Plant Biotechnology Journal

Volume 6 | Issue 9
December 2008

WILEY-
BLACKWELL

www.plantbiotechjournal.com | ISSN 1467–7644

Plant Biotechnology Journal (2008) **6**, pp. 887–896 doi: 10.1111/j.1467-7652.2008.00369.x

Matrix attachment region elements have small and variable effects on transgene expression and stability in field-grown *Populus*

Jingyi Li[1,†], Amy M. Brunner[2], Richard Meilan[3] and Steven H. Strauss[1,*]

[1]*Department of Forest Science, Oregon State University, Corvallis, OR 97331-5752, USA*
[2]*Department of Forestry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0324, USA*
[3]*Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907-2061, USA*

Received 1 August 2007;
revised 20 June 2008, 25 June 2008;
accepted 26 June 2008.
**Correspondence* (fax 541-737-3093;
e-mail Steve.Strauss@oregonstate.edu)
†*Present address*: Department of Plant and Microbial Biology, The University of California, Berkeley, CA 94720-3102, USA

Keywords: dormancy, expression stability, gene silencing, genetic engineering, poplar, transformation, vegetative propagation.

Summary

Matrix attachment regions (MARs) are thought to buffer transgenes from the influence of surrounding chromosomal sequences, and therefore to reduce transgene silencing and variation in expression. The statistical properties of more than 400 independent transgenic events produced in *Populus*, with and without flanking MAR elements from the tobacco root gene *RB7*, were analysed. The expression of two reporter genes in two poplar clones during three phases of vegetative growth, and the association of T-DNA characteristics with expression, was examined. It was found that MARs did not show a consistent effect on transgene expression levels; they had no effect on the green fluorescent protein (*GFP*) reporter gene, but reduced expression in the Basta resistance (*BAR*) reporter gene by 23%. The presence of MARs reduced expression variability within transformant populations, apparently by reducing the number of silenced or weakly expressing events. Transgene expression was highly stable over vegetative growth cycles that spanned 3 years of growth in the glasshouse and field, but MARs showed no association with the strength of correlations in expression over the years. Nonetheless, MARs increased the correlation in expression between a p35S::*GFP* and p*rbcS*::*BAR* transgene linked on the same vector, but the effect was small and varied between the years. The presence of MARs had no effect on the transgene copy number, but was positively associated with T-DNA truncations, as well as with the formation of direct over inverted repeats at the same chromosomal locus.

Introduction

A number of studies have investigated the effects of matrix attachment regions (MARs) on transgene expression and stability (Allen *et al*., 2000; Miguel *et al*., 2007). This work has been motivated by the wide variation in expression of transferred genes. Both the level and pattern of transgene expression show extensive variation in independent insertion events, and expression can become modified over time or silenced entirely. Variation in transgene expression is generally attributed to position effects associated with insertion sites (Kohli *et al*., 2003). When transgenes are integrated in or near to transcriptionally inert regions, such as repetitive DNA and heterochromatin, transgenes are prone to be silenced. When transgenes are inserted near transcriptional enhancers, expression is likely to be increased in particular tissues.

MARs are operationally defined as DNA elements that bind specifically to the nuclear matrix *in vitro* (Allen *et al*., 2000). They have been proposed to buffer genes from the effects of surrounding chromatin on gene expression. Most previous studies have shown that the presence of MARs results in higher expression in transformant populations (Allen *et al*., 1993, 1996; Han *et al*., 1997; Petersen *et al*., 2002; Maximova *et al*., 2003), although the degree appears to depend on the transformation method, the transgene studied and the origin of MARs. The effects of MARs on the variability of transgene expression within populations of independent transformants are less consistent (summarized in Allen *et al*., 2000); MARs

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

either reduce expression variability or have a very limited or undetectable effect. Most studies of gene silencing have shown that MARs tend to reduce or prevent the occurrence of transgene silencing (Conner *et al*., 1998; Ulker *et al*., 1999; Vain *et al*., 1999; Brouwer *et al*., 2002; Maximova *et al*., 2003). For example, MARs reduce the loss of β-glucuronidase (*GUS*) expression in transgenic tobacco plants from one generation to the next during sexual reproduction (Ulker *et al*., 1999). The chicken A1 element has been shown to stabilize *GUS* expression over time in both hemizygous and homozygous plants in three independent events of transgenic tobacco (Mlynarova *et al*., 2003). After removal of the A1 element, the *GUS* gene was silenced over time in two of the three events.

Despite these positive effects of MAR elements on transgene expression and/or stability, most studies have used a small number of transgenic events (4–30), restricting the ability to draw strong inferences about their benefits for a large transgenic population, as may be produced in a functional genomics or applied biotechnology programme. Many plant species and genotypes remain difficult to transform; as a result of the substantial sizes of dual MAR elements when added to transgenic vectors, the inclusion of flanking MAR elements is undesirable unless they provide substantial benefits. In trees, for which stable transgene expression is required over decades – especially for biosafety-related genes such as those that reduce fertility (Brunner *et al*., 2007) – MARs may have the potential to enable technology acceptance by regulatory authorities. However, MARs in trees have received only very limited and short-term study (Han *et al*., 1997). In this study, we found that the expression level of poplar was remarkably stable over several years, but was not associated with the presence of MAR elements. MARs did, however, tend to reduce variability in expression among transformants, improve the coordination of gene expression in linked transgenes, and improve the structure and integrity of transgenic loci. Unfortunately, these statistically beneficial effects were not large and were inconsistent, suggesting that it may not be worthwhile to add MAR elements to transgenic constructs for many experimental and technological goals.

Results

Effect of MARs on expression level, variance and stability over time

To examine the effect of the tobacco *RB7* MAR on transgene expression, the MARs were inserted flanking the T-DNAs in one of the two transgenic constructs (Figure 1). A total of 1616 transgenic poplar plants derived from 404 independent transgenic events were studied for their green fluorescent protein (*GFP*) and Basta resistance (*BAR*) gene expression levels for three consecutive years (2003–2005). The measurements in 2003 were performed in the glasshouse before the plants were transferred to the field. Using an enzyme-linked immunosorbent protein assay, expression of the *BAR* gene at the protein level was measured on 168 transgenic events produced from the poplar clone 717-1B4. Based on the enzyme-linked immunosorbent assay (ELISA) performed on *BAR* transgenic poplar plants produced for another study, quantified protein levels from the *BAR* transgene were well correlated with the observed herbicide damage (Li *et al*., 2008).

A wide range of transgene expression levels were observed for both *GFP* and *BAR* transgenes in both transgenic populations with and without flanking MARs (Figure 2). No significant difference in mean *GFP* expression was observed between clones 353-38 and 717-1B4 ($P > 0.05$). In addition, there was no significant interaction effect between MAR and clone ($P > 0.05$). Therefore, the two clones were pooled together when testing MAR effects.

The mean *GFP* expression levels for the constructs with and without MARs were 295 [$n = 207$; standard deviation (SD) = 67] and 303 ($n = 197$; SD = 81), respectively, differing only by 3%. This difference was not statistically significant ($P > 0.05$).

Figure 1 Schematic maps of the T-DNA region of binary vectors. Both constructs contained backbone and border sequences from the pGreenII binary vector, the selectable marker gene neomycin phosphotransferase type II (*NPTII*) driven by the nopaline synthase (*NOS*) promoter, the green fluorescent protein (*GFP*) gene driven by the *35S* promoter, and the Basta resistance (*BAR*) gene driven by the poplar *rbc*S promoter. LB, left border; RB, right border; TP, transit peptide from *Arabidopsis* small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) for targeting *BAR*-encoded proteins to chloroplasts. (a) Construct pG3KGB without flanking matrix attachment regions (MARs). (b) Construct pG3MKGB with flanking MARs. The orientation of MAR elements is indicated by arrows. The T-DNA regions are not drawn to scale.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd,

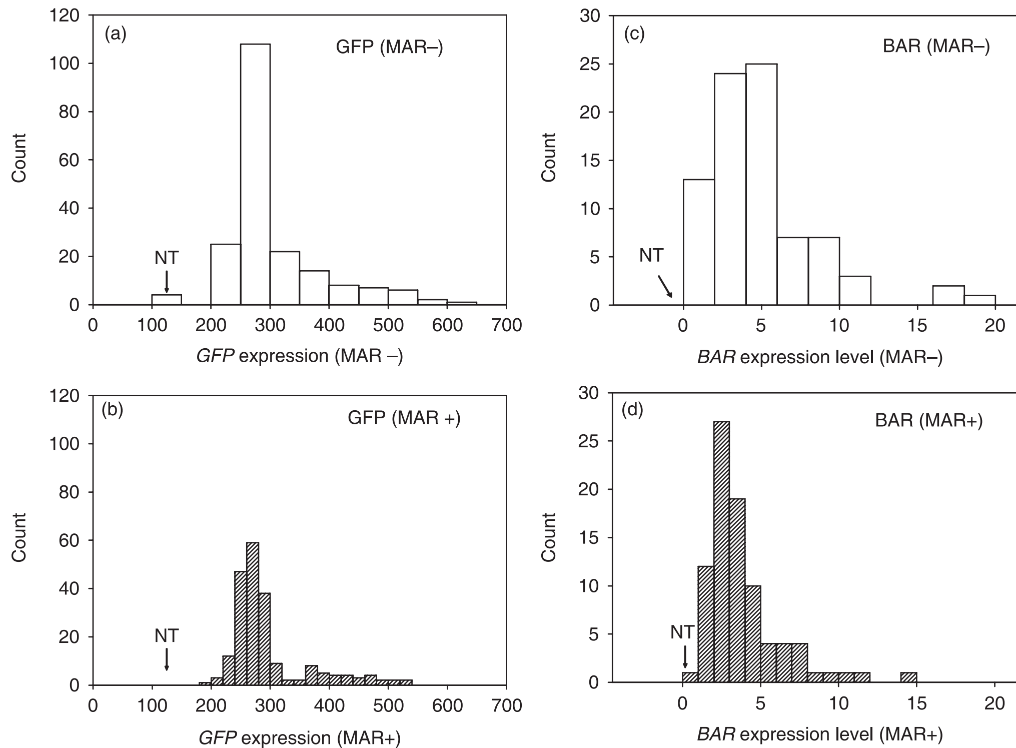

Figure 2 Distribution of green fluorescent protein (*GFP*) and Basta resistance (*BAR*) gene expression levels of transgenic events with and without matrix attachment regions (MARs). (a, b) *GFP* expression distribution of transgenic events without (a) and with (b) MARs. (c, d) *GFP* expression distribution of transgenic events without (c) and with (d) MARs. For each event, the expression level was averaged over 3 years. Mean background levels of non-transgenic plants (NT) are indicated.

In contrast, the mean *BAR* expression level without MARs (mean = 4.9, SD = 3.6, $n = 82$) was significantly higher ($P < 0.05$) than that of the MAR-containing transgenic events (mean = 3.8, SD = 2.4, $n = 86$), a difference of 23%.

MARs did not have a significant effect on the distributions of *GFP* expression among events in any of the 3 years studied (Kolmogorov–Smirnov test, $P > 0.05$; Figure S1, see 'Supporting information'). However, of the 404 transgenic events, four events without MARs (7gb41, 7gb319, 7gb227, 3gb102) were initially silenced (i.e. *GFP* expression was indistinguishable from that for non-transgenic controls), and remained so in the field. Correspondingly, the distribution appeared to have a greater dispersion for events without MARs. This was reflected in a significantly greater variance and SD in the non-MAR group. However, after the four transgenic events with silenced *GFP* expression were removed from the variance analysis, the two groups did not differ significantly in expression variance ($P = 0.11$). The effect of MARs on the expression distributions appeared to be greater for *BAR* than for *GFP* (statistically significant based on the Kolmogorov–Smirnov test, $P < 0.05$; Figure 2; Figure S2, see 'Supporting information'). Similar to their effect on *GFP* variance among events, MARs significantly reduced the variance of *BAR* expression ($P < 0.05$), resulting in less dispersion in the distribution for MAR-containing events. Three of the 717-1B4 transgenic events silenced for *GFP* expression did not show initial silencing of the *BAR* gene; however, initial silencing of the *BAR* gene occurred for event 3gb102 (data not shown). There were no new cases of gene silencing observed during the years of field study, and no significant MAR effects for quantitative expression variation over the years when random and linear effects were analysed (Li, 2006).

Effect of MARs on correlated expression of linked *GFP* and *BAR* transgenes

Because the *GFP* and *BAR* transgenes were contained in the same T-DNA, it was reasonable to expect that the two transgenes would show correlated expression levels among transgenic events. The correlation coefficients between the *GFP* and *BAR* expression levels in both MAR- and non-MAR-containing transgenics were therefore determined for each of the 3 years. There was a weak positive association ($r = 0.09$–0.16) between the expression levels of *GFP* and *BAR* in transgenic plants without MARs, but the association was not statistically significant ($P > 0.05$). The weak correlation was strongly influenced by three events that had been silenced for

© 2008 The Authors

GFP expression prior to the first measurement. The exclusion of the three events as outliers resulted in a stronger correlation (r = 0.18–0.25; P = 0.02–0.09). In the MAR-containing transgenic events, the correlation between *GFP* and *BAR* expression levels was much stronger in the years 2004 and 2005 (r = 0.44 and 0.38, respectively), and highly statistically significant ($P < 0.0001$). However, the correlation for the year 2003 was weaker (r = 0.12; P = 0.28) and close to the correlation in the non-MAR transgenic plants.

On the basis of the observation that the MAR sequence acts as a buffering sequence to prevent T-DNA truncations in multiple-copy transgenic populations (discussed below), the expression correlation between the two transgenes was also analysed for single-copy transgenic populations alone; the results were similar to those obtained when the whole transgenic population was analysed. In the single-copy transgenic plants without MARs, the correlation was very weak in each of the 3 years of study (r = –0.02–0.09). In the MAR-containing populations, the correlation for the year 2003 was also weak (r = –0.01, P = 0.97), but the correlation was much stronger in the years 2004 (r = 0.31, P = 0.007) and 2005 (r = 0.20, P = 0.08).

Effect of MARs on transgene integration and structure

The integrated transgene copy number was determined for 396 transgenic events using comparative real-time polymerase chain reaction (PCR). The presence of MARs was not statistically associated with the distribution of the copy number (Figure 3). For events transformed with the construct without MARs, 82.7% contained a single copy of the insert, similar to the frequency seen for events containing MARs (83.6%).

Employing a similar strategy to that used by Kumar and Fladung (2000), the T-DNA structure was determined for 55 events with multiple copies of integrated transgenes. The method uses inverse PCR, where primers point away from each other in the T-DNAs (Figure 4). Thus, no amplification should be achieved in the case of single-copy integration or when multiple copies are integrated at different loci. The presence of amplified bands will vary, however, depending on whether the multiple copies of T-DNA are arranged as direct or inverted repeats at a given locus. The presence and size of PCR products also indicate whether the repeats are complete or incomplete (Table 1; Figure S3, see 'Supporting information').

Analysis of two-copy events

Of the 55 multiple-copy events (two to four copies), 40 contained two copies and 15 contained three or four copies (Figure 5). For the 40 two-copy events, 21 events carried MARs and the remaining 19 events did not. Of the 19 non-MAR events, eight (42%) showed integration of the two copies at the same locus. Of these eight events, six formed direct repeats and two formed inverted repeats, arranged head-to-head (Figure 5). Seven (33%) of the 21 MAR-containing events had repeat formation at the same locus, and all were arranged as direct repeats. For the six non-MAR events with direct repeats, five showed complete direct repeats without the obvious presence of intervening DNA sequence (the size and patterns of the amplified fragments followed predictions), whereas the other event had a complete direct repeat with approximately 400 bp of intervening DNA between the repeats. A large truncation was also observed in two events containing head-to-head inverted repeats. About 1.9 kb of T-DNA sequence was truncated at the left border of one copy, and 2.7 kb at the right border of the other copy. In addition, when two copies were arranged as inverted repeats, T-DNA truncations occurred in both copies and on different borders (i.e. if the left border of one copy was truncated, a truncation of the other copy occurred on the right border). Frequent truncations were observed for MAR-containing events with direct repeats. Of the seven events with direct repeats, three had complete direct repeats and the remaining four had a large truncation (~1.4 kb) in the MAR sequences at the junctions of two repeats (Figure 5; Figure S3).

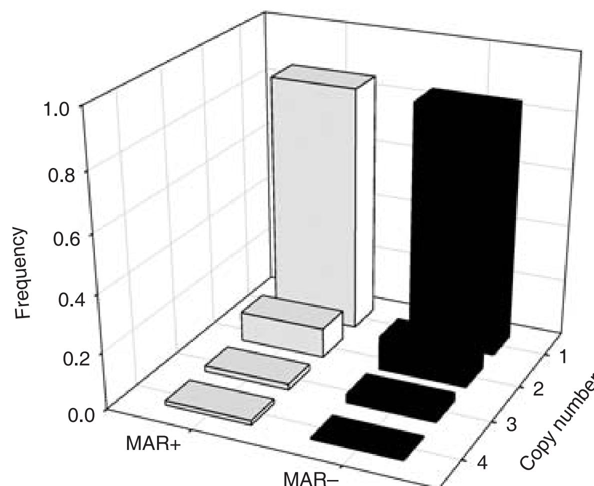

Figure 3 Distribution of transgene copy number for transgenic events with and without matrix attachment regions (MARs). Copy number distributions for 201 transgenic events containing MARs (+) and for 195 transgenic events lacking MARs (–).

Analysis of events with three and four copies

When more than two copies are integrated, there are certain cases in which the PCR strategy is not able to determine the

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd,

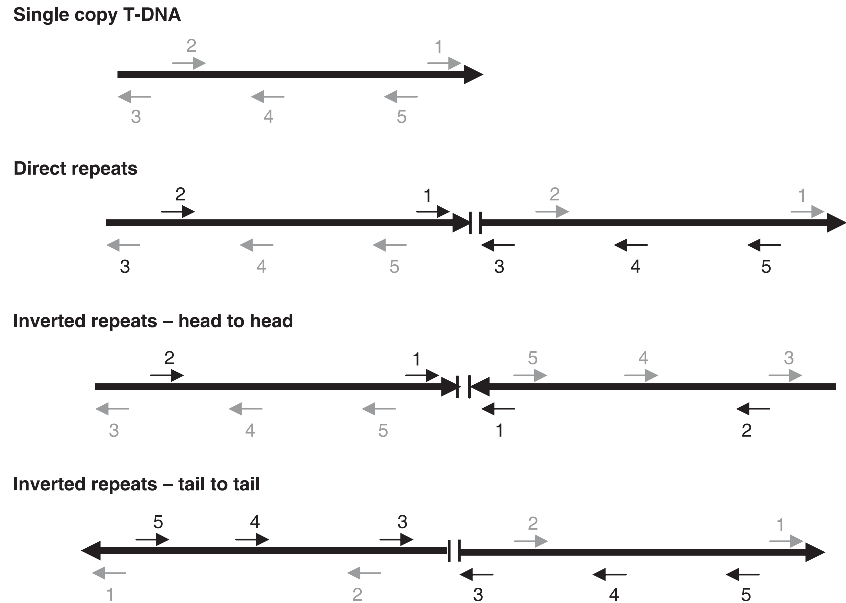

Figure 4 Diagrams of T-DNA repeats and polymerase chain reaction (PCR) primers to determine the T-DNA structure. The relative positions of five PCR primers (1–5) along the T-DNA are indicated. Larger arrows represent T-DNA and point towards the right-hand border. Primer positions that would not be capable of contributing to the amplification of a product in the different insertion configurations are shown in grey. The number and size of the expected bands that would amplify for each insertion configuration with different primer pairs are listed in Table 1.

Table 1 Primer pairs used for polymerase chain reaction (PCR) amplification to determine the T-DNA repeat structure

Primer pair		NR	DR	IR-TT	IR-HH
P1	1 + 3	–	a (2.1)	b (2.5)	c (1.6)
P2	1 + 4	–	a + 1.6 (3.7)	b + 3.2 (5.7)	c (1.6)
P3	1 + 5	–	a + 3.8 (5.8)	b + 7.4 (9.9)	c (1.6)
P4	2 + 3	–	a + 3.1 (5.2)	b (2.5)	c + 6.2 (7.8)
P5	1 + 2	–	–	–	c, c + 6.2, c + 3.2
P6	3 + 4	–	–	b, b + 3.2, b + 1.6	–

DR, direct repeats; IR-HH, inverted repeats arranged head-to-head; IR-TT, inverted repeats arranged tail-to-tail; NR, no repeats.
Letters a, b and c represent the band size (kb), and the numbers in parentheses indicate the expected band size when there is no genomic DNA separating two repeats of T-DNA.
The estimated sizes of the amplified fragments shown in the table are based on the pG3KGB construct. For pG3MKGB, the size is 2.4 kb larger for all corresponding bands. The amplification patterns listed here were based on the presence of complete repeats. Other possible combinations resulting from T-DNA truncations are not listed here, but are explained in the text.

T-DNA structures for all integrated copies. For instance, if three copies of the insert are arranged as direct repeats with the same size of intervening DNA between two repeats, the amplified patterns are the same as two direct repeats at the same locus. However, the amplified bands and size still enable the determination of whether at least two copies are integrated at the same locus. To our surprise, PCR-amplified products were detected for all transgenic events with more than two copies (Figure 5; Figure S3), indicating that, for all of these events, at least two copies of the insert formed repeats at the same chromosomal locus.

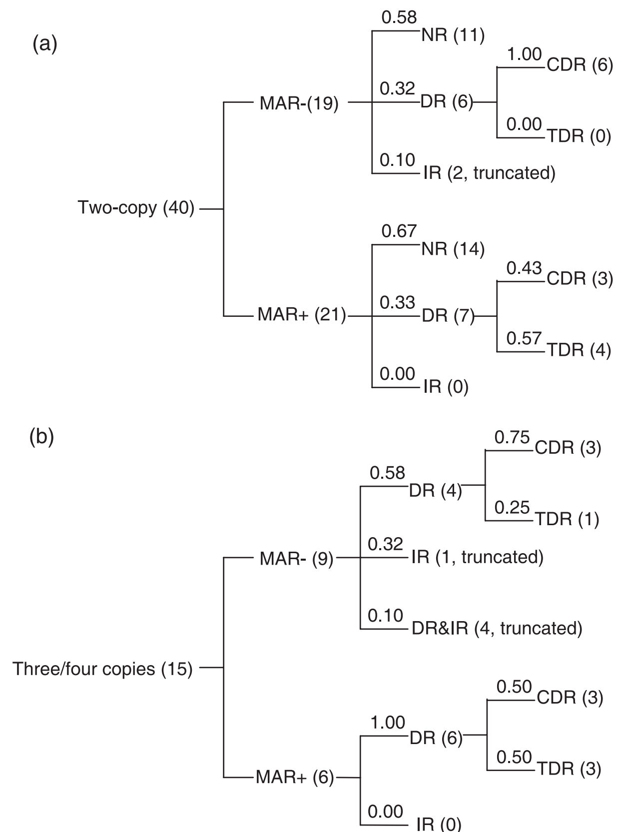

Figure 5 (T-)DNA structures of transgenic events containing two copies (a) or more than two copies (b) of the transgenes with or without matrix attachment regions (MARs). The values in parentheses indicate the number of events in each category. The corresponding frequency for each category is shown above the line. CDR, complete direct repeat; DR, direct repeat; IR, inverted repeat; MAR–, without MARs; MAR+, with MARs; TDR, truncated direct repeat.

© 2008 The Authors

Of the nine non-MAR-containing events, four (44.5%) contained direct repeats (Figure 5), one formed inverted repeats arranged head-to-head and the other four (44.5%) contained both direct and inverted repeats. All six events with flanking MARs formed direct repeats at the same locus. T-DNA truncations were also seen in about 50% of the events with repeat formation, especially when MARs or inverted repeats were present (Figure 5). In one event which was silenced for both *GFP* and *BAR* expression, one of the amplified products from primer pair P4 could not be explained with any kind of repeat formation, and it seems probable that T-DNA rearrangements in the inverted repeat copies, or non-specific binding, may have occurred (Figure S3).

In summary, the presence of flanking MARs did not affect significantly the frequency of multiple insertions at a given locus. For two-copy transgenic events, the frequencies of