A Rapid and Direct Approach to Identify Promoters That Confer High Levels of Gene Expression in Monocots

Meizhu Yang, Robert Bower, Mark D. Burow, Andrew H. Paterson,* and T. Erik Mirkov*

ABSTRACT

Direct screening of genomic libraries for highly expressed genes is an efficient way to identify promoters that confer high levels of gene expression. To test the efficacy of this approach to isolate promoters for directing high levels of gene expression in sugarcane, 11 genomic clones were isolated after screening a sugarcane genomic library with radioactively labeled first strand cDNA probes synthesized from sugarcane leaf mRNA. Hybridization analysis with pooled first strand cDNA indicated that eight of the clones probably contain genes with expression levels similar to or higher than ubiquitin (which was also isolated in this screen). The coding regions of these 11 genomic clones were isolated from a cDNA library, sequenced, and found to represent parts of five different genes including elongation factor 1α, α-tubulin, an aquaporin, a proline-rich protein, and one novel gene. Southern and Northern analyses showed that the sugarcane proline-rich protein-encoding gene (SPRP1) was the best candidate to isolate a promoter that would direct high levels of expression, although the other four are also good candidates. Two genomic clones of the sugarcane proline-rich gene and a sugarcane elongation factor 1α (SEF1α) gene, which contain both the promoter and coding regions, were subcloned, sequenced, and promoter regions defined by comparison of cDNA and genomic DNA sequences. Both the SEF1α and SPRP1 promoter/β-glucuronidase (GUS) fusions resulted in high levels of GUS expression when reintroduced into embryogenic sugarcane callus, and also resulted in high levels of transient GUS expression in wheat embryos. This approach has value as an efficient technique to find promoters that confer high-level gene expression in monocots.

Sugarcane is one of the world’s most important crops, with the 2001–2002 world cane sugar crop forecast at a near record 126.8 million megagrams (FAS, 2001). Modern cultivated sugarcane is believed to be derived from interspecific hybrids between the domesticated species Saccharum officinarum L. and its wild relative S. spontaneum L. Chromosome numbers of sugarcane cultivars range from 100 to 130 with approximately 10% derived from S. spontaneum (Simmonds, 1976). Inter-specific hybridization has reduced disease problems, increased biomass yield and sugar yield, and improved adaptability for growth under various stress conditions (Roach, 1972; Srivastava et al., 1994).

The complicated genetics and long breeding cycle for sugarcane make transformation an attractive approach to include in the toolbox for sugarcane improvement. The most commonly used method for transformation of sugarcane is particle bombardment of embryogenic callus combined with a herbicide resistance gene as a selectable marker (Gallo-Meagher and Irvine, 1993, 1996; Bower and Birch, 1992; Bower et al., 1996). Production of transgenic sugarcane plants by intact cell electroporation (Arencibia et al., 1995) and by Agrobacterium-mediated transformation of embryogenic callus has also been reported (Enriquez-Obregon et al., 1998; Arecibia et al., 1998).

In conjunction with an efficient transformation system, the availability of a range of promoters with varied strengths and tissue specificities is critical to the success of transgenic approaches to crop improvement. The CaMV 35S promoter has been widely used for high-level constitutive expression in dicots, but confers lower levels of expression in monocots. Promoters isolated from monocots generally show higher activity in monocots, and adding an intron between the promoter and the reporter gene can increase transcription levels (Wilkinson et al., 1995; Rathus et al., 1993; Maas et al., 1991; Callis et al., 1987). The rice actin promoter Act1 (McElroy et al., 1990, 1991; Wang et al., 1992; Zhang et al., 1991) and the maize ubiquitin promoter Ubi1 (Christensen and Quail, 1996) achieved far better expression than the CaMV 35S in most monocots tested. Among promoters tested in sugarcane, the Emu promoter and the maize ubiquitin promoter showed higher levels of expression than the CaMV 35S promoter (McElroy et al., 1991; Gallo-Meagher and Irvine, 1993; Rathus et al., 1993), curiously the opposite of their relative strengths in dicots (Callis et al., 1990; Mitra and Higgins, 1994). Variation in transgene expression levels between different species and promoters may be due to different abundance of transcription factors, recognition of promoter sequences or intron splicing sites (Wilkinson et al., 1995) or other factors.

