential involvement of individual elements in the

---

**Figure 5.** MARs and Predicted Loops in the Maize adh1 Region.

Maize YAC 334B7 is represented by a bar with open boxes (genic DNA). Putative exons are shown as vertical boxes. Diamonds and filled arrowheads designate MITEs and long interspersed nuclear element-like retroelements, respectively. The thin lines correspond to regions occupied by repetitive retrotransposon DNAs. Putative chromatin loops are shown by curves above the boxes. The anchors and the asterisks indicate the positions of domain-defining and intragenic MARs, respectively. The boxes below the YAC designate the restriction enzyme fragments tested in the binding assays; multiple overlapping fragments were tested, and the MAR signs were placed above the shortest restriction fragments recovered from the matrix-bound fraction. The scale is in kilobases. The numbers above the boxes correspond to the putative genes as defined in Tikhonov et al. (1999) and as referred to in the text.
matrix binding interaction is illustrated for two adjacent Tourist elements revealed in the sorghum shrunken2/a1 region, close to the 3’ end of the shrunken2 gene homolog (Chen et al., 1997; GenBank accession number AF010283). Their structure, the location of the various primers used, and the amplified regions are illustrated in Figures 8A and 8B. The binding capacity of the individually amplified fragments are shown in Figures 7F and 7G. Bands 2 and 4, corresponding to the individual MITEs, can bind the matrix; an amplified fragment containing only the terminal repeats at the junction of the two elements (band 5) does not bind to the matrix. The whole fragment (band 1) binds as well, although with an apparent lower affinity. The reason for this is not clear at present, although one possibility is the lower overall outcome of this particular labeling (as determined by the intensity of the input fragment). Nonetheless, this apparently weaker fragment retained persistent matrix binding capacity when tested under three different competition concentrations. Its matrix-affinity curve as well as the binding curves of all examined MITEs with MAR activity place them in the class of the durable MARs (data not shown).

DISCUSSION

Domain Organization along the Maize Contig

Comprehensive analysis of MAR distribution in the low-copy-number regions of the maize contig provided an insight into the potential of these regions to fold into putative structural loops. The general arrangement along the maize contig shows genes occupying individual loops that are separated from neighboring retroelement repeats. Three exceptions have been observed: the 334B7.6 and 334B7.7 genes, which are positioned in the same loop (Figures 5 and 9), and two genes that apparently cohabit with retroelements inside their structural domains.

The genic sequence of the maize 334B7.4 gene has been scattered over 42 kb of genomic space as a result of the insertion of three large blocks of retroelements (Tikhonov et al., 1999). The closest external MARs would delimit a putative structural domain of ~60 kb (Figure 5). However, a MAR located inside the gene displayed matrix-affinity characteristics of the durable MARs (Figure 3C). This could indicate that structural attachment point(s) might exist inside the genic sequence. Thus, this region might be involved in a structural role, particularly if the gene is not functional anymore, as suggested by the location of a MAR and a MITE inside an exon (Tikhonov et al., 1999). In the absence of expression data regarding the functional capacity of the 334B7.4 gene, however, it is not possible to decide whether it is located within a large structural domain, flanked by and interrupted by retroelements, or whether it serves as an attachment structure itself.

The only other gene in the studied region sharing a loop with retrotransposable elements is the putative 334B7.9 gene (Figure 9, loop ‘)’. There are no data regarding whether this is a functional gene, but both 334B7.4 and 334B7.9 genes provide attractive systems to further pursue a relationship between genes’ structural organization and their expression.

Domain Structures in Sorghum

In the 78-kb colinear region of sorghum, 18 MARs were identified. Most of the genes, with the apparent exception of genes 110K5.9 and 110K5.10, seem to be positioned in individual loops (Figures 6 and 9). Seven of the MARs colocalize with fragments containing MITEs. It is interesting that a fragment in the space between genes 110K5.9 and 110K5.10 did not bind to the matrix, despite the fact that it contained...
a MITE; thus, the two genes apparently share the same domain.

