The Plant Journal (1996) **10**(6), 1163–1168

SHORT COMMUNICATION

Gene identification in a complex chromosomal continuum by local genomic cross-referencing

Zoya Avramova[1], Alexander Tikhonov[1], Phillip SanMiguel[1], Young-Kwan Jin[1], Changnong Liu[1], Sung-Sick Woo[2], Rod A. Wing[2] and Jeffrey L. Bennetzen[1,*]
[1]*Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA, and*
[2]*Crop Biotechnology Center, Texas A&M University, College Station, TX 77843, USA*

Summary

Most higher plants have complex genomes containing large quantities of repetitive DNA interspersed with low-copy-number sequences. Many of these repetitive DNAs are mobile and have homology to RNAs in various cell types. This can make it difficult to identify the genes in a long chromosomal continuum. It was decided to use genic sequence conservation and grass genome co-linearity as tools for gene identification. A bacterial artificial chromosome (BAC) clone containing sorghum genomic DNA was selected using a maize *Adh1* probe. The 165 kb sorghum BAC was tested for hybridization to a set of clones representing the contiguous 280 kb of DNA flanking maize *Adh1*. None of the repetitive maize DNAs hybridized, but most of the low-copy-number sequences did. A low-copy-number sequence that did cross-hybridize was found to be a gene, while one that did not was found to be a low-copy-number retrotransposon that was named *Reina*. Regions of cross-hybridization were co-linear between the two genomes, but closer together in the smaller sorghum genome. These results indicate that local genomic cross-referencing by hybridization of orthologous clones can be an efficient and rapid technique for gene identification and studies of genome organization.

Introduction

Higher plant genomes are composed of interspersed DNAs of various reiteration frequencies (Flavell *et al*., 1974), but neither the composition nor the arrangement of these sequences is well understood. In a 280 kb region surrounding the *Adh1* gene of maize, over 80% of the sequences are repetitive DNAs that fall into at least 37 classes, as determined by cross-hybridization (Springer *et al*., 1994). This pattern is observed generally in the maize genome, where short (2–20 kb) blocks of low-copy-number and unmethylated DNA alternate with long (up to 200 kb) blocks of intermixed repetitive and methylated DNA (Bennetzen *et al*., 1994). The different classes of repeats were considered to be randomly scrambled with no apparent pattern (Springer *et al*., 1994). Recent studies, however, have indicated that the mosaic of mixed dispersed repetitive elements has originated primarily from the insertion of retrotransposons into one another (SanMiguel *et al*., 1996). Our goal now is to develop tools for the identification of genes and other functionally important elements within such a complex chromosomal continuum.

Our initial attempts to localize genes were based on the common approaches for identification of coding sequences. However, Northern and reverse Northern analyses, as a well as cDNA cloning techniques, indicated that most of the repetitive elements surrounding the *Adh1* gene were homologous to transcripts (Avramova *et al*., 1995). Hence, in maize, a transcriptional criterion will often enrich for sequences that are <u>not</u> standard Mendelian genes.

A second approach utilized diagnostic DNA sequencing of low-copy-number regions within the contiguous region. However, most of the results failed to uncover significant homology with sequences in the standard data bases (GenBank, EMBL), nor were other obvious features of genes identified (unpublished). In fact, a low-copy-number region 3′ to the gene was found to be a retroelement, arguing that copy number alone is a not a reliable indicator of the presence of a gene.

Recombinational mapping of numerous grass genomes with common DNA markers has indicated extensive conservation of both gene content and gene map order (Ahn *et al*., 1993; Bennetzen and Freeling, 1993; Hulbert *et al*., 1990; Moore *et al*., 1995). In contrast, the interspersed highly repetitive DNAs in different grass species usually do not cross-hybridize detectably, due to differential presence and/or a more rapid evolution (Hulbert *et al*., 1990). Recent studies have shown that between-species conservation of gene content and order is also present at the level of single large-insert clones (Chen *et al*., 1996; Dunford *et al*., 1995; Kilian *et al*., 1995).

A possible one-step approach for the identification of functionally important features in large genomic continuums is based on the idea that genes, regulatory

Received 2 June 1996; revised 9 September 1996; accepted 10 September 1996.
*For correspondence (fax +1 317 496 1496; e-mail maize@bilbo.bio.purdue.edu).