Sugarcane, as well as other monocots, may be a useful source of promoters that can provide specific transgene expression patterns in valuable monocot crop species, but there are few published results on sugarcane promoters (Wei et al., 1999). We seek to begin to fill this
knowledge gap, using a direct, broadly applicable approach to discover new promoters for high-level or tissue-specific transgene expression.

**MATERIALS AND METHODS**

**Lambda Genomic Library Construction**

Genomic DNA (averaging about 100 kilobases (kb) in length) was isolated from frozen leaves of sugarcane cultivars CP65-357 and CP72-1210 by a cetyltrimethylammonium bromide (CTAB) method (Honeycutt et al., 1992), partially digested with SmaI and size fractionated by preparative agarose gel electrophoresis. DNA in the 10- to 23-kb range was purified and ligated to BamHI-predigested Lambda DASH II vector at an equal ratio of insert:vector. Ligation was performed at 16°C overnight. The ligated DNA was packaged with a Gigapack III Gold Packaging Extract (Stratagene, La Jolla, CA). About 1 × 10^6 phages were amplified once and the amplified library was used for further screening.

**Preparation of Pooled cDNA Probe and Screening of a Genomic Library**

RNA was extracted from young leaf tissue (greenhouse grown, 5 mo old) of sugarcane cultivar CP72-1210 according to a protocol developed by Yang Si in Dr. Paterson’s lab (personal communication). About 1 g plant tissue was frozen in liquid nitrogen and ground into fine powder. This powder was then transferred into a 50-mL conical tube containing 10 mL ice-cold RNA extraction buffer [200 mM Tris–HCl pH 8.5, 2% (w/v) SDS, 10 mM Na2-ethylenediaminetetraacetate (EDTA), 1% (w/v) sodium deoxycholate and 1% (w/v) polyvinyl pyrrolidone 40]. The powder in solution was blended in a polytron at high speed for 1 min after adding 10 mL PCI (phenol:chloroform:isoamyl-alcohol = 25:24:1). Next, 450 μL of sodium acetate (3.3 M pH 5.2) was added and mixed well. This mixture was kept on ice for about 15 min to let RNA diffuse into the aqueous phase. The (upper) aqueous phase was separated by centrifugation at 3500 × g for 20 min and transferred to a fresh conical tube. The RNA was precipitated with isopropanol and redissolved in 800 μL H2O, mixed with 200 μL 10 M LiCl and incubated on ice for 5 to 12 h. This solution was then centrifuged at 12 000 × g for 15 min, the pellet resuspended in 400 μL H2O, and mixed with 600 μL 5 M KOAc (pH not adjusted). The mixture was incubated again on ice for 3 h and centrifuged at 12 000 × g for 20 min. Finally, the RNA pellet was resuspended in 200 μL H2O, and then precipitated with ethanol.

Poly A + RNA was isolated from total RNA by means of Oligo (dT) Cellulose (New England Biolabs, Beverly, MA) according to the manufacturer’s instructions. About 0.5 to 1 μg poly A + RNA was transcribed into first strand cDNA by BRL Superscript reverse transcriptase according to the manufacturer’s instructions, except that 9 μL of 6000 CI/Mol [α-32P]dCTP per reaction was used. The probe was denatured and added to hybridization buffer for library screening.

The genomic library was plated on 100-× 15-mm Petri dishes at a density of about 5000 plaques per plate. Twenty plaque lifts (Hybond N+ membrane, Amersham, Piscataway, NJ) were hybridized with pooled first strand cDNA probe. Prehybridization and hybridization were performed at 65°C according to the manufacturer’s instructions (Amersham).