**Comparison of the Predicted Domain Structures in Maize and Sorghum**

The degree of similarity between the orthologous adh-containing regions in maize and sorghum has been reported at the DNA sequence level (Tikhonov et al., 1999). The possible folding into structural domains (loops) of these colinear continuums has been assessed here. Despite the presence of more genes (nine and 14 putative genes on the maize and sorghum contigs, respectively), the sorghum contig contains approximately three times less DNA than does the orthologous maize region. Of the 225 kb surrounding maize adh1, 160 kb is occupied by highly repetitive DNAs represented by 23 long terminal repeat retrotransposons, organized in six blocks of various sizes (Tikhonov et al., 1999). To be able to compare the gene-containing structures of the two regions, we have graphically “subtracted” the retroelement blocks from the maize contig (Figure 9).

It is shown that the gene-containing domains and the sizes of the putative loops containing orthologous genes in the two species are comparable. This fact suggests that the structural folding of the homologous genomic regions, excluding the blocks of highly repetitive DNA, might be very similar. Thus, despite an enormous difference in the sizes of the loops occupied by the two orthologous genes, 110K5.4 and 334B7.8 (Tikhonov et al., 1999), the individual MITEs are shown by the boxes, and the terminal inverted repeats as blank arrows. The position of the primers is shown by the black arrowheads. The numbers at the primers correspond to the genomic location of the respective sequence. The PCR products are shown by the numbered lines under the boxes, corresponding to the numbering of the binding fragments in Figures 7F and 7G. The cross-hatched box inside the empty box illustrates a MITE nested within another MITE.

**Figure 7.** In Vitro Nuclear Matrix Binding Assays of MITEs.

Fragments corresponding to MITEs are shown by arrows in (A) to (G). Open arrowheads denote fragments from the vector; m, l, and P denote lanes containing size-marker fragments, input fragments, and fragments recovered from the pellet, respectively.

(A) Clone p13MC (Table 1), digested with PmlI, SnaBl, and Scal, released nested Tourist elements located 3’ to the sorghum 110K5.3 gene.

(B) Alternative digestion of the same clone with Stul alone, showing the binding activity of one of the Tourist elements.

(C) Clone p17 (Table 1) after digestion with NotI and Stul, generating a solo Tourist at the 5’ of the sorghum 110K5.5 gene.

(D) Maize clone p95d (Table 1) after digestion with AflIII and Stul, releasing the MITE inside the putative exon of maize gene 334B7.4.

(E) Clone pB2wx carries the 128-bp Tourist-Zm1 element (Bureau and Wessler, 1992; GenBank accession number S48688). Digestion with EcoRI released a 184-bp fragment; its binding is shown by the arrow.

(F) and (G) PCR-amplified MITEs. Fragment 1 contains three MITEs (two nested and a closely adjacent one) from the spacer region between maize genes 334B7.7 and 334B7.8 (Tikhonov et al., 1999). The individual MITEs are shown by the boxes, and the terminal inverted repeats as blank arrows. The position of the primers is shown by the black arrowheads. The numbers at the primers correspond to the genomic location of the respective sequence. The PCR products are shown by the numbered lines under the boxes, corresponding to the numbering of the binding fragments in Figures 7F and 7G. The cross-hatched box inside the empty box illustrates a MITE nested within another MITE.
maize region. These regions differ in their domain organization as well: 110K5.8 is predicted to be in an individual small loop (Figure 9, loop h) on the sorghum contig, whereas 110K5.9 shares a putative loop with 110K5.10 (Figure 9, loop i to j). The orthologous maize gene, 334B7.7, also is positioned in a loop as a pair. However, because the two corresponding neighboring genes 5’ to 334B7.7 are missing in maize, its partner in the loop is its current 5’ neighbor, gene 334B7.6 (Figure 9, loop g’ to j’). Therefore, sharing a structural domain with a closely positioned 5’ neighboring gene seems to be a preserved feature for the two orthologous genes, 334B7.7 and 110K5.10. The possible significance of such an arrangement is unclear, but it points again to a possible relationship between the positioning of a gene within a structural unit and its functional requirements.