(a)

```
madh1   AATGGCGACC GCGGGGAAGG TGATCAAGTG CAAAGGTCCG CCTTGTTTCT CCTCTGTCTC TTGATCTGAC    70
                                                  IVS  I

sadh                                       AAG CTTCCTAGGG AGTTTTTTTT CCCCGATGAA    33
madh1   TAATCTTGGT TTATGATTCG TTGAGTAATT TTGGGGA--- ----GTCCAC ---------- T..-------   138

sadh    CAGTGCCGCA GTGGCGCTGA TCTTGTATGT TGTCCTGCAA TCGCGGTGAA CTTGTTCTTT TTCTATCCTT   103
madh1   ---------- ---------- ---------C -A-------- ---T------ ---A-GTC-- --A-----.-   207

sadh    TAACCCCCAT GAAAATGCTA TTTATCTTTC TTACATCTTC CAGTTCCAGC ACTGCTATTA CC.....GTC   168
madh1   -C--TA---- -----GA--- G-A------- -CGATGTAA- ATCG------ ---------- --GTGTG---   277

sadh    CATCCGACAG TCTGGCTGGA C......... .......... .......... ... ...... ..........   189
madh1   ---------- --------A- -ACATCATAC GATATTGAGC AAAGATCTAT CTTCCCTGTT CTTTAATGAA   347

sadh    .......... .......... .......... ...TGACACT ACTTATGGAG .CATTGCTTT CTTTGAATTT   225
madh1   AGACGTCATT TTCATCAGTA TGATCTAAGA ATG-TG--AC TTGC-A---- G-G--T---- ----------   417

sadh    AACTAACTCG TTGAGTACTG GCTCTGTTTC TCGGACGGAA GACATTTGCC TAATCCACCA TGTCCATTCG   295
madh1   ---------- -----...-- --C------- -------T-- -G-C-----T GCTC-ACA-- ----------   484

sadh    AATTTTGCCG GTGCTTTAGC AAGGGCGAAA AGTTTGCGTC TTGATGGTT. AGCTTGACTA TGTGATTGCT   364
madh1   ------A--- .--.------ ---------- -------A-- ------A--T ---------- --C-------   552

sadh    TTCCTTGGAC CCGTGCAGCT GCGGCGGCAT GGGAGGCCGG CAAGCGACTG TCGATCGAGG AGGTGGAGGT   434
madh1   -----.---- ---------- ----T----- ---------- -----C---- ---------- ----------   621

sadh    AGCGCCTCCG CAGGCTATGG AGGTGCGCGT CAAGATCCTC TTCACCTGCC TCTGCCACAC CGACGTCTAC   504
madh1   ---------- -----C---- ---------- ---------- -------CG- ---------- ----------   691

sadh    TTCTGGGAGG CCAAGGTATC TAATCAGCCA TCCCACATTT GGGTCTTGTC AGTAGATATG ATGCAGCAAC   574
madh1   ---------- ---------- ---------- -----TT-G- -ATCT----- ---------- --A--A----   761
                   IVS  II

sadh    TCGCGGTTGA CTTGCGCTTT CTTGGCGGCT TATCTGTCTT AGGGGCAGAC TCCGGTGTTC CCTCGGATCT   644
madh1   ---------- -------C-- ---------- ---------C ---------- ---C------ ----------   831

sadh    TCGGTCATGA GGCTGGAGGG TATGTTCTAT TCTCCGATTT ACTTCACTAT GTTGCTGGCT AAGTGGCTAT   714
madh1   -T--C----- ---------- ---------- --C------- ---------- -----..... ...--A----   893
                       IVS  III

sadh    AGATGTACTG TGTTTATACT TTACAAATGT ATTATGTTCG AGCGCCTGAA TTTATGGTTG TCGTTGGTGG   784
madh1   ------G--- --------T- -----T--T- --------TT T---T----- -------G-A -G--------   963

sadh    TCTTTGTTTA ATGGTTTACT AGATGCATGT GGAAGAGTCA TTAAGAAACA GTTTCTGTTT GAAACGGTAT   854