**Construction of the cDNA Library**

Poly A (+) RNA was isolated from leaf total RNA of sugarcane CP72-1210 with a Poly Quick mRNA isolation kit (Stratagene) based on the manufacturer’s instructions. Single stranded and double stranded cDNA were produced from 5 μg poly A (+) RNA. The library was constructed in the Uni-ZAP XR vector (Stratagene) according to the manufacturer’s instructions. The packaged phages were plated on the E. coli cell line XL1-Blue MRF+. About 1 × 10^6 primary clones were amplified and this amplified cDNA library was further screened by DNA probes.

**Southern and Northern Analysis**

Ten micrograms of total genomic DNA isolated from sugarcane leaves (Honeycutt et al., 1992) was digested with various restriction enzymes. Southern blotting, radioactive labeling, and autoradiography were as described by Chittenden et al. (1994). The 10 to 30 μg total RNA isolated from root, stem, and leaf of sugarcane variety CP72-1210 were separated by a 1.2% formaldehyde/agarose gel containing 7% formaldehyde and 1× MOPS buffer. The RNA was transferred to a positively charged nylon membrane (Hybond N+, Amersham) with 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and fixed by baking the filter for 2 h at 80°C. The membrane was hybridized with a cDNA probe. Hybridization and washing was the same as for the Southern hybridization.

**DNA Sequencing and Primer Extension**

All genomic subclones and cDNA clones were cloned in the pBluescript SK plasmid vector with T3 and T7 promoters flanking the insert. The clones were sequenced by means of T3 and T7 primers and further sequenced by designing internal primers. The sequencing reaction was performed according to ABI PRISM diRhodamine Terminator Cycle Sequencing Ready Reaction Kit and run on an ABI 377 (Applied Biosystems, Foster City, CA). The transcription initiation sites were determined by primer extension analysis according to the method developed by Dias (1995) with some minor modifications. The primer-extended cDNA products were analyzed by electrophoresis on a 6% (w/v) urea-polyacrylamide gel in parallel with a sequencing ladder generated with the same primer and corresponding genomic clone as a template. Manual sequencing was conducted with a Sequitherm Cycle Sequencing Kit (Epicentre Technologies, Madison, WI) with [α-32P]dATP.

**Microprojectile Bombardment**

Bombardment of embryogenic sugarcane callus for each test promoter was assessed essentially as described by Gallo-Meagher and Irvine (1993) with the following exceptions. The helium pressure was increased from 0.3 MPa (40 psi) to 0.6 MPa (90 psi) and 0.7-μm tungsten particles were used instead of 1-μm tungsten particles. Wheat embryos (cv. Westonia) were cultured on MS medium containing 2.5 mg/l/2,4-D for 12 d at 24°C in the dark before transfer to an osmoticum medium containing the same ingredients plus 0.4 M sorbitol for 4 h. The scutellar surface was bombarded on the basis of the bombardment conditions and precipitation procedures described by Bower et al. (1996). Precipitations for each test promoter consisted of 2 μg pDP687 (Bower et al., 1996), which drove anthocyanin accumulation, plus 8 μg of the test promoter plasmid.

**Test Promoter Plasmids**

Plasmids pSEF/GUS, pSPRP/GUS, and pSPRP2.4/GUS were constructed (See Fig. 1) by replacing the CaMV 35S...
promoter contained in pBI221 (Clontech, Palo Alto, CA) with 2.0 kb of sequences upstream of the ATG start codon (pSEF/GUS and pSPRP/GUS) or 2.4 kb of sequences upstream of the ATG start codon (pSPRP2.4/GUS). Plasmid pACH27 (pMUBI/GUS; Christensen and Quail, 1996) was included to enable comparison of the SEF1α and SPRP1 promoters to a strong monocot promoter.