The apparent tendency of high-copy-number retroelements to insert into other retroelements, rather than into genic sequences, was reported by SanMiguel and co-workers (SanMiguel et al., 1996). Here, we noted that the insertion sites in the maize genome of the initial retroelements (those at the base of the stacks [SanMiguel et al., 1996]) map right at, or in a very close proximity to, MARs that flank genes. This observation suggests that MARs may act both as potential target sites and as barriers for the genes against invasion by transposable elements.

**Intronic MARs**

It has been reported that under conditions of nonsaturating amounts of matrix proteins (i.e., binding sites), only DNAs with the highest affinity would exhibit binding. Equilibrium binding studies, at saturating levels of matrix proteins, have revealed that the process is cooperative and reversible (Mielke et al., 1990). Competitor DNA, added to prevent nonspecific binding, is used to measure the matrix affinity of a MAR-containing fragment. On this basis, MARs have been arbitrarily defined as strong and weak, and it has been suggested that they may have different functions (Allen et al., 1993; Bode et al., 1996). However, the criteria for defining strong and weak MARs have not been unified.

The detection of MARs in the introns of several genes (Cockerill and Garrard, 1986a; Mielke et al., 1990; Avramova and Paneva, 1992; Romig et al., 1994) pointed to yet another feature that divides MARs into different categories. Intronic MARs, by definition, cannot block progression by polymerase II, implying either that they do not function in vivo or that their binding to the matrix is regulated. In any case, intronic MARs are not considered to be defining structural domains (Bode et al., 1996). An important question, therefore, is what makes the two types of MARs different.

It has been shown that a simple classification of MARs as strong and weak would not discriminate between intronic and external MAR elements (Romig et al., 1994). We therefore chose to examine the various MARs for their affinity to the nuclear matrix, under increasing concentrations of competitor DNA, instead of making a one-point strong/weak comparison. Tentatively, the tested MARs were divided into two classes on the basis of their binding patterns. One group contains elements that could bind the matrix in the presence of the highest competitor concentrations. These MARs were termed durable, and they were usually found
outside the genes. The second group, unstable MARs, were competed off their matrix binding sites under similar conditions. Most of the intronic MARs were found in this class. It should be emphasized that the difference in the competition curves is not regarded as an absolute criterion separating the two classes of MARs. Some of the tested MAR fragments, particularly longer restriction fragments and fragments encompassing internal and external genetic sequences, exhibited curves intermediate to the two profiles, indicating a continuum of MAR affinities. The continuous series of MAR binding behavior suggest that various components collaborate and contribute to the interactions. Therefore, the difference in the curves suggests that intronic and external MARs might differ in the nature of their association contacts with the nuclear matrix. Persistent binding activity might be more relevant for MARs with a structural function, whereas a capacity for an easier dissociation of even strong MARs might be a prerequisite for a controlled association with the matrix.

Intronic MARs were found in several sorghum and maize genes. It is interesting that orthologous loci in maize/sorghum (334B7.4/110K5.4, 334B7.7/110K5.10, and 3334B7.8/110K5.11) have preserved MARs in their respective introns. These MAR-containing introns also display extensive sequence conservation (Tikhonov et al., 1999), distinguishing them from the external MARs.

