Planta (2009) 229:549–558
DOI 10.1007/s00425-008-0852-8

ORIGINAL ARTICLE

# Efficient silencing of reporter transgenes coupled to known functional promoters in sugarcane, a highly polyploid crop species

**Stephen R. Mudge · Kenji Osabe · Rosanne E. Casu · Graham D. Bonnett · John M. Manners · Robert G. Birch**

Received: 24 June 2008 / Accepted: 24 October 2008 / Published online: 15 November 2008


**Abstract** Sugarcane is a crop of great interest for engineering of sustainable biomaterials and biofuel production. Isolated sugarcane promoters have generally not maintained the expected patterns of reporter transgene expression. This could arise from defective promoters on redundant alleles in the highly polyploid genome, or from efficient transgene silencing. To resolve this question we undertook detailed analysis of a sugarcane gene that combines a simple pattern in genomic Southern hybridization analysis with potentially useful, sink-specific, expression. Sequence analysis indicates that this gene encodes a member of the SHAQYF subfamily of MYB transcription factors. At least eight alleles were revealed by PCR analysis in sugarcane cultivar Q117 and a similar level of heterozygosity was seen in BAC clones from cultivar Q200. Eight distinct promoter sequences were isolated from Q117, of which at least three are associated with expressed alleles. All of the isolated promoter variants were tested for ability to drive reporter gene expression in sugarcane. Most were functional soon after transfer, but none drove reporter activity in mature stems of regenerated plants. These results show that the ineffectiveness of previously tested sugarcane promoters is not simply due to the isolation of non-functional promoter copies from the polyploid genome. If the unpredictable onset of silencing observed in most other plant species is associated with developmental polyploidy, approaches that avoid efficient transgene silencing in polyploid sugarcane are likely to have much wider utility in molecular improvement.

**Keywords** Allelic variation · Promoter function · *Saccharum* · Transgene silencing

**Abbreviations**

| | |
|---|---|
| BAC | Bacterial artificial chromosome |
| EST | Expressed sequence tag |
| GUS | *β*-Glucuronidase |
| LUC | Firefly luciferase |
| *Nos* | *Nopaline synthase* |
| SNP | Single nucleotide polymorphism |
| *Ubi-1* | *Ubiquitin-1* |

## Introduction

In practical plant improvement, it is widely observed that only a small proportion of transgenic events show stable expression of the introduced gene in precisely the intended pattern to confer the added trait in the absence of other undesired changes (Birch 1997). This has been true even for input traits such as insect and herbicide resistance that need only coarse control over the level of constitutive transgene expression (Finnegan and McElroy 1994). The required precision is expected to be higher in metabolic engineering that diverts metabolic flows into novel and high-value products, because diversion of these flows in inappropriate tissues will interfere with plant development.

Sugarcane is an attractive crop for improvement aimed at sustainable biomaterial production, because of high

**Electronic supplementary material** The online version of this article (doi:10.1007/s00425-008-0852-8) contains supplementary material, which is available to authorized users.

S. R. Mudge (✉) · K. Osabe · R. G. Birch
Botany Department, School of Integrative Biology,
The University of Queensland, Brisbane 4072, Australia
e-mail: s.mudge@uq.edu.au

S. R. Mudge · R. E. Casu · G. D. Bonnett · J. M. Manners
CSIRO Plant Industry, Queensland Bioscience Precinct,
306 Carmody Road, St Lucia 4067, Australia

biomass productivity, built-in containment features and an efficient gene transfer system (Birch 2007). However, its unusual genetic feature of highly stable polyploidy poses some challenges. Modern cultivars are highly polyploid, aneuploid hybrids between *Saccharum officinarum* L (octoploid, with $2n = 80$ chromosomes) and *S. spontaneum* L (ploidy level of 5–16, with $2n = 40$–$128$ chromosomes). Sugarcane cultivars typically have $2n = 100$–$130$ chromosomes, of which approximately 15–20% are derived from *S. spontaneum* and 5% are recombinants derived from both species (D'Hont et al. 1996; Cuadrado et al. 2004). Therefore, sugarcane cultivars have approximately 10–13 homo(eo)logous copies of most loci.

Molecular marker studies and sequencing of genomic PCR products indicate high heterozygosity (Jannoo et al. 1999; Cordeiro et al. 2000). Sequences within the sugarcane expressed sequence tag (EST) databases indicate that multiple haplotypes are expressed (Grivet et al. 2001, 2003), but little is known about the differential expression of alleles.

This complexity of the sugarcane genome has implications for the isolation of functional sugarcane promoters. Several sugarcane promoters have been isolated and shown to drive transgene activity in callus or young plants but not in the expected pattern in mature plants (Hansom et al. 1999; Wei et al. 2003; Yang et al. 2003). Several heterologous and synthetic promoters have also been shown to drive strong expression in callus but little or no expression in mature plants, with evidence for both transcriptional and post-transcriptional gene silencing (PTGS) effects (Hansom et al. 1999). In contrast, the maize Ubi-1 promoter has driven sustained reporter transgene expression in sugarcane (Hansom et al. 1999).

It is possible that the previously isolated sugarcane promoters simply corresponded to defective promoter copies from the highly polyploid genome. We set out to test this possibility by isolating a full suite of promoter alleles for a sugarcane gene that is expressed predominantly in the mature stems. We established the allelic variation for this gene, identified the expressed alleles, isolated the promoter that drives the most strongly expressed allele and tested its ability to drive transgene expression in sugarcane. This promoter did not drive detectable reporter gene activity in mature stems of transgenic plants, whereas the native allele continued to be expressed.

From these results, efficient transgene silencing is an important limitation to facile molecular improvement in sugarcane. The clarity of transgene silencing in this stable polyploid provides a powerful experimental system in which to understand mechanisms and develop solutions that are likely to be widely applicable in genetically 'simpler' crops, which have the problem of unpredictable onset of silencing associated with developmentally changing ploidy levels through endoreduplication.

## Materials and methods

### Choice of target locus

Sugarcane EST clones that were potentially up-regulated in mature stems, based on microarray data (Casu et al. 2004), were characterized in greater detail by sequencing and by Northern analysis of expression patterns. A sugarcane gene designated *ScR1MYB1* (Genbank accession EU719199) that is up-regulated in mature stems was identified as detailed below.

### RNA extraction and Northern analysis

RNA was extracted from tissues of field- or glasshouse-grown sugarcane (*Saccharum* hybrid) cultivar Q117 (obtained from BSES Meringa Sugar Experiment Station, Gordonvale, QLD, Australia) by CsCl ultracentrifugation (Chirgwin et al. 1979). Following denaturation, approximately 10 µg RNA was fractionated on a 1.2% TBE agarose gel, and transferred onto Hybond $N^+$ membrane (Amersham) using capillary transfer with $10 \times$ SSC. Hybridization was done for 16 h at 65°C in 0.5 M $NaH_2PO_4$, 1 mM EDTA, 7% SDS; using $^{32}$P-dCTP-labeled probes prepared using a Megaprime kit (Amersham). Following hybridization, membranes were washed to high stringency ($0.1 \times$ SSC; 0.1% SDS at 65ºC), and analyzed using a PhosphorImager SI (Molecular Dynamics).

### DNA extraction and Southern analysis

Sugarcane genomic DNA was isolated using a CTAB-based method (Rogers and Bendich 1988). Ten microgram of each sample was digested with restriction enzymes, fractionated on a 0.8% agarose TAE gel, and transferred onto Hybond $N^+$ membrane (Amersham) using capillary transfer with either 0.4 M NaOH or $10 \times$ SSC. Labeling of probes and hybridization conditions were as described for Northern analysis.

### DNA sequencing

DNA sequencing of plasmid templates was done using a BigDye Terminator 3.1 DNA sequencing kit (Applied Biosystems), and separations by the Australian Genomic Research Facility (Brisbane, QLD, Australia).

### Analysis of allelic variation and expression

The degree of allelic variation for ScR1MYB1 was assessed by cloning and sequencing multiple PCR products amplified from Q117 using a forward primer beginning at the start codon (MYB F: 5′-ATGGCTAGGAAATGCTCC

AG-3′) and a reverse primer beginning at the stop codon (MYB R: 5′-TTATTCTCCTAGACGCCAGT-3′). Similarly, RT-PCR products derived from mature stem (amplified using the above primer set or the MYB R primer in combination with MYBF220: 5′-GCCTACTAYGGAG CTGCTGC-3′) were sequenced to determine which alleles are expressed. For RT-PCR, cDNA was synthesized using Superscript III (Invitrogen) and oligo dT primer using RNA from internodes 11–12 or 30. In all cases, the PCR was done using Expand High Fidelity polymerase (Roche) according to the manufacturer's instructions, and all amplified products were cloned into pGEM-T Easy (Promega) prior to sequencing.

### Isolation of upstream promoter regions and PCR amplification

Promoter alleles corresponding to *ScR1MYB1* were amplified using a PCR-based Genome Walker strategy (Clontech), using primers GSP1 (5′-ATTGTTTCCACAA CTGGAGCATTTCCTAGC-3′) and GSP2 (5′-AACAA GGGCACAAGAAGTTGCGGTGTAGTA-3′) derived from the 5′ end of the *ScR1MYB1* transcript. To maximize the chances of amplifying all promoter alleles in the polyploid genome of cultivar Q117, we used a total of eight GenomeWalker libraries (prepared using restriction enzymes *Dra*I, *Eco*RV, *Pvu*II, *Sca*I, *Stu*I, *Bsr*BI, *Ecl*136II, and *Ssp*I). These and all other PCR amplification reactions used Expand High Fidelity polymerase (Roche) according to the manufacturer's instructions. Amplified products were cloned into pGem-T Easy (Promega) prior to sequencing. The transcriptional start site was determined by 5′RACE using the GeneRacer Kit (Invitrogen) according to the manufacturer's instructions.

### Construction of promoter–reporter gene fusions

Promoter sequences with native 5′UTRs were amplified from Genome Walker clones, by use of PCR primers designed to incorporate *Pst*I and *Nco*I restriction sites. These restriction sites enabled cloning of the promoters in front of the GUS*Plus* (CAMBIA) and *luc*$^{+}$NF (Promega) reporter genes. Constructs containing the same reporter genes driven by the maize *Ubi-1* promoter (Christensen et al. 1992) were included as positive controls. All constructs included the *nos* terminator.

### Sugarcane transformation

Transformation and regeneration of sugarcane cultivar Q117 (obtained from BSES Meringa Sugar Experiment Station) was done as previously described (Bower et al. 1996). Plants were grown in a containment glasshouse under natural light intensity, at 28°C and watered twice a day. Each plant was grown as a single stalk in a pot of 20 cm diameter and a density of 18 pots/m$^2$, fertilized with Osmocote® at 5 g/month for the first and the second months, then 10 g/month. Leaves were numbered from one for the top visual dewlap (TVD) with higher numbers for older leaves. Internodes were numbered according to the leaf attached to the node immediately above.

### Reporter gene assays

For histochemical GUS staining, tissues were vacuum infiltrated in assay buffer (50 mM sodium phosphate, pH 7.0; 0.1% Triton X-100, 0.5 mM potassium ferrocyanide, 0.5 mM potassium ferricyanide, 10 mM EDTA, 0.05% 5-bromo-4-chloro-3-indolyl glucuronide) and incubated at 37°C for 18 h. Green tissues were destained with ethanol. For fluorimetric GUS assays (Jefferson 1987), fluorescence measurements were taken at 0, 60, 120 and 180 min timepoints using a Labsystems Fluoroskan Ascent microplate reader (Pathtech Diagnostics, Melbourne, VIC, Australia). LUC assays were done as described (Luehrsen and Walbot 1993), using a BMG POLARstar OPTIMA luminometer.

## Results

### Identification of a novel *MYB* gene that is up-regulated in mature stem

EST clone MCSA063B04 was identified as putatively up-regulated in mature versus immature stem based on microarray data (Casu et al. 2004). Initial sequencing of this clone revealed a putative methionine sulfoxide reductase (Genbank accession number CF572917), which we found by Northern analysis not to be up-regulated in mature stems (data not shown). Sequencing of the full insert in the EST clone revealed the presence of likely concatenated cDNAs, of which one comprised a full open reading frame with a predicted translation product similar to SHAQKYF R1MYB transcription factors. This transcript was shown by Northern analysis to be strongly up-regulated in mature versus immature stem (Fig. 1a). It was also up-regulated in mature roots (Fig. 1b), and it was not detected in leaf tissue. Northern analysis using RNA from vascular, parenchyma and rind tissue dissected from internode eight of field-grown Q117 (Rae et al. 2005) revealed the presence of transcript in all three tissue types within the stem (Fig. 1c). The detection of multiple bands by Northern analysis reflects alternative splicing of the transcript (see below).

Southern analysis using the same probe revealed a relatively simple banding pattern by sugarcane standards (Fig. 2), given that the restriction enzymes *Eco*RI, *Hin*dIII

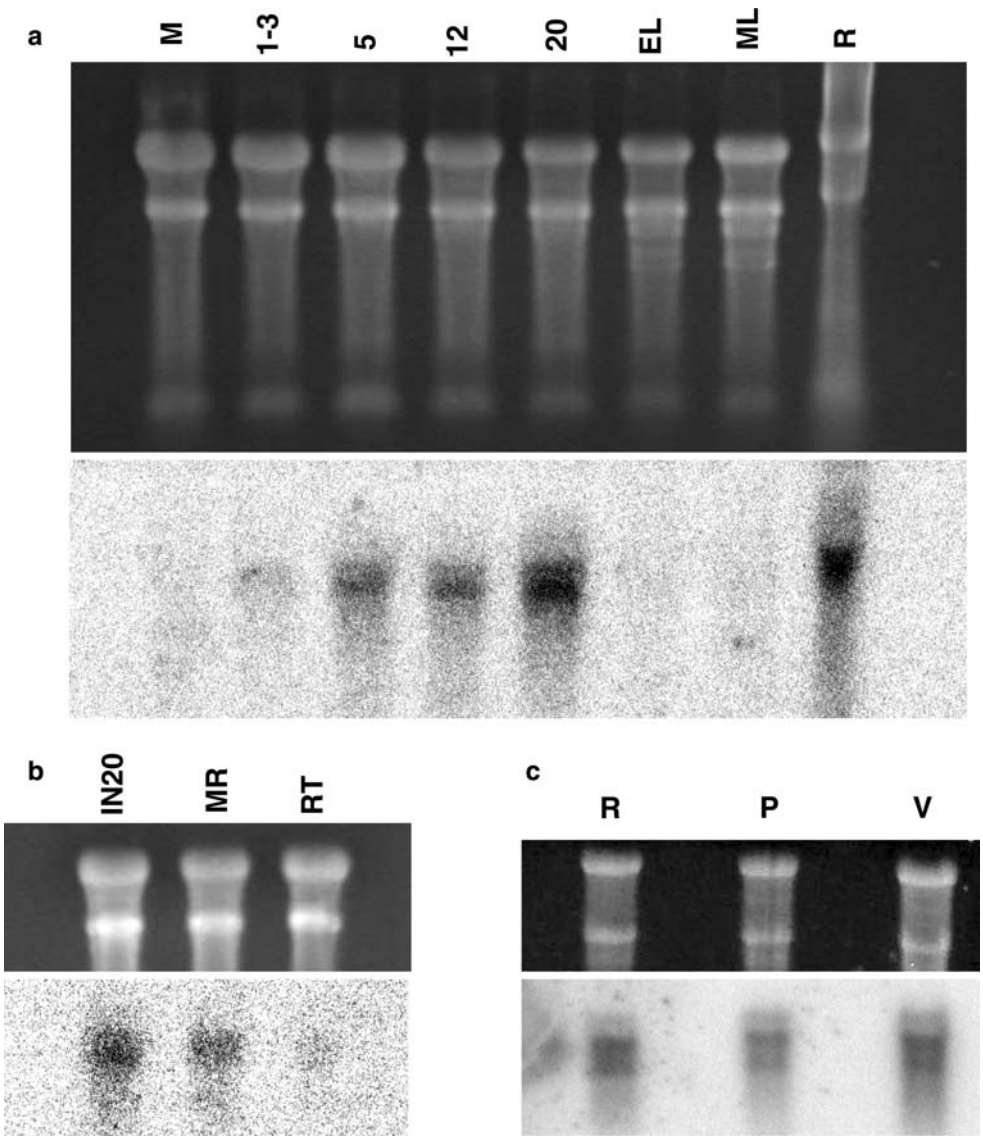


**Fig. 1** Expression of *ScR1MYB1* in different sugarcane tissues. **a** Northern analysis using total RNA isolated from meristems (*M*), internodes 1 to 3 (1–3), internode 5 (5), internode 12 (12), internode 20 (20), expanding leaf (*EL*) and mature leaf (*ML*) from field-grown Q117 plants, plus root (*R*) from glasshouse-grown Q117 plants, probed with the 838 bp *Hin*dIII fragment of the *ScR1MYB1* cDNA. **b** Northern analysis using total RNA from internode 20 (IN20), mature roots (*MR*) and root tips (*RT*) from glasshouse-grown Q117 plants, probed with the 838 bp *Hin*dIII fragment of the *ScR1MYB1* cDNA. **c** Northern analysis using total RNA from dissected rind (*R*), storage parenchyma (*P*) and vascular strands (*V*) from internode eight of field-grown Q117 plants, probed with the full insert of EST clone MCSA063B04

and *Sac*I all cut once within the region covered by the probe. The sugarcane gene locus corresponding to this isolated nucleotide sequence is hereby designated as *ScR1MYB1*.

## Sugarcane is highly heterozygous at the ScR1MYB1 locus

Allelic variation within cv. Q117 was analyzed by cloning and sequencing multiple amplicons obtained from genomic PCR and RT-PCR. These amplicons contained either the full ScR1MYB1 coding region (~1.3 and 0.9 kb products for genomic and RT-PCR, respectively), or a product lacking the first 220 nt of exon 1. Thirty-two sequence variants were detected among 51 genomic PCR products from five independent amplifications, and 106 RT-PCR products from four amplifications on RNA from mature stem (internode 12). However, in control reactions on a known template we detect approximately 1 error per 1.3 kb amplicon under the same PCR conditions. PCR recombination can also lead to an overestimation of allelic diversity (Cronn et al. 2002; Yu et al. 2006). Using a conservative approach of counting only sequence variants detected multiple times in independent PCRs, we obtained a total of eight haplotypes from Q117. Evidence from genomic Southern analysis, BAC library screening, and analysis of the corresponding locus in *Sorghum bicolor* (see below), indicates that these haplotypes most likely represent alleles at a single locus. An alignment of the genomic nucleotide sequences of these eight alleles is shown in Suppl. Fig. 1. The frequency of single nucleotide polymorphisms (SNPs) detected in introns and exons was similar (perhaps reflecting the alternative splicing of the larger intron 1; see below). Approximately 40% of SNPs lead to an amino acid substitution, and the eight putative alleles encode different predicted proteins (Fig. 3). Substantial variability towards the downstream end of the

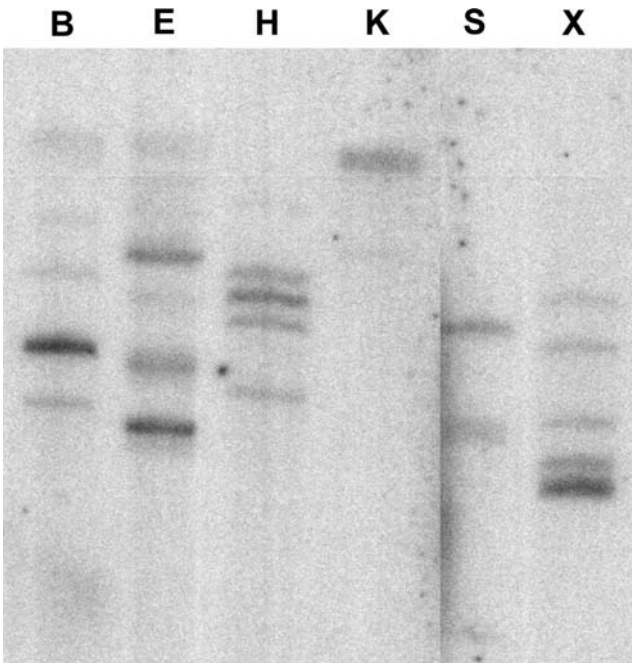


**Fig. 2** Genetic complexity of *ScR1MYB1* in sugarcane, as revealed by Southern analysis. Genomic DNA from cultivar Q117 was digested with the restriction enzymes *Bam*HI (*B*), *Eco*RI (*E*), *Hin*dIII (*H*), *Kpn*I (*K*), *Sac*I (*S*) and *Xba*I (*X*), and probed with the 838 bp *Hin*dIII fragment of the *ScR1MYB1* cDNA

coding region resulted in predicted protein lengths ranging from 292 to 313 amino acids.

As an alternative measure of allelic variation in sugarcane, we isolated and sequenced multiple clones containing the full coding region from a genomic BAC library of cultivar Q200 (thereby avoiding any PCR-derived artifacts). We screened approximately 95,000 clones with an average insert size of 115 kb (S. Brumbley, BSES, Brisbane, QLD, personal communication). Assuming a non-replicated somatic cell (2C) genome size of 6,400 Mb (Edme et al. 2005) and using the formula of Clarke and Carbon (1976), there is an 82% probability of any Q200 gene being present this library, but only a 14% probability of 10 homo(eo)logous copies of this sequence being present. We identified six alleles among 13 hybridizing clones in this library (Table 1). An additional allele (X2) is represented several times in the pool of cloned RT-PCR products from mature stem of this cultivar. Thus, at least seven alleles are present in Q200. Several alleles appear to be specific to each cultivar.

## Multiple ScR1MYB1 alleles are expressed in sugarcane

The frequency of occurrence among sequenced RT-PCR products indicates that transcripts of the Z1 allele predominate (40%) in internode 12 tissue of Q117, with three other alleles each contributing ≥10% (Table 2). Interestingly, the same alleles predominate among Q200 transcripts. These cultivars do not have the same parents, but do share common ancestors in earlier generations.

Sequence analysis of the RT-PCR products also reveals the presence of alternatively spliced transcripts. In ScR1MYB1 and closely related sequences from the rice and *Arabidopsis* genomes, there are two introns, with the position of intron 1 being highly conserved and the position of intron 2 being less conserved. In approximately one-third (34/106) of RT-PCR products from Q117, intron 1 was unspliced, and in rare cases (2/106 clones), a small part of intron 1 remained (Suppl. Fig. 2). Similarly, in Q200 40% (26/65) of all RT-PCR products were unspliced. In all cases, intron 2 was fully spliced. The presence of all or part of intron 1 in these transcripts results in a smaller predicted protein (152 and 124 amino acids, respectively), that does not include any of the MYB domain but does include the N-terminal zinc finger and serine-rich domains (Fig. 3). The presence of both spliced and unspliced transcript explains the double band observed on the Northern blot probed with the ScR1MYB1 probe (Fig. 1).

## ScR1MYB1 promoter alleles fail to drive reporter transgene expression in mature stems

The presence of a full open reading frame for ScR1MYB1 in our cDNA clone facilitated amplification of upstream promoter sequences. Using the GenomeWalker single-specific-primer PCR procedure, we identified a total of eight putative promoter alleles among 41 cloned products of at least 1 kb in size. Three of these (A1, A2 and B) were identified multiple times in multiple PCRs. An additional three (A3, A6 and A5) had unique insertions or deletions (of 240, 18 bp and at least 1.2 kb, respectively) unlikely to arise from PCR errors (Suppl. Fig. 3). The remaining two alleles (A4 and A7) differ only by a few SNPs. They were recovered because of rare events in which the adapter primer ligated to a random blunt end rather than the blunt restriction site used to construct the library.

These eight putative promoter alleles show 99% identity for approximately 280 bp upstream of the start codon. A putative TATA box predicted using the program HCtata (Milanesi et al. 1996) occurs approximately 133–135 bp upstream of the start codon, and 23 bp upstream of the transcription start site determined by 5′-RACE. Beyond 280 bp upstream, the sequence of the B allele diverges and approximately 60 bp further upstream, the A5 sequence diverges, suggesting that major rearrangements have occurred in these promoter alleles.

To investigate the relationship between promoter alleles and coding sequence alleles, we have amplified across the promoter:coding sequence junction using either conserved or allele-specific PCR primers in both the promoter and

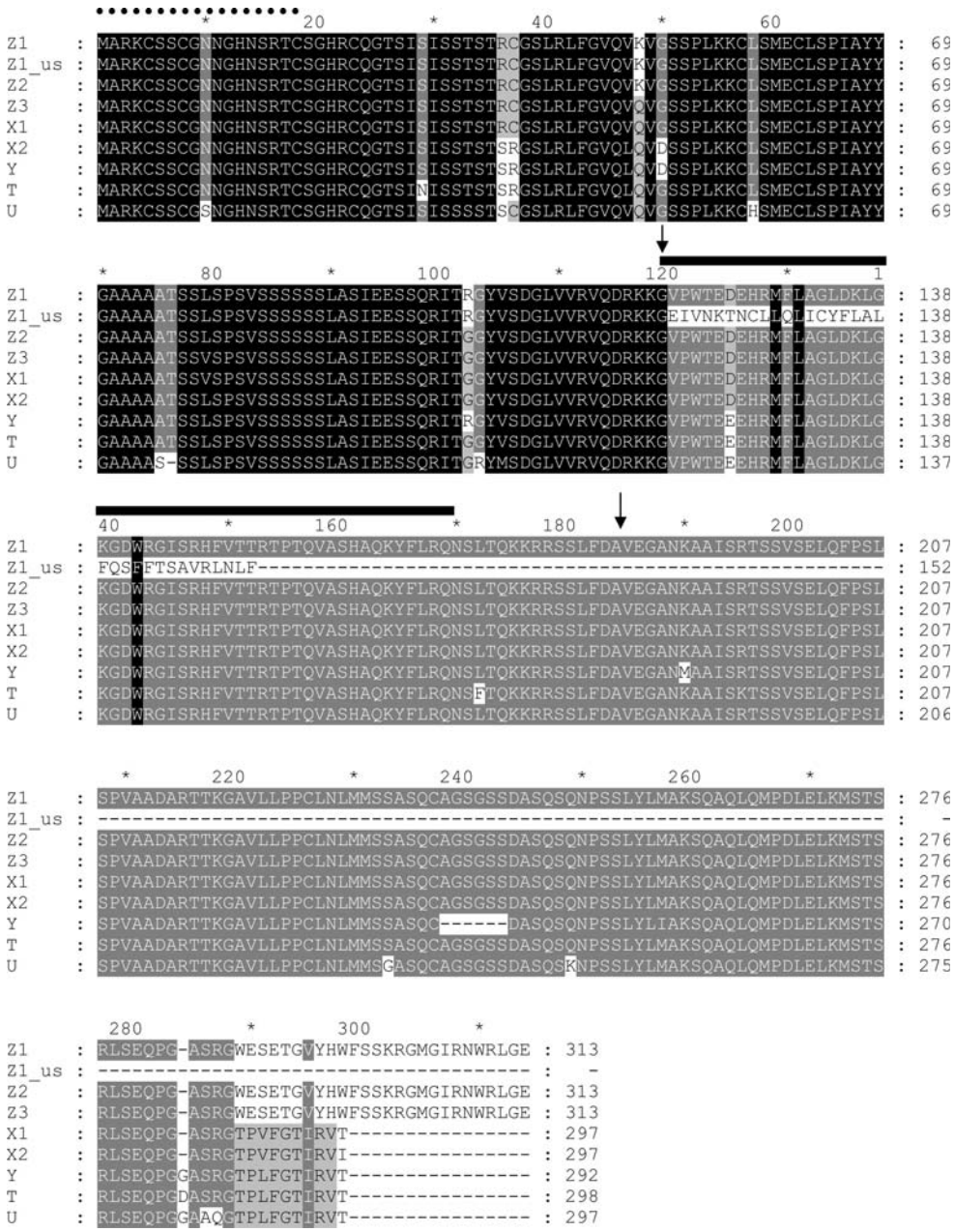


**Fig. 3** Alignment of the proteins encoded by the eight putative alleles of *ScR1MYB1* identified in cultivar Q117. The MYB DNA binding domain is marked by a *solid bar* above, the CCHC-type zinc finger is marked by a *dashed line*, and the positions of introns 1 and 2 are marked by *vertical arrows*. "*Z1 us*" represents the sequence encoded by the partially spliced Z1 transcript

coding sequence. Among 16 clones generated with a Z1 specific primer, all contained the Z1 coding sequence linked to the A1 promoter. All five clones containing Z1 from amplifications using conserved primers were also linked to A1, indicating that the A1 promoter allele is uniquely linked to the most strongly expressed (Z1) coding sequence in Q117. In Q200 BAC clones, the Z1 and Z2 coding sequences were always linked to the A1 promoter sequence. The A1 promoter:Z1 coding sequence was identified five times among the 13 BAC clones, suggesting that the high proportion of Z1 transcripts corresponds with a high genomic copy number of this allele in Q200. We can conclude that the A1 promoter is functional for *ScR1Myb1* expression in both cultivars.

Our analysis also indicates that the X1 transcribed sequence is linked to both the A4 and B promoter sequences in Q117, and thus either of both of these promoters is likely to be functional in their native state. At present we have not identified the promoter(s) linked to the Z2 or X2 transcribed alleles in Q117.

At least ten independent transgenic lines containing each of the eight promoter alleles fused to the LUC and/or GUS-*Plus* reporter genes were assayed for reporter gene activity. All eight promoters drove GUS activity in some regenerating shoots in tissue culture. The frequency at which this activity was observed varied, even between clusters of shoots of the same transgenic line. No reporter gene activity was detected in stem internodes 3, 6, 12, or 15 from

**Table 1** Identity of clones obtained from Q200 BAC library, showing the promoter and coding sequence present in each clone, plus the frequency at which the clone was identified in the library

| Allele (promoter:coding sequence) | Number of clones |
|---|---|
| A1:Z1 | 5 |
| A1:Z2 | 1 |
| A4:X1 | 2 |
| A10:U3 | 2 |
| A9:U2 | 1 |
| A11:S | 2 |

glasshouse-grown plants with approximately 20 internodes, using either the fluorimetric GUS assay (<10 pmol $min^{-1}$ $mg^{-1}$ protein) or a luminometer LUC assay (0 RLU $mg^{-1}$ protein). In the same assay *Ubi*-GUS*Plus* lines typically showed >200 pmol $min^{-1}$ $mg^{-1}$ protein and *Ubi*-LUC lines showed >50,000 RLU $mg^{-1}$ protein in mature stem extracts.

Southern analysis of A1 promoter:GUS*Plus* transgenic lines showed a range of integration complexities and presence of an intact expression cassette (data not shown), typical for this routine sugarcane transformation method (Bower et al. 1996). The lack of reporter gene activity in mature stems was not due to low co-transformation rates.

In the course of PCR analysis of the promoter:coding sequence links, we revealed two additional promoters (A8 and A9) linked to weakly expressed alleles (T and U) which comprise ~1% of the transcripts from the *ScR1MYB1* locus in Q117. The A8 and A9 upstream sequences have not been tested for promoter activity using reporter genes in transgenic plants.

### The lack of reporter gene activity does not reflect changes in native promoter activity

Northern analysis of the mature transgenic plants using the ScR1MYB1 probe indicates that the lack of reporter gene activity is not accompanied by reduced expression of the native transcript (Fig. 4). Furthermore, cloning and sequencing of 60 RT-PCR amplicons revealed the same four alleles predominantly expressed in a tested A1 promoter:GUS*Plus* transgenic line as in the non-transformed controls described above (Table 2). Thus, the native A1 promoter function appears undiminished in this transgenic line, despite failure of reporter gene activity to be sustained from an introduced copy of the same promoter-5′ leader sequence.

## Discussion

The failure of previously isolated sugarcane promoters to drive the anticipated reporter activity patterns in mature transgenic sugarcane may be due to the isolation of defective promoter alleles from the highly polyploid sugarcane genome. To test this, we aimed to isolate and test the full suite of promoter alleles for a sugarcane gene. We chose a gene that is expressed predominantly in mature stems. This gene proved to encode a member of the SHAQKYF subfamily of R1MYB transcription factors.

**Table 2** Summary of allele expression patterns in mature stem (internode 12) of non-transformed Q117, Q200 and transgenic Q117 plants, as determined by RT-PCR

| Allele | Q117 | | Q200 | | Q117 *A1*:GUS$^+$ | |
|---|---|---|---|---|---|---|
| | Number of occurrences | Frequency (%) | Number of occurrences | Frequency (%) | Number of occurrences | Frequency (%) |
| Z1 | 42 | 40 | 34 | 52 | 30 | 50 |
| Z2 | 15 | 14 | 4 | 6.1 | 18 | 30 |
| Z3 | 4 | 3.8 | 0 | 0 | 0 | 0 |
| X1 | 12 | 11 | 9 | 14 | 3 | 5 |
| X2 | 11 | 10 | 12 | 18 | 0 | 0 |
| Y[a] | 4 | 3.8 | 0 | 0 | 0 | 0 |
| T[a] | 2 | 1.9 | 0 | 0 | 2 | 3.3 |
| U[a] | 1 | 0.9 | 0 | 0 | 1 | 1.7 |
| U2[a] | 0 | 0 | 2 | 3 | 0 | 0 |
| U3[a] | 0 | 0 | 1 | 1.5 | 0 | 0 |
| S[a] | 0 | 0 | 3 | 4.5 | 0 | 0 |
| Others | 15 | 14 | 1 | 1.5 | 6 | 10 |
| Total | 106 | (100) | 66 | (100) | 60 | (100) |

[a] Identified in a single cultivar

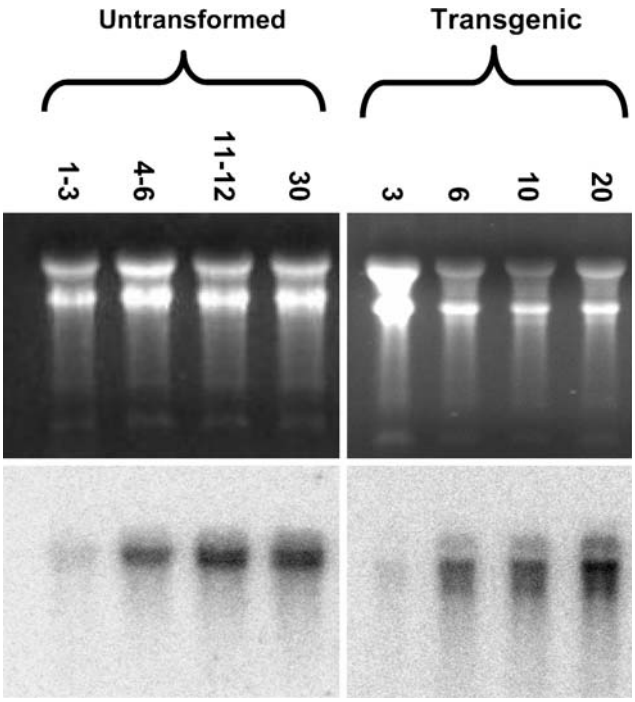


**Fig. 4** Northern analysis of *ScR1MYB1* expression in glasshouse-grown untransformed Q117 (*left*), and Q117 transformed with a *ScR1MYB1* promoter:GUS*Plus* construct (*right*). RNA was extracted from a range of internodes from immature to mature stem, as indicated *above the lane*. The RNA loading is shown in the *upper panel*, and the amount of transcript detected using the 838 bp *ScR1MYB1* probe is shown in the *lower panel*

The first described member of the SHAQKYF subfamily was MYBSt1 from potato (Baranowskij et al. 1994), and many homologues have since been described. These proteins have a single MYB DNA binding domain and an N-terminal CCHC-type zinc finger. Several members of the subfamily have recently been shown to be nuclear-targeted transcriptional activators (Lu et al. 2002; Rubio-Somoza et al. 2006a, b).

Sugarcane is highly heterozygous at the ScR1MYB1 locus, with eight coding region variants and ten putative promoter variants identified in cultivar Q117. The relatively simple Southern hybridization pattern (Fig. 2) and the number of hybridizing clones in a genomic BAC library indicate a single locus in sugarcane. The closely related diploid *Sorghum bicolor* genome (Sbi1 release, June 2008; http://www.phytozome.net/sorghum) also contains a single orthologous locus. Therefore, we believe that the sequence variants (haplotypes) described here are alleles. While the PCR-based approach in polyploids is subject to recombination artifacts (Cronn et al. 2002; Yu et al. 2006), our conservative analysis in Q117 agrees well with the heterozygosity revealed using the Q200 BAC library. This level of heterozygosity means that even at this locus, which was selected based on its apparent simplicity, most copies in the polyploid genome of these cultivars are different when analyzed at the nucleotide level. High levels of heterozygosity have been indicated previously in sugarcane, using isozymes (Glaszmann et al. 1989), restriction fragment length polymorphisms (Jannoo et al. 1999) and microsatellite markers (Cordeiro et al. 2000), although these latter markers were not targeted towards coding regions of the genome. McIntyre et al. (2006) used an RT-PCR approach on a 417 bp region within the coding sequence of a family III sucrose phosphate synthase gene in a commercial cultivar and a *S. officinarum* clone, and estimated five and three genomic alleles, respectively.

To our knowledge, this work on ScR1MYB1 is the first to examine the level of heterozygosity of a sugarcane gene at both the DNA and RNA levels. Multiple alleles are expressed, at diverse levels ranging from 40% to <1% of total transcripts from this locus. Of the six sugarcane *ScR1MYB1* EST sequences in GenBank (release 164), two correspond to the strongly expressed Z1 allele and three correspond to the moderately expressed X1 or X2 alleles identified in the present study, one of which is partially spliced. Previous studies of heterozygosity within sugarcane ESTs have reported similar numbers of putative expressed alleles. For 6-phosphogluconate dehydrogenase gene B (*PgdB*), a minimum of two, four, and six haplotypes were reported in three different cultivars (Grivet et al. 2001), while four and up to six haplotypes were reported for alcohol dehydrogenase genes *Adh1_1* and *Adh2* (Grivet et al. 2003).

A cumulative effect of strong total expression from multiple weak alleles could explain instances of low transgene expression from the promoter of a strongly expressed sugarcane gene (Hansom et al. 1999). Indeed, weak or defective promoter alleles do exist in sugarcane (Table 2). However, in the present study, we have identified the promoter variant that drives a strongly expressed ScR1MYB1 allele in the cultivars analyzed. Surprisingly, this isolated promoter does not drive detectable reporter gene activity in mature stems of transgenic plants, despite the fact that the native copy remains functional in these plants.

One possible reason for the lack of reporter gene activity is that the length of promoter used (1.1 kb) does not contain all of the *cis* regulatory elements required for activity in mature stem. However, the variability in GUS detection in regenerating shoots is more suggestive of the onset of a silencing mechanism that progresses to completely block formation of the transgene product in mature stems. Both PTGS involving targeted degradation of transgene transcripts, and associated promoter methylation have been observed in sugarcane, even in constructs containing the commonly used maize ubiquitin promoter (Hansom et al. 1999; Ingelbrecht et al. 1999; Wei et al. 2003).

We previously have shown that transgene silencing is independent of copy number in sugarcane (Hansom et al.

1999). The present study shows that silencing is not simply triggered by multiple copies of the promoter and 5′ untranslated leader, since the native promoter function is not affected in lines with silenced transgene copies. Evidently, silencing of the transgene does not trigger silencing of the native gene which shares the same (101–103 bp) 5′ UTR sequence. Methods to avoid transgene silencing (Graham et al. 2008) will be important to obtain desired patterns of expression in a high proportion of transformed sugarcane lines.

In other crops, it is common for thousands of transformed lines to be screened to identify a few chance events that escape both transgene silencing and any undesired incidental effects of the transformation process. The problem is exacerbated because the onset of silencing is unpredictable, creating a risk of instability in the introduced trait after commercial release (Finnegan and McElroy 1994; Birch 1997).

There is increasing support for the idea that plants are likely to have integrated silencing mechanisms in the exploitation of polyploidy (Chen and Ni 2006; Barow 2006). In genetically simpler plant species, developmental polyploidy through endoreduplication may actually confound studies into silencing, through patchy and erratic developmental changes in ploidy, with associated cellular switches that are incidentally involved in transgene silencing. The apparent high efficiency of transgene silencing in sugarcane may simply be a clearer early establishment of silencing in this stable polyploid species. This makes sugarcane a good experimental system in which to understand silencing mechanisms and develop practical approaches to reduce the nuisance of transgene silencing, which are likely to be widely applicable in other plants.

The observation that multiple alleles are expressed at medium to high levels also has broader implications for sugarcane breeding. If this is a general phenomenon, as appears likely from the combination of our results and initial analysis of diversity within sugarcane ESTs (Grivet et al. 2001, 2003), then multiple alleles at a locus are likely to contribute to phenotypic traits. If combinations and ratios of alleles at a locus substantially influence phenotype, the application of marker-assisted selection will be much more challenging in sugarcane than in systems where selection is applied for markers linked to a single favorable allele.

**Acknowledgments** This work was funded by the Australian Sugar Research and Development Corporation. We thank Angelo Fallarino (Botany Department, SIB, UQ) for technical assistance, Stevens Brumbley and BSES for providing the Q200 BAC library, Anne Rae (CSIRO Plant Industry) for providing RNA samples from dissected stem tissues, and Chris Grof (CSIRO Plant Industry) for providing the Q117 Genome Walker libraries. Thanks also to Lynne McIntyre (CSIRO Plant Industry) for helpful discussions.

## References

Baranowskij N, Frohberg C, Prat S, Willmitzer L (1994) A novel DNA-binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. EMBO J 13:5383–5392

Barow M (2006) Endopolyploidy in seed plants. Bioessays 28:271–281

Birch RG (1997) Plant transformation: problems and strategies for practical application. Ann Plant Physiol 48:297–326

Birch RG (2007) Metabolic engineering in sugarcane: assisting the transition to a bio-based economy. In: Verpoorte RA, Alfermann W, Johnson TS (eds) Applications of plant metabolic engineering. Springer, Berlin, pp 249–281

Bower R, Elliott AR, Potier BAM, Birch RG (1996) High-efficiency, microprojectile-mediated cotransformation of sugarcane, using visible or selectable markers. Mol Breed 2:239–249

Casu RE, Dimmock CM, Chapman SC, Grof CPL, McIntyre CL, Bonnett GD, Manners JM (2004) Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. Plant Mol Biol 54:503–517

Chen ZJ, Ni ZF (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays 28:240–252

Chirgwin JM, Przybyla AE, Macdonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic-acid from sources enriched in ribonuclease. Biochemistry 18:5294–5299

Christensen AH, Sharrock RA, Quail PH (1992) Maize polyubiquitin genes—structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol 18:675–689

Clarke L, Carbon J (1976) Colony bank containing synthetic Col El hybrid plasmids representative of entire *Escherichia coli* genome. Cell 9:91–99

Cordeiro GM, Taylor GO, Henry RJ (2000) Characterisation of microsatellite markers from sugarcane (*Saccharum sp.*), a highly polyploid species. Plant Sci 155:161–168

Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. Theor Appl Genet 104:482–489

Cuadrado A, Acevedo R, de la Espina SMD, Jouve N, de la Torre C (2004) Genome remodelling in three modern *S-officinarum x S-spontaneum* sugarcane cultivars. J Exp Bot 55:847–854

D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp) by molecular cytogenetics. Mol Gen Genet 250:405–413

Edme SJ, Comstock JC, Miller JD, Tai PYP (2005) Determination of DNA content and genome size in sugarcane. J Am Soc Sugar Cane Technol 25:1–16

Finnegan J, McElroy D (1994) Transgene inactivation: plants fight back!. Bio/Technol 12:883–888

Glaszmann JC, Fautret A, Noyer JL, Feldmann P, Lanaud C (1989) Biochemical genetic-markers in sugarcane. Theor Appl Genet 78:537–543

Graham MW, Mudge SR, Sternes PR, Birch RG (2008) Understanding and avoiding transgene silencing. In: Stewart CN, Touraev A, Citovsky V, Tzfira T (eds) Plant transformation technologies. Wiley-Blackwell, New York, (in Press)

Grivet L, Glaszmann JC, Arruda P (2001) Sequence polymorphism from EST data in sugarcane: a fine analysis of 6-phosphogluconate dehydrogenase genes. Genet Mol Biol 24:161–167

Grivet L, Glaszmann JC, Vincentz M, da Silva F, Arruda P (2003) ESTs as a source for sequence polymorphism discovery in sugarcane: example of the *Adh* genes. Theor Appl Genet 106:190–197

Hansom S, Bower R, Zhang L, Potier B, Elliott A, Basnayake S, Cordeiro G, Hogarth DM, Cox M, Berding N, Birch RG (1999) Regulation of transgene expression in sugarcane. In: Proceedings of XXIII International Society Sugar Cane Technologists Congress, New Delhi, India, 22–26 February 1999. vol. 2. pp 278–290

Ingelbrecht IL, Irvine JE, Mirkov TE (1999) Posttranscriptional gene silencing in transgenic sugarcane. Dissection of homology-dependent virus resistance in a monocot that has a complex polyploid genome. Plant Physiol 119:1187–1197

Jannoo N, Grivet L, Seguin M, Paulet F, Domaingue R, Rao PS, Dookun A, D'Hont A, Glaszmann JC (1999) Molecular investigation of the genetic base of sugarcane cultivars. Theor Appl Genet 99:171–184

Jefferson R (1987) Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep 5:387–405

Lu C-A, Ho TD, Ho S-L, Yu SM (2002) Three novel MYB proteins with one DNA binding repeat mediate sugar and hormone regulation of a-amylase gene expression. Plant Cell 14:1963–1980

Luehrsen K, Walbot V (1993) Firefly luciferase as a reporter for plant gene expression studies. Promega Notes 44:24–29

McIntyre CL, Jackson M, Cordeiro GM, Amouyal O, Hermann S, Aitken KS, Eliott F, Henry RJ, Casu RE, Bonnett GD (2006) The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. Mol Breed 18:30–50

Milanesi L, Muselli M, Arrigo P (1996) Hamming clustering method for signals prediction in 5′ and 3′ regions of eukaryotic genes. Comput Appl Biosci 12:p399–p404

Rae AL, Grof CPL, Casu RE, Bonnett GD (2005) Sucrose accumulation in the sugarcane stem: pathways and control points for transport and compartmentation. Field Crops Res 92:159–168

Rogers S, Bendich A (1988) Extraction of DNA from plant tissues. In: Plant molecular biology manual. Kluwer Academic Publishers, Dordrecht, pp 1–10

Rubio-Somoza I, Martinez M, Abraham Z, Diaz I, Carbonero P (2006a) Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. Plant J 47:269–281

Rubio-Somoza I, Martinez M, Diaz I, Carbonero P (2006b) HvMCB1, a R1MYB transcription factor from barley with antagonistic regulatory functions during seed development and germination. Plant J 45:17–30

Wei HR, Wang ML, Moore PH, Albert HH (2003) Comparative expression analysis of two sugarcane polyubiquitin promoters and flanking sequences in transgenic plants. J Plant Physiol 160:1241–1251

Yang MZ, Bower R, Burow MD, Paterson AH, Mirkov TE (2003) A rapid and direct approach to identify promoters that confer high levels of gene expression in monocots. Crop Sci 43:1805–1813

Yu W, Rusterholtz KJ, Krummel AT, Lehman N (2006) Detection of high levels of recombination generated during PCR amplification of RNA templates. Biotechniques 40:499–507