madh1   ---------- C--T------ ---------- ---------- .G------T- ----T----- ----T-----  1032

sadh    ATCAACGTTT G.....AATA TTATCTGTGT GGACGTCAGA TGTTGTGG.. TATTGGCAGT GGACTTAG..   915
madh1   -C-----G-- -GATATT--- ---------- ----A----- ----C---GT --C------- ------T-AC  1102

sadh    ......ATTT ATAAT..TTT CAGCAGCAGT TTCTTCAGCT AATTTACTCT CG.AGTATTT TTCAGTTTAC   976
madh1   AGATTT--C- --G--TC--- --TT------ ---------- ---------- ..T-C----- ------A---  1170

sadh    AAAGACA.GT ACAACTAGGG TTGTTCAGTC CTTAATTGGC ATATATGAGA TTTTGCTTTA ATTTGAGATT  1045
madh1   -------C-- ---G------ ----GT--.. .......... .......... ......AA-C G---A----C  1212

sadh    CGTTATCTGA GTCAAATTTG CTTATTCTAG CGCCCTGAAA TTTCTTGATT TTGACAGTAT CATAGAGAGT  1115
madh1   T-----A--- -G------A- ---------- -CG------- ---------- ---C------ ----------  1282

sadh    GTTGGAGAGG GTGTGACTGA CGTAGCTCCC GGTGACCATG TCCTTCCTGT GTTCACTGGG GAGTGCAAGG  1185
madh1   ---------- ---------- ---------- ---------- ---------- ---------- ----------  1352

sadh    AGTGCGCCCA CTGCAGATCG GCAGAGAGCA ACATGTGTGA TCTGCTCAGG ATCAACACTG ACCGGGGTGT  1255
madh1   ---------- -----AG--- ---------- ---------- -T-------- ---------- ----C-----  1422

sadh    GATGATTGGC GATGGCAAGT CACGGTTTTC AATCAATGGA AAGCCTATCT ACCATTTTGT TGGGACTTCA  1325
madh1   ---------- ---------- -G-------- ---------G ---------- ----C----- ---------C  1492

sadh    ACCTTCAGCG AGTACACTGT CATGCATGTT GGTTGTGTTG CAAAGATCAA CCCTGAGGCT CCCCTTGATA  1395
madh1   ---------- -------C-- ---------C ---------- ---------- ----C----- ----------  1562

sadh    AAGTTTGTGT TCTTAGTTGT GGTATTTCTA CTGGTAAGTT CATTTATTAC ATTTGGGTAT GGATGTTAGC  1465
madh1   -------C-- C-----C--- ---------- ---------- --A---C--- ----T---G- -----C-G-A  1632
                                                  IVS  IV

sadh    ATATATTTAT GTTGAGATGC TGAGTTATAT AA....TTTC TATGTTTAGG TCTTGGTGCA TCAATTAATG  1531
madh1   G--C------ C--------- -------C-C --ATTC---- -C-------- ---------- ----------  1702

sadh    TTGCAAAACC ACCAAAGGGT TCTACAGTGG CTATTTTCGG TCTAGGAGCA GTTGGTCTTG CCGTAAGTGT  1601
madh1   ---------- T--G------ --G------- --G------- -T-------C ---------- ----------  1772
                                                                          IVS  V

sadh    TGAAATTATT CACTTGTTCT ATCTTCTATG ACCTTTCCAT TACAATGAAA ATGTGTGTTG GGTTTGCATC  1671
madh1   -----CG--- TG---..... ...------- --A----A-- -G-----G-- ---------- ----------  1835

sadh    TGATTACGCT GCACATGATT .AGGCTGCAG                                             1700
madh1   -------C-- --G----GT- A---------                                             1894
```