Little is known about the role, if any, of intronic MARs. It has been suggested that MARs, particularly those in introns, may act as a sink for absorbing the positive supercoils generated in front of a progressing polymerase transcription complex, increasing the efficiency of transcript elongation by removing a geometric impediment to progress (Benham et al., 1997). Other data show an involvement of the intronic κ-immunoglobulin MAR in the regulated demethylation of the gene upon activation (Kirillov et al., 1996) and in somatic hypermutation (Betz et al., 1994). An intronic MAR from a potato gene, when placed between the gene encoding human interferon β and neo, caused activation of transcription in transfected mouse L cells (Mielke et al., 1990). The findings that intronic MARs are preserved in homologous genes (Avramova and Paneva, 1992) and in orthologous introns in maize and sorghum (see earlier text) support the idea of a conserved function, possibly in gene regulation. Specific MAR binding soluble factors might be involved in mediating a regulated binding of intronic MARs to the nuclear matrix (Boulikas and Kong, 1993; Bode et al., 1996).

**MARs and MITEs**

Inverted repeats are cruciform-forming sequences (Schroth and Ho, 1995), and MARs are often enriched in inverted repeats (Boulikas and Kong, 1993). Plant genomes harbor short interspersed repeats, MITEs, which have terminal inverted repeats and have a potential to form secondary stem–loop structures (Bureau and Wessler et al., 1994; Charrier et al., 1999). Nineteen of the 36 MARs discovered in the adh-homologous genomic regions of the two grasses were found to colocalize on the same fragments with MITEs. Frequent close associations between MITEs and MARs in the shrunkened2/a1-homologous regions of rice and sorghum have been noticed as well (Avramova et al., 1998), suggesting a possible correlation between the potential secondary DNA structure of these elements and MAR function.

Hence, the capacity of isolated MITEs to bind to the matrix was tested. Approximately 20 elements contained within MAR fragments from the four genomic regions were examined. The MITEs were obtained either by excision with restriction nucleases or by PCR amplification, and in all tested cases, the MITEs displayed matrix binding activity. In several cases, MITEs were found in a nested arrangement or immediately adjacent to each other, and our results showed that the set as a whole and its individual components display a capacity for matrix binding. This had been observed earlier for the maize s’ adh1 MAR, before recognizing that this region was made of two nested Tourist elements. Each of these elements bound to the matrix, although with reduced activity compared with the nested structure (Avramova and Bennetzen, 1993).

The apparent involvement of these mobile elements in a genome structural function concerns the basic question of whether transposable elements and repetitive DNAs may play any role in host genomes. As is also true of the MARs associated with some members of 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 element might be selected for new local functions, such as gene regulation, recombinational initiation (Galliano et al., 1995), telomere structure (Danilevskaya et al., 1994; Eickbush, 1997), or MAR function.

**Loop Organization of Plant Genes**

A major consequence of the model suggesting that chromatin in the nucleus is folded into discrete structural domains (loops) is the attractive possibility that each domain constitutes an independent unit of genetic function. Several classes of structural elements have been suggested to play this role (reviewed in Geyer, 1997), but there is controversy regarding the chromatin-boundary function for some of them (Avramova and Tikhonov, 1999).

MARs are good candidates for playing a structural, domain-defining role, and their distribution in eukaryotic genomes typically places them at the borders of gene domains. Adjacent MARs commonly outline potential loops of 3 to 400 kb (Boulikas and Kong, 1993), exhibiting a general inverse relationship between domain size and gene activity (reviewed in Bode et al., 1996). In plants, relatively small (3 to 10 kb) loops seem to represent a common pattern for the organization of plant genes (Breyne et al., 1992; van der Geest et al., 1994; Chinn and Comai, 1996; Avramova et al., 1998). In a 16-kb region of the small Arabidopsis
genome, for example, loops averaging ~5 kb were found to harbor two genes each (van Drunen et al., 1997).

Recently, it was reported that the maize adh1 gene is located within a predicted 90-kb loop, based on the size of a genomic fragment obtained by preferential cleavage at superhypersensitive sites (Paul and Ferl, 1998). Although it is not possible to interpret this result in terms of a structural loop, it localizes the closest nuclease-susceptible sites to the adh1 gene. Our sequence data indicate that ~90 kb is the approximate distance between the promoter of adh1 gene and the closest flanking promoter of maize gene 334B7.5, suggesting that the nuclease hypersensitive sites map at potentially active promoter regions that often colocalize with MARs (reviewed in Bode et al., 1996).