Assays of Promoter Activity

GUS activity of bombarded embryogenic sugarcane callus was essentially as described by Gallo-Meagher and Irvine (1993). Fluorometric GUS assays were replicated three times for each treatment and each treatment was replicated three times. Therefore, nine data points were used to calculate each of the treatment means. Bombarded wheat (Triticum aestivum L.) embryos remained on the osmoticum medium overnight then were transferred to MS medium plus 2.5 mg/L 2,4-D and incubated at 24°C in the dark until 48 h post bombardment. They were then rated for the number of red anthocyanin accumulating cells (Ludwig et al., 1990) on the bombarded surface, as an internal control for variability in bombardment efficiency, after which they were incubated in GUS assay buffer (Jefferson, 1987) overnight at 37°C. Substantial variation is often observed in transient expression levels of bombarded wheat tissues because of variation between bombardments. Use of a second reporter gene (anthocyanin) is an effective method of reducing the impact of variability between shots on the accuracy of estimation of promoter strength by bombardment of tissues as it enables an estimation of the effectiveness of each shot. The number of blue cells on the bombarded surfaces was rated by the same scale as for the anthocyanin accumulating cells. The categories were 0 = no red or blue cells; 1 = 0 to 10 red or blue cells/embryo; 2 = 10 to 30 red or blue cells/embryo; 3 = more than 30 red or blue cells/embryo. Each promoter construct was rated on two replicate bombardments consisting of 20 precultured embryos per bombardment and a Promoter Activity Rating (PAR) was derived by calculating the mean ratio of the blue cell number rating divided by the red cell number rating for each plate of bombarded embryos. For both sugarcane and wheat, the maize ubiquitin promoter-intron/GUS (Christensen and Quail, 1996) construct (pMUBI/GUS) was included to enable comparison of the SEF1α and SPRP1 promoters to a strong monocot promoter.

RESULTS

Initial Genomic Library Screening for Highly Expressed Genes

Labeled first strand cDNA was used to screen the genomic library for clones that hybridized to abundant transcripts. Two steps were used in the first screening. In the first step, 1 × 10⁶ phages separated on 20 plates were screened with the first strand cDNA probe to identify clones with a strong hybridization signal (Fig. 2). Because some clones with a strong signal might contain rDNA, or other highly repetitive sequences, further screening was required. In the second step, the same filters were stripped and probed with poly A (-) RNA. Most of the clones showed little hybridization signal. Twenty-two genomic clones which showed a very strong hybridization signal in the primary screening, but did not hybridize with poly A (-) RNA, were selected for secondary and tertiary screening. These screens identified 12 clones that were selected for further characterization.
coding regions of highly expressed genes. The filter was then stripped and hybridized with a subclone from the cDNA of ubiquitin p6t7.2b1 (Christensen and Quail, 1996), revealing that clone 15-1 contained the ubiquitin gene.

**cDNA Library Screening with Probes of Selected Genomic Clones**

The eight genomic clones 9-1, 10-1, 14-1, 14-2, 16-1, 17-2, 18-1 and 19-1, which had similar or higher expression levels than 
ubi, were selected as candidates for promoter isolation. The restriction fragments of λ phage genomic clones containing the coding region were used as probes to screen the sugarcane leaf cDNA library. About 10 to 20 cDNA clones were isolated from the sugarcane leaf cDNA library for each genomic clone. Hybridization results showed that clones 10-1 and 14-1 contained the same gene. Also, 14-2, 17-2, and 18-1 hybridized with the same cDNA clones. In total, five different genes were found following cDNA library screening.

All the cDNA clones were sequenced. GenBank BLASTN and X similarity searches based on partial sequences of both 5' and 3' ends of cDNA inserts indicated that four out of the five cDNAs clones were similar to known genes (Table 1), specifically elongation factor 1α, α-tubulin, an aquaporin, and a proline-rich protein. One of the clones (14-2) did not show significant similarity with any sequence in GenBank.

**Southern Analysis of Genomic DNA and Coding Region Identification**

DNA from the 12 genomic clones selected above was digested with BamHI, EcoRI, and BamHI + EcoRI and run on a 0.8% (w/v) agarose Tris-borate-EDTA (TBE) gel (Fig. 3A). A Southern blot was made from this gel and hybridized with pooled first strand cDNA derived from poly A (+) RNA (Fig. 3B). The fragments which hybridized with pooled cDNA should contain the coding regions of highly expressed genes. The filter was then stripped and hybridized with a subclone from the cDNA of ubiquitin p6t7.2b1 (Christensen and Quail, 1996), revealing that clone 15-1 contained the ubiquitin gene.