sequences, structurally important elements, etc. will be more conserved in evolution than will other sequences. The similarity of gene content and order (synteny) of the maize and sorghum genomes and the almost complete lack of similarity among their respective highly repeated DNAs (Hulbert *et al.*, 1990), suggested that microsynteny could be employed as a tool for gene identification in the two species.

Hence, we cloned the region orthologous to maize *Adh1* from sorghum on a bacterial artificial chromosome (BAC). We found that cross-hybridization between the regions flanking the *Adh1* orthologs of maize and sorghum specifically identified genes. This very simple and rapid approach uses genomic microsynteny and the power of natural selection to identify islands of conserved genetic function in a sea of repetitive DNA, and should be applicable to any syntenous species that contain divergent repetitive DNAs.

Results and Discussion

Identification and cloning of an Adh1 *ortholog from sorghum*

A prerequisite for employing local genomic cross-referencing is to have cloned orthologous genomic regions from two species. We chose sorghum to assist in maize genome analysis because it has an approximately 3.5-fold smaller genome, and shares more than 90% homology with maize low-copy-number number sequences and little, if any, with most maize repeats (Hulbert *et al.*, 1990).

A fragment of the maize *Adh1* gene was used as a probe to screen a recombinant cosmid library containing sorghum genomic DNA. One *Adh1*-homologous cosmid was isolated and a 1.7 kb *Hin*dIII fragment from this clone was sequenced. The results (Figure 1) confirmed that this sorghum clone was highly similar to the maize *Adh1*. Over a distance of 1700 sequenced nucleotides, the maize *Adh1* gene and the sorghum *Adh1* homolog were 90% identical with a predicted 97% amino acid identity (Figure 1).

The 1.7 kb *Hin*dIII fragment of this sorghum *Adh1*-homolog was used as a probe to screen a BAC library containing large inserts of sorghum DNA (Woo *et al.*, 1994). Out of 14 112 clones screened, four clones with homology to maize *Adh1* were identified. Limited DNA sequencing of the *Adh1*-homologous regions of these BACs clones indicated that three of them contained other members of the *Adh1* family, while one of them (No. 110K5), carried the gene that was orthologous to maize *Adh1*.

Cross-hybridization of maize and sorghum DNAs flanking Adh1

The whole BAC No. 110K5, containing a 165 kb insert of sorghum DNA, was labeled and used as a probe in hybridization to a contiguous series (contig) of lambda clones containing the maize *Adh1* region (Springer *et al.*, 1994). Only about 11% of the sequences in the 280 kb maize contig cross-hybridized to the sorghum probe. In most cases, the positive bands corresponded to the low-copy-number regions on the contig (Figure 2).

Sorghum BAC No. 110K5 was restriction mapped and fragments of various lengths were subcloned and used as hybridization probes to the maize contig. These individual fragments cross-hybridized with sequences in the same linear order in each chromosomal region, indicating extensive microsynteny (Figure 2). In general agreement with the 3.5-fold smaller size of the sorghum genome (Arumuganathan and Earle, 1991), the contiguous 120 kb region of maize (fragments 39–93) cross-hybridized with a contiguous 62 kb region of sorghum.

The region 5′ to *Adh1* from the maize contig, composed exclusively of repetitive DNAs, did not share detected homology to the region 5′ to the sorghum *Adh*. A low-copy-number sorghum sequence (fragment B), located about 15 kb 5′ to the sorghum *Adh* gene, was recombina-

(b)

```
sadh                AAAWEAGK RLSIEEVEVA PPQAMEVRVK ILFTCLCHTD VYFWEAKGQT   48
madh1   MATAGKVIKC KA-V------ P--------- ---------- ----S----- ----------   60

sadh    PVFPRIFGHE AGGIIESVGE GVTDVAPGDH VLPVFTGECK ECAHCRSAES NMCDLLRINT  108
madh1   ---------- ---------- ---------- ---------- -----K---- ----------  120

sadh    DRGVMIGDGK SRFSINGKPI YHFVGTSTFS EYTVMHVGCV AKINPEAPLD KVCVLSCGIS  168
madh1   ---------- ---------- ---------- ---------- -----Q---- ----------  180

sadh    TGLGASINVA KPPKGSTVAI FGLGAVGLAA                                   198
madh1   ---------- ---------V ----------                                   210
```

Figure 1. Comparison of the sequences of maize *Adh1* and a sorghum *Adh1* homolog.
(a) Nucleotide comparisons, starting with the ATG in maize that signals initiation of translation in maize *Adh1* (Dennis *et al.*, 1984). Underlining indicates introns (IVS, intervening sequence). Only differences from the sorghum sequence are shown for the maize sequence. sadh, sorghum *Adh*; madh1, maize *Adh1*.
(b) Predicted amino acid comparisons, starting with the initiator methionine of maize *Adh1* (Dennis *et al.*, 1984). Only differences from the sorghum sequence are shown for the maize sequence. sadh, sorghum *Adh*; madh1, maize *Adh1*.