A difficult unresolved matter is the relationship between genome structure and regulated gene expression. An approach providing information on a possible link between chromatin folding into structural units and the organization of the genes inside has been reported here. From this perspective, MARs have been analyzed and considered as structural elements only outlining putative domains. The comparative analysis enabled us to observe preservation of common structural features of gene organization, despite the very different nature of the surrounding sequences. An important corollary is that if MARs define structural do-
Marital artificial chromosome (BAC) 110K5 contains ~165 kb of the colinear sorghum region (Woo et al., 1994). YAC 334B7 and BAC 110K5 were subcloned, and the clones analyzed here are shown in Table 1. All 225 kb of the maize YAC and the colinear 78 kb from the sorghum BAC have been screened for matrix attachment region (MAR) distribution, after their complete sequence analysis (Tikhonov et al., 1999). The sizes of the cloned fragments, the cloning sites, and the vectors they were cloned in are shown in Table 1.

The clones were subjected to digestions with various combinations of restriction endonucleases, as shown in Table 2. The purpose was to create multiple overlapping fragments, all of which were tested for matrix binding activity. Their positions on the respective contigs are shown in Figures 5 and 6. Restriction endonucleases and T4 polynucleotide kinase were obtained from New England Biolabs (Beverly, MA) or Promega (Madison, WI). Calf intestinal alkaline phosphatase used for the dephosphorylation of DNA fragments was obtained from Pharmacia (Piscataway, NJ).

**Isolation of Miniature Inverted Repeat Transposable Elements**

The location of the genomic clones containing colocalized miniature inverted repeat transposable elements (MITEs) and MARs is shown in Table 1. Restriction digestion of the clones was conducted as indicated in the legend to Figure 7.

Oligonucleotide primers (summarized in Table 3) were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). Polymerase chain reactions (PCRs) were performed as follows: one cycle at 95°C for 3 min; 35 cycles at 55°C for 30 sec, 72°C for 2 min, 94°C for 45 sec; and one cycle at 75°C for 5 min. Samples were stored at 4°C.