**cDNA Library Screening with Probes of Selected Genomic Clones**

The eight genomic clones 9-1, 10-1, 14-1, 14-2, 16-1, 17-2, 18-1 and 19-1, which had similar or higher expression levels than 
ubi, were selected as candidates for promoter isolation. The restriction fragments of λ phage genomic clones containing the coding region were used as probes to screen the sugarcane leaf cDNA library. About 10 to 20 cDNA clones were isolated from the sugarcane leaf cDNA library for each genomic clone. Hybridization results showed that clones 10-1 and 14-1 contained the same gene. Also, 14-2, 17-2, and 18-1 hybridized with the same cDNA clones. In total, five different genes were found following cDNA library screening.

All the cDNA clones were sequenced. GenBank BLASTN and X similarity searches based on partial sequences of both 5' and 3' ends of cDNA inserts indicated that four out of the five cDNAs clones were similar to known genes (Table 1), specifically elongation factor 1α, α-tubulin, an aquaporin, and a proline-rich protein. One of the clones (14-2) did not show significant similarity with any sequence in GenBank.
Expression Analysis of cDNA Clones STUB, SPRP1, SAQ1, and SEF1α

The steady state mRNA accumulation of the isolated cDNAs in different tissues was determined by Northern RNA analysis. In this experiment, RNA was isolated from young tissue of root, stem, and leaf of sugarcane plants grown in the greenhouse. Fig. 4 shows the mRNA accumulation of the four cDNA clones. The STUB1 gene was transcribed most highly in the stem, less in the roots and lowest in the leaf. In contrast to STUB1, the highest mRNA accumulation for SPRP1 was in the leaf, less in the stem, and there was little mRNA accumulation in the roots. mRNA levels for SAQ1 and SEF1α were nearly constitutive, with slightly higher mRNA accumulation in the stem for SAQ1 and in the roots for SEF1α. An 18S rRNA clone was used as a probe to standardize signal levels per lane.

Genomic Southern Hybridization with Four cDNAs

Southern analysis was performed to investigate the copy number of genomic sequences encoding the isolated genes (data not shown). Both sorghum and sugarcane genomic DNA were digested with either HindIII, EcoRI, or XbaI. Sorghum was chosen for this experiment because it is closely related to sugarcane, but is not a recent polypoid, so gives a better indication of gene copy number per genome. The SPRP1 gene showed low copy number, perhaps indicating single copy status in sorghum, and also had the lowest number of restriction fragments in sugarcane. The remaining four genes each hybridized with several genomic restriction fragments in sorghum, suggesting that the genes are members of small multigene families.

Isolation, Identification, and Characterization of a Sugarcane Proline-Rich Protein Gene and its 5’ Upstream Sequence

The cDNA clone SPRP1 had an insert of approximately 1.2 kb (Accession number AF331853). Both nucleotide and amino acid sequences are similar to proline-rich proteins from maize, Zea mays L. (Accession number Y17332) and wheat (Raines et al., 1991). The nucleotide sequences revealed 73% identity between sugarcane and maize, and 70% identity between sugarcane and wheat. Comparison of the deduced amino acid sequences showed that sugarcane PRP has 78% similarity with maize and 77% similarity with wheat. Translation analysis showed that this cDNA clone was not full length, lacking the 5’ end but with 184 base pairs (bp) of 3’ noncoding sequence. The 5’-end sequence of SPRP was obtained from a longer cDNA clone (SPRP2, accession number AF331852), containing 99 bp of 5’-coding sequence with one possible translation start site (ATG). As expected, overlapping sequences were found between the SPRP1 and SPRP2 cDNAs. There is 93% nucleotide sequence identity between these two cDNAs in the 300-bp overlapping region—this degree of divergence suggests that the two cDNAs may belong to two different members of the SPRP gene family in sugarcane. The sugarcane gene isolated here has common structural features with the previously published wheat PRP sequence (Raines et al., 1991) and a maize proline-rich protein, including a hydrophilic N-terminal region that is common to proline-rich cell wall proteins (John and Keller, 1995), high proline content, and a hydrophobic signal peptide. This result in combination with results from the Northern analysis, suggests that the SPRP protein may be a cell wall protein abundant in leaf tissues.