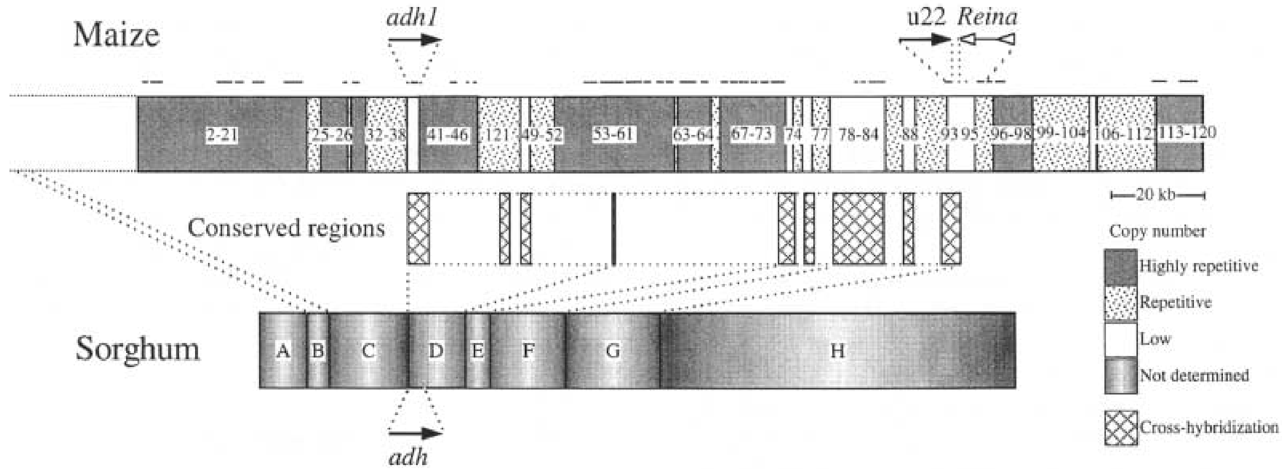

Figure 2. Conserved sequences identified in a microsyntenous region of the maize and sorghum genomes.
The upper bar represents a 290 kb segment of the maize genome. The fill indicates the copy number of various sequence blocks, as previously determined (Springer *et al.*, 1994). Open fill indicates copy numbers less than 100, dotted fill designates copy numbers in the hundreds, and dark fill indicates copy numbers in the thousands. The lines above the maize bar indicate areas with homology to transcripts found in maize roots, leaves and/or tassels (Avramova *et al.*, 1995). The arrows above the bar indicate two transcripts; those encoded by the maize *Adh1* gene and a gene (u22) with homology to a rice seedling cDNA. The double arrows indicate a novel retrotransposon which we have named *Reina*. The numbers within the bar are fragment numbers from the contiguous series generated in Springer *et al.* (1994) and further refined in Avramova *et al.* (1995). The central boxes indicate the placement and size of cross-hybridizing fragments on the maize *Adh1* contig. Lines connect these boxes to the probe fragments from sorghum BAC No. 110K5 (lowest bar). Letters within the lowest bar present consecutive designations that we have arbitrarily applied to these fragments. The arrow below this bar indicates the position and orientation of the sorghum homolog of maize *Adh1*. The dotted lines from sorghum fragment B indicate that this fragment is not homologous to sequences on the maize YAC, but does co-segregate with maize *Adh1* in recombinational mapping (see text).

tionally mapped on a set of maize recombinant inbred lines (Burr *et al.*, 1988) and found to co-segregate 100% with *Adh1* (data not shown). Hence, a homolog of this sorghum fragment is present in maize, presumably at a site 5′ to the maize contig that we have cloned. The advantage of using a smaller genome for the study of a larger and more complex genome is illustrated by this example where a region of approximately 15 kb in sorghum spans a region of more than 90 kb in maize, thus facilitating chromosome walks and studies of genome structure, function and evolution.

Characterization of cross-hybridizing and non-cross-hybridizing fragments

Most of the maize fragments that cross-hybridized with the sorghum BAC were low-copy-number DNAs. To further determine the nature of low-copy number sequences that did, or did not, cross-hybridize between maize and sorghum, we subcloned and sequenced fragments 93 and 95. A portion of the sequence of cross-hybridizing fragment 93 (Figure 3) showed homology with a clone from a rice cDNA library made with RNA from etiolated seedling shoots (DDBJ accession No. RICS1659a). A key difference between the two sequences is that the maize genomic sequence contains apparent introns relative to the rice cDNA (Figure 3).