**Table 2. Restriction Enzyme Digestions for Matrix Binding Assays**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Assay 1</th>
<th>Assay 2</th>
<th>Assay 3</th>
<th>Assay 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>p11CM</td>
<td>XbaI, EcoRI, ApaLI, CiaI, MluI</td>
<td>MluI, MfeI, BamHI, NruI</td>
<td>MfeI, EcoRI, XbaI, BsiWI, NcoI, NruI</td>
<td>BsiWI, Clal, MluI, NcoI, PacI, SpeI</td>
</tr>
<tr>
<td>P13MC</td>
<td>MluI, XbaI, Bsu36I</td>
<td>MluI, Clal, ApaLI, SpeI</td>
<td>MluI, SacII, Bsu36I, SmaBI, MfeI, NcoI</td>
<td>MluI, Clal, ApaLI, SpeI</td>
</tr>
<tr>
<td>p5.4</td>
<td>Clal, SacI</td>
<td>Clal, BstXI</td>
<td>Clal, BstXI</td>
<td>Clal, BstXI</td>
</tr>
<tr>
<td>p2</td>
<td>NotI</td>
<td>NotI</td>
<td>NotI</td>
<td>NotI</td>
</tr>
<tr>
<td>p16</td>
<td>HindIII, Eco136, Stul, Clal</td>
<td>ApaLI, Ndel, NotI, Clal</td>
<td>ApaLI, Ndel, NotI, Clal</td>
<td>ApaLI, Ndel, NotI, Clal</td>
</tr>
<tr>
<td>pL243d</td>
<td>Ncol, NotI, XhoI, BamHI</td>
<td>Spel, AphiI, Stul, BsiWI, XhoI</td>
<td>Spel, AphiI, Stul, BsiWI, XhoI</td>
<td>Spel, AphiI, Stul, BsiWI, XhoI</td>
</tr>
<tr>
<td>p52MAR</td>
<td>XhoI, XbaI</td>
<td>BamHI, XhoI, BsrGI</td>
<td>BamHI, XhoI, BsrGI</td>
<td>BamHI, XhoI, BsrGI</td>
</tr>
<tr>
<td>p32MAR1</td>
<td>BamHI, BgIII, EcoRI</td>
<td>BamHI, EcoRI</td>
<td>BamHI, EcoRI</td>
<td>BamHI, EcoRI</td>
</tr>
<tr>
<td>pL95d</td>
<td>Bsu36I, HindIII, CiaI, XhoI, NotI</td>
<td>BsrGI, BsiWI, EcoRI, NotI</td>
<td>BsrGI, BsiWI, EcoRI, NotI</td>
<td>BsrGI, BsiWI, EcoRI, NotI</td>
</tr>
<tr>
<td>pL201</td>
<td>Bsu36I, XhoI, BsrGI, NotI</td>
<td>ApaLI, Avrl, NotI</td>
<td>ApaLI, Avrl, NotI</td>
<td>ApaLI, Avrl, NotI</td>
</tr>
<tr>
<td>pL21</td>
<td>Bsu36I, XhoI, BsrGI, NotI</td>
<td>BgIII, Ncol, Stul, BamHI</td>
<td>BgIII, Ncol, Stul, BamHI</td>
<td>BgIII, Ncol, Stul, BamHI</td>
</tr>
<tr>
<td>p3.3BamHI</td>
<td>AatII, SmaBI, EcoRI</td>
<td>BsrGI, BsiWI, EcoRI, NotI</td>
<td>BsrGI, BsiWI, EcoRI, NotI</td>
<td>BsrGI, BsiWI, EcoRI, NotI</td>
</tr>
<tr>
<td>PB2wx</td>
<td>EcoRI</td>
<td>NotI, Bsp120I</td>
<td>NotI, Bsp120I</td>
<td>NotI, Bsp120I</td>
</tr>
<tr>
<td>pGHNC5</td>
<td>NotI, Bsp120I</td>
<td>NotI, Bsp120I</td>
<td>NotI, Bsp120I</td>
<td>NotI, Bsp120I</td>
</tr>
</tbody>
</table>

*aEnzyme that produced partial digestion.*

**Isolation of Nuclei and Nuclear Matrices**

Nuclei were prepared following the procedure described in Avramova and Bennetzen (1993). Sorghum nuclei were isolated from 15-day-old etiolated seedlings. Ten grams of seedling tissue was frozen in liquid nitrogen and ground in three portions to a fine powder in a mortar. All subsequent steps were performed at 0 to 4°C. Intactness and purity of nuclei were monitored under a light or fluorescent microscope after staining with 4',6-diamidino-2-phenylindole. The nuclear preparation from the final pellet was adjusted to 70% glycerol, aliquoted at 3 A260 OD, and kept at -270°C.

Nuclear matrices were isolated by the 2 M NaCl extraction protocol, as described in Cockerill and Garrard (1986a), with minor modifications. Three A260 units of nuclear preparation were thawed on ice and spun down for 3 min at maximum speed in a microcentrifuge, and the pellet was washed in 1 mL of 10 mM NaCl, 3 mM MgCl2, 10 mM Tris-HCl, pH 7.4, 0.5 mM phenylmethylsulfonfyl fluoride, with the addition of leupeptin and aprotinin (1 mg/mL each), pelleted at maximum speed, and resuspended in 250 mL of the same buffer. DNase I and RNase A were added (5 mg of each), and the nuclei were incubated for 1 hr at 37°C. An equal volume of ice-cold 4 M NaCl was added, and the tube was incubated on ice for 10 to 20 min with two to three inversions of the tube. The nuclear proteins that were insoluble in 2 M NaCl were pelleted by centrifugation at maximum speed for 3 min at room temperature. The supernatant was removed carefully, and the pellet (nuclear matrix) was washed three times, as described previously, in matrix binding buffer (MBB; 10 mM Tris-HCl, pH 7.4, 50 mM NaCl, 20 mM EDTA, 250 mM sucrose, and 0.25 mg/mL BSA).
Finally, matrices from 3 A₂₆₀ units of the nuclear preparation were dispersed in 100 μL of MBB.