Four genomic clones of the SPRP genes were found by screening of the genomic library. An 8.0-kb EcoRI fragment from one genomic clone was subcloned into the pBluescript SK plasmid vector. Detailed sequencing was done on this 8.0-kb subclone which contains both the promoter and coding region. A total of 1.8 kb of upstream sequence from the translation start site of SPRP1 was sequenced from this clone (Accession number AF331851). Sequence analysis revealed that the promoter contains several important cis elements. There is a consensus TATAAA box 172 bp upstream from the translation start codon ATG. This result indicated that the deduced translation start site might actually function in vivo. A sequence (5’-CCATC) resembling a CAAT box is found 37 bp upstream of the TATA box.

Table 1. cDNA clones which are highly expressed in sugarcane and their predicted genes.

<table>
<thead>
<tr>
<th>Name of cDNA clone</th>
<th>Corresponding genomic clone</th>
<th>Homologous gene after BLAST searching (% nucleotide sequence identity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STUB1</td>
<td>Hordeum vulgare alpha-tubulin 3 (90%)</td>
<td>accession number AJ133399</td>
</tr>
<tr>
<td>SPRP1</td>
<td>Zea mays proline-rich protein (84%)</td>
<td>accession number Y17332</td>
</tr>
<tr>
<td>SAQ1</td>
<td>Zea mays plasma membrane MIP protein (pip2) (96%)</td>
<td>accession number AF131201</td>
</tr>
<tr>
<td>SEF1α</td>
<td>Zea mays Elongation factor 1α (93%)</td>
<td>accession number U76259</td>
</tr>
</tbody>
</table>

Fig. 4. Northern analysis of four cDNA clones isolated by genomic and cDNA library screening. Total RNA was isolated from young roots (R), stems (S), and leaves (L). The 18S rRNA signal was used for standardizing the signal level per lane. Autoradiographs were exposed for 1 d for STUB, half day for SPRP1 and 3 d for SAQ1 and SEF1α.
Isolation, Identification, and Characterization of a Sugarcane Elongation Factor 1α gene and Its 5' Upstream Sequence

A cDNA clone (Accession number AF331850) isolated during library screening was close to full length with the 5' end being 18 bp downstream of the ATG translation start site. The sugarcane SEF1α clone shows 93% similarity to the maize nucleotide sequence and 99% similarity to the maize deduced amino acid sequence (Berberich et al., 1995), respectively. From the 17-kb genomic insert of the SEF1α in the phage genomic clone, a 9.5-kb EcoRI fragment containing the entire coding sequence and the complete promoter region was subcloned in pBluescript SK. The 4537-bp genomic sequence (Accession number AF331849) of the entire gene including the 5' upstream region matched base by base to the corresponding sugarcane and maize EF1α cDNA sequence (Accession numbers AF331850 and U7259). The comparison between genomic and cDNA sequences showed two introns. One of them is located within the 5' untranslated region and is 597 bp in length. In the Arabidopsis A1 EF1-α gene, an intron was also found in the 5' noncoding region and is important for the expression of EF 1α gene in leaves (Curie et al., 1991, 1993). The second intron (779 bp) is located in the coding region. Like other plant introns, these two introns in sugarcane SEF1α have GT and AG nucleotides at the 5' and 3' ends. They are also strongly enriched in AT nucleotides throughout the intron, a feature that is thought to be a requirement of efficient splicing of plant introns (Liu and Filipowicz, 1996).

To map the 5' end of the SEF1α gene, a primer extension reaction was done with a 30-bp primer near the translation start site. On the basis of manual sequencing of the genomic clone, the transcription start site (tsp) is 130 bp upstream of the translation start site (excluding the intron located in the 5' UTR).

To characterize sequences involved in the regulation of SEF1α in sugarcane, about 1.2 kb of 5'-flanking DNA was sequenced. Like other promoters, this region was AT-rich, with a putative TATA box (TATAAA) located 33 bp upstream of the deduced transcription start site and a typical CAAT box found 40 bp upstream of the TATA box. The base composition plot of the entire SEF1α gene and promoter reveals a distribution of AT rich regions typical of plants, with four AT rich regions: the promoter, two introns, and the 3' untranslated sequences. There is only one small GC rich region, in the first exon (untranslated leader sequence).

Estimation of Promoter Strength in Sugarcane and Wheat by Transient Expression Assays

A comparison of the ability of each promoter to drive transient expression of the uidA (GUS) gene was made by microprojectile bombardment of sugarcane and wheat tissues with promoter-GUS fusion plasmids. In sugarcane embryogenic callus, the SEF1α and SPRP promoters directed levels of GUS expression that were comparable to the maize ubiquitin promoter (Fig. 5).

For wheat embryos, the results were used to calculate a Promoter Activity Rating (PAR) for each promoter. Use of the anthocyanin marker gene as an internal control to correct for variation in shot efficiency in transient expression assays meant that minimal variation was observed for the PAR observed in replicate bombardments of each construct. The rating system enabled the direct comparison of the test promoters with the known strong monocot maize ubiquitin promoter. In wheat, the activity of the promoters was rated as follows: SEF1α PAR = 1, SPRP1 PAR = 1, SPRP2.4 PAR = 1, maize ubiquitin promoter PAR = 1.5. It was noted that blue foci varied in size and intensity between cells bombarded with different constructs, with the pMUBI/GUS bombarded cells showing the heaviest blue staining, followed by pSPRP2.4/GUS, pSEF/GUS and pSPRP/GUS in decreasing order (Fig. 6).

DISCUSSION

We have identified and studied several highly expressed genes and their endogenous promoters toward development of constructs ideal for high-level expression of transgenes in sugarcane, using an approach that is easily generalized to other taxa. There were four steps in this approach. First, pooled mRNA from the specific tissue of interest was used as a probe to screen a phage genomic library. Second, the coding region of these genes was estimated by Southern hybridization of restriction-enzyme digested λ phage DNA clones with pooled first strand cDNA, and corresponding cDNA clones were isolated from a sugarcane cDNA library. Third, the copy number of genes was estimated by Southern analysis and the expression level of genes was confirmed by Northern analysis. Finally the promoters were characterized by sequencing of the cDNA and genomic clones, GenBank searches, primer extension analysis, and transient expression analysis.

This approach may help to find new promoters that confer high levels of expression of transgenes in sugarcane or other monocots for which endogenous promot-
Fig. 6. Embryogenic wheat callus showing GUS activity after bombardment with promoter-GUS fusion constructs. (A) Promoter SPRP1, (B) Promoter SPRP 2.4, (C) Promoter SEF1α, and (D) maize ubiquitin-1.

ers have not yet been described. The maize ubiquitin promoter drives high levels of transgene expression in sugarcane and the endogenous sugarcane ubiquitin gene was identified by the screening method described here, supporting the validity of our approach. A significant advantage of this approach to promoter isolation is the capacity to isolate several different promoters at the same time without cDNA library construction (although we did construct a cDNA library to better characterize the genomic clones that we found). This approach can be modified by adding two steps for tissue-specific promoter isolation. The first phage selection can be used to identify any genomic clones that contained highly expressed genes in one tissue. A second screening of phage DNA using a Southern blot or dot blot methods, which are more sensitive than plaque lifts, can be used to eliminate the clones which are expressed in nontarget tissues. A disadvantage of the phage approach is that the DNA isolation is time consuming and low yielding compared with plasmid DNA isolation. Long-range PCR may help to isolate tissue-specific promoters more easily and quickly. Hundreds of phages can be picked in the first screening and selected in the second screening by means of mRNA from different tissues. Growing EST (expressed sequence tags) databases for many taxa may facilitate prediction of the promoter region.

Southern blot analysis of sorghum [Sorghum bicolor (L.) Moench] and sugarcane genomic DNA using SPRP1 cDNA revealed one strongly hybridizing band and many fainter bands, in both S. propinquum and S. bicolor. This suggests that sorghum may contain only one homologous copy of this gene, and perhaps additional ancient paralogs. The wheat gene encoding the proline-rich protein appears to have similar genomic organization (Raines et al., 1991). The high-level gene expression and low copy nature suggest that the promoter of SPRP1 may be especially useful for driving transgene expression in sugarcane and other monocots. We have sequenced about 1.8 kb upstream of the translation start site and shown that the sugarcane sequence exhibits the features of promoters of other proline-rich protein genes. Curiously, the genomic nucleotide sequence shows 100% identity with the 3′ untranslated region of the SPRP1 cDNA but only 97% identity to the coding region. Similarly, the nucleotide sequence identity of the 3′ untranslated region between the genomic clone and SPRP2 is 96%, much higher than the sequence identity (83%) of the coding region near the 3′ end. The high level of variation in the SPRP coding region is unexpected since members of most gene families are more conserved in the coding region. We did not obtain sequence for the entire coding region of the genomic clone because of a highly repetitive sequence in the middle of the gene. We did not find any introns in the genomic DNA regions that were sequenced.

We also found EF1α to be highly expressed in suga-
cane, but occurring in several copies. Multigene family organization of this gene has also been shown in Arabidopsis (four copies; Curie et al., 1991, 1993) and maize (six copies; Berberich et al., 1995). The Arabidopsis thaliana A1 promoter has been isolated, its expression pattern determined (Curie et al., 1991, 1993), and its use for directing strong transient expression in transfected protoplasts demonstrated (Axelos et al., 1989; Curie et al., 1991). One member of the gene family encoding the α subunit translation factor and the corresponding genomic clone has been isolated from maize (Berberich et al., 1995). The comparison between our genomic sequence and the maize EF1α genomic sequence indicated that the sugarcane and maize genes have high similarity in the coding region (95%) and less in the intron regions (70%). The maize genomic clones obtained by Berberich et al. (1995) only contained part of the first intron and 5′ untranslated region as well as the entire coding region, but we isolated the entire EF1α gene plus the promoter region. The structure of the EF1α gene in sugarcane is similar to Arabidopsis, both having two introns with one located in the 5′ untranslated region and the other one in the coding region.

Fusion of the isolated promoters to the uidA reporter gene and use of transient GUS expression assays proved an effective method of confirming each promoter’s activity in sugarcane and wheat cells. This is significant, since isolation of promoters from a highly polyploid plant, such as sugarcane, carries increased risk of isolating a nonfunctional promoter from a nonactive copy of the gene. The assay procedure enables a crude (semi-quantitative) assessment of promoter strength compared to a known strong monocot promoter (ubi). This is a useful preliminary screen to select promoters for further testing in stable transformation experiments. The use of scutellar tissue (including early callus) from precultured wheat embryos as target tissue may have biased the assessment of relative promoter strength between promoters showing differing patterns of tissue specificity. Nonetheless, the results show that the SEF1α and SPRP promoters are strongly active in driving transient expression in two important monocot crop species, wheat and sugarcane.

ACKNOWLEDGMENTS

We would like to thank Dr. Yang Si for assistance in library construction, Dr. Veera Padmanabhan for assistance in promoter analysis and BJ Moran for assistance in transient expression assays. This work was supported by the Texas Higher Education Coordinating Board Advanced Technology Program grant nos. 999902-029 and 999902-188 to TEM.

REFERENCES