The adjacent low-copy number region (fragment 95) did not cross-hybridize with the sorghum BAC. Its complete sequence, and that of adjacent fragments (GenBank accession No. U69258) indicated the presence of a novel low-copy-number retroelement that we have named *Reina* (retroelement inserted near *Adh*) (Figure 4). Hence, in the two cases investigated, we have illustrated the power of local genomic cross-referencing: the low-copy-number region that cross-hybridized between the two species is a gene, while the non-hybridizing low-copy-number sequence is a mobile DNA. These results emphasize the degree to which this cross-hybridization process is superior to copy-number determinations for the identification of genes.

A highly repetitive fragment from the maize contig that cross-hybridized to sorghum

An apparent exception to the general observation that only low-copy-number DNAs cross-hybridized between the two species was seen: a 5.9 kb maize DNA fragment (number 56), carrying at least three different classes of dispersed repeats (San Miguel *et al.*, 1996; Springer *et al.*, 1994), was found to hybridize to the sorghum clone. However, detailed mapping and hybridization studies of this fragment identified an internal 450 bp sequence which proved to be low-copy-number in subsequent Southern analysis (unpublished observations). Hence, local genomic cross-referencing allowed the identification of a small, conserved, low-copy-number DNA buried in a large block of highly reiterated sequences that could have been overlooked in

```
1
GCCATGAGGC TAACATCTGA GACTTGGTGT TTTCAGATTA TCTCCGGAAA ATCATAGTGC AGGGAAGGTC TTCCAAGTCA ACTAAGGTTG GCGACATCAT GACTGAAGAG GTATTACCCC
                                        TA cCTCaGGAAA ATCATAGTGC AGGGAcGaTC TTCCAAGTCA ACTAAaGTTG GaGACATCAT GACTGAAGAG ..........
                                        1
121
CACTAATCTT TTTTGTGTGT GTGGTTTGCA AGATAAGCTA GAACATGATT AGCATAAGAC CAGTACCTGG TGACGATATG ACCTATATAA GTGCCGAAAG ATGCTTGGAC ACAACTTAAC
.......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

241
CGTGAACCGT TATGATGCCA GAACAAACTG ATCACAGTGA ATCCCGACAC CAAGGTCCTG CAGGCAATGC AGCTCATGAC AGGTAACCTG GTTGCCCGTG TTCTTCTGTC AGAAAAATAC
.......... .......... .AACcAgCTG ATCACgGTGA AgCCtGAtAC CAgaGTCCTG CAaGCAATGC AGCTgATGAC AG........ .......... .......... ..........
                       73
361
TACTAGAATC GTGAGAGCAC TGGAAAATTG GTTATGATGC CGCTGTACCA ATCGTCTGCG CAGAAAACCG CGTCAGACAC ATCCCGGTGA TCGACGGCAC CGGGATGCT. GGGATGGTCT
.......... .......... .......... .......... .......... .......... ...AgAAgCG CaTCAGgCAC ATCCCGGTGA TCGACGGCAC gGGcATGgTt GGGATGGTCT
                                                                     133
480                                                 538
.CATCGGCGA CGTCGTGCGC GCGGTGGTGG CCGCGCAA.G GGAGGAGCTG AACCGGCTC
cCATtGGnGA CaTtGTcCGC GCcGTGGTca gCGaGCAccG GGAaGAGCTG AACCGGCTC
                                                    251
```

Figure 3. Comparisons of the sequence of a rice cDNA with the sequence of a region near maize *Adh1* that cross-hybridizes with the orthologous region from the sorghum genome.
The sequence of the cross-hybridizing region of fragment 93 is shown on the upper line. The lower line is a homologous sequence of a rice seedling cDNA clone found in DDBJ (accession No. RICS1659a). Lower case letters indicate sequence differences between the rice cDNA and the maize genomic DNA. Boxes indicate proposed exons and the vertical lines mark the intron/exon junctions present in the maize genomic sequence.

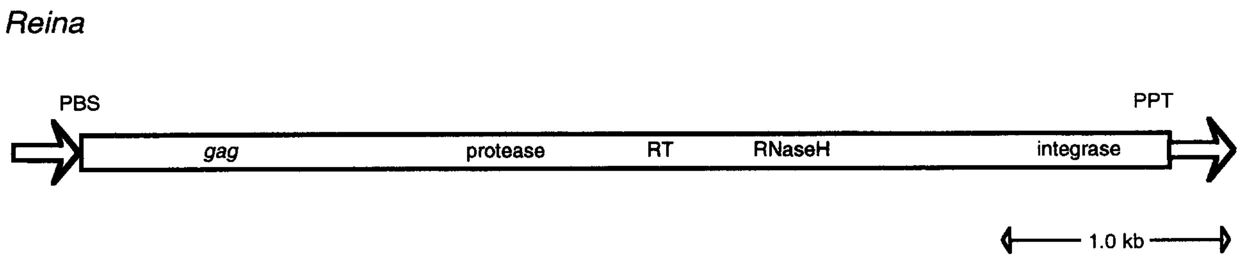

Figure 4. Schematic representation of the *Reina* retrotransposon.
The non-hybridizing low-copy-number regions from the maize contig overlapping fragments 93–95 were sequenced (accession number) and found to belong to a 5.4 kb retrotransposon. Sequences, corresponding to the LTRs, primer binding site (PBS), and polypurine/polypyrimidine tracks (PPT) were identified, enabling us to place the 5′ and 3′ LTRs, respectively. Internal sequences showed extensive homologies to the *gag* and *pol* (RT/RNase H) genes. The location of the integrase coding potential at a position 3′ to *pol* indicates that *Reina* is a Ty3/*gypsy*-class retrotransposon.

other types of analyses. Sequencing of the 450 bp fragment did not reveal similarity to other known sequences, but the fact of its conservation in the two species suggests that it may be functionally important.

Local genomic cross-referencing

The process of local genomic cross-referencing, which uses genic sequence conservation and genomic microsynteny as tools to identify important genetic features, will be of value in any syntenous species. Indeed, the power of microsynteny in gene identification has been demonstrated in a few cases, using rodent and human genomes (Koop, 1995). However, these studies were based on sequencing and computer sequence comparisons of enormous (megabase) genomic regions of the two species. In our investigation, conserved regions were identified with a simple, time- and cost-effective cross-hybridization analysis of large syntenous regions. To confirm and extend our studies, we mapped the sorghum BAC and performed additional hybridizations with the individual fragments. What is important, however, is that the identification of the conserved sequences on the maize contig was accomplished completely in one step, with the first hybridization to the whole BAC clone. Genes hidden in blocks of repetitive DNA can be uncovered, while low-copy number sequences that are not conserved genes can be eliminated from the genome search, whether or not they are transcribed. A particularly exciting potential value of this approach will be its use in identifying conserved features other than genes.

Experimental procedures

Cosmid and BAC library screenings

The sorghum cosmid library (constructed as a *Sau*3AI partial digest of Tx430 DNA in a derivative of pHC79 (Hohn and Collins, 1980), generously provided by S. Hulbert, Kansas State University), was screened with the 2.3 kb *Hin*dIII fragment that is central to the maize *Adh1*-S gene (Dennis *et al.*, 1984). A 1.7 kb *Hin*dIII fragment of a sorghum cosmid clone that was homologous to maize *Adh1* was used as a hybridization probe to screen a sorghum BAC library by the techniques previously described (Woo *et al.*, 1994).

Gel blot hybridizations

Cross-hybridization between sorghum and maize fragments was determined using gel-purified restriction fragments of sorghum BAC No. 110K5 as probes. Fragments were labeled by the technique of Feinberg and Vogelstein (1984) and hybridized to gel blot replicas of restriction-digested lambda clones containing an overlapping contig series of fragments from the maize *Adh1* region (Springer *et al.*, 1994). Hybridization, washing and exposure conditions were as previously described (Jin and Bennetzen, 1994).

DNA sequencing

DNAs were sequenced as previously described (Jin and Bennetzen, 1994) or from the ends of *Tn*1000 insertions generated in subcloned fragments (Strathmann *et al.*, 1991). Some sequencing reactions were analyzed in an ALF express automated sequencer (Pharmacia).

Acknowledgments

We are grateful to Dr S.H. Hulbert for providing a sorghum cosmid library. This work was supported by research grants from the USDA/NRIGP (No. 93-37300-8769 to Z.A. and No. 94-37300-0299 to J.L.B.) and the Texas Agricultural Experiment Station to (R.A.W).

References

Ahn, S., Anderson, J.A., Sorrells, M.E. and Tanksley, S.D. (1993) Homologous relationships of rice, wheat and maize chromosomes. *Mol. Gen. Genet.* **241**, 483–490.

Arumuganthan, K. and Earle, E.D. (1991) Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208–218.

Avramova, Z., SanMiguel, P., Georgieva, E. and Bennetzen, J.L. (1995) Matrix attachment regions and transcribed sequences within a long chromosomal continuum containing maize *adh1*. *Plant Cell,* **7**, 1667–1680.

Bennetzen, J.L. and Freeling, M. (1993) Grasses as a single genetic system: genome composition, collinearity and compatability. *Trends Genet.* **9**, 259–261.

Bennetzen, J.L., Schrick, K., Springer, P.S., Brown, W.E. and SanMiguel, P. (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome,* **37**, 565–576.

Burr, B., Burr, F.A., Thompson, K.A., Albertson, M.C. and Stuber, C.W. (1988) Gene mapping with recombinant inbreds in maize. *Genetics,* **118**, 519–526.

Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S.-S., Zhang, H., Wing, R.A. and Bennetzen, J.L. (1996) Microcolinearity in the maize, rice and sorghum genomes. *Proc. Natl Acad. Sci. USA,* in press.

Dennis, E.S., Gerlach, W.L., Pryor, A.J., Bennetzen, J.L., Inglis, A., Llewellyn, D., Sachs, M., Ferl, R.J. and Peacock, W.J. (1984) Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize. *Nucl. Acids Res.* **12**, 3983–3995.

Dunford, R.P., Kurata, N., Laurie, D.A., Money, T.A., Minobe, Y. and Moore, G. (1995) Conservation of fine-scale DNA marker order in the genomes of rice and the Triticeae. *Nucl. Acids Res.* **23**, 2724–2728.

Feinberg, A.P. and Vogelstein, B. (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**, 266–267.

Flavell, R.B., Bennett, M.D., Smith, J.B. and Smith, D.B. (1974) Genome size and proportion or repeated nucleotide sequence DNA in plants. *Biochem. Genet.* **12**, 257–269.

Hohn, B. and Collins, J. (1980) A small cosmid for efficient cloning of large DNA fragments. *Gene,* **11**, 291–298.

Hulbert, S.H., Richter, T.E., Axtell, J.D. and Bennetzen, J.L. (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proc. Natl Acad. Sci. USA,* **87**, 4251–4255.

Jin, Y.-K. and Bennetzen, J.L. (1994) Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the *Bs1* retroelement of maize. *Plant Cell,* **6**, 1177–1186.

Kilian, A., Kudrna, D.A., Kleinhofs, A., Yano, M., Kurata, N., Steffenson, B. and Sasaki, T. (1995) Rice-barley synteny and its application to saturation mapping of the barley *Rpg1* region. *Nucl. Acids Res.* **23**, 2729–2733.

Koop, B.E. (1995) Human and rodent DNA sequence comparisons: a mosaic model of genomic evolution. *Trends Genet.* **11**, 367–371.

Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. (1995) Grasses, line up and from a circle. *Curr. Biol.* **5**, 737–739.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z. and Bennetzen, J.L. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science,* in press.

Springer, P.S., Edwards, K.J. and Bennetzen, J.L. (1994) DNA class organization on maize *Adh1* yeast artificial chromosomes. *Proc. Natl Acad. Sci. USA,* **91**, 863–867.

Strathmann, M., Hamilton, B.A., Mayeda, C.A., Simon, M.I., Meyerowitz, E.M. and Palazzolo, M.J. (1991) Transposon-facilitated DNA sequencing. *Proc. Natl Acad. Sci. USA,* **88**, 1247–1250.

Woo, S.-S., Jiang, J., Gill, B.S., Paterson, A.H. and Wing, R.A. (1994) Construction and characterization of a bacterial artificial chromosome library for *Sorghum bicolor. Nucl. Acids Res.* **22**, 4922–4931.