Matrix Binding Assay

An aliquot (30 μL) of the matrix preparation described above was used in each binding assay, in a total volume of 50 μL. Sheared (0.3 to 2 kb) competitor Escherichia coli DNA was added to the suspension of nuclear matrices to a final concentration of 50 to 400 μg/mL. Depending on the size of the tested plasmid, 20 to 100 ng of DNA was dephosphorylated and end-labeled with γ-32P-ATP by T4 polynucleotide kinase. Five percent of the labeling reaction was added to the binding assay, and the volume was adjusted to 50 μL with MBB. Microcentrifuge tubes were placed in a horizontal position on a tube rotary shaker and incubated at room temperature. After 5 to 12 hr, the matrices were collected by centrifugation at maximum speed for 3 min, and the supernatants were saved. The pellets were washed by resuspending once in 1 mL of MBB, followed by centrifugation as described earlier, and finally once with 1 mL of Tris-EDTA, pH 8.0. Final pellets were dispersed in 25 μL of 0.5% SDS and 2 mg/mL proteinase K in Tris-EDTA. After incubation at 37°C for 5 to 12 hr, the reactions were incubated at 70°C for 15 min. One-half of each reaction was loaded onto a 0.7 to 1.2% agarose gel in 0.5× Tris-borate-EDTA buffer. After electrophoresis, agarose gels were scanned by densitometer or by exposing them directly to a PhosphorImager plate (Molecular Dynamics, Sunnyvale, CA). Both methods yielded comparable results.

Image Analysis

Exposed x-ray films were analyzed by ImageQuant version 3.3 (Molecular Dynamics, Sunnyvale, CA) and/or Scion Image version 3, a Windows version of NIH Image (Scion Corp., Frederick, MD). The mean gray value of each fragment recovered from the pellet was computed, and the values for the background were subtracted. The mean gray values for the background were determined by measuring a band of the same size from a region adjacent to the respective fragment in the pellet. The mean gray value of each respective fragment from the “input” lanes was used as a reference value, taken as 100%. MAR activity was quantified as a percentage of each matrix-bound fragment relative to its value in the input fraction. These values were then plotted against the step-gradient concentration of the competitor DNA.

ACKNOWLEDGMENTS

The authors are grateful to Phillip SanMiguel (Purdue University) for subcloning the maize YAC and to Steve Spiker (North Carolina State University, Raleigh) and Susan Wessler (University of Georgia, Athens) for providing the clones containing the MAR from the tobacco root-specific gene RB7 and the clone pB2wx carrying the Tourist-Zm1 element, respectively. This research was supported by the U.S. Department of Agriculture Cooperative State Research Education and Extension Service grants No. 98-35300-6167 to Z.V.A. and No. 97-35300-4594 to J.L.B.

REFERENCES


Structural Domains and Matrix Attachment Regions along Colinear Chromosomal Segments of Maize and Sorghum
Alexander P. Tikhonov, Jeffrey L. Bennetzen and Zoya V. Avramova
Plant Cell 2000;12;249-264
DOI 10.1105/tpc.12.2.249

This information is current as of July 11, 2017

References

Permissions
https://www.plantcell.org/cgi/alerts/ctmain

eTOCs
http://www.plantcell.org/cgi/alerts/ctmain

CiteTrack Alerts
http://www.plantcell.org/cgi/alerts/ctmain

Subscription Information
http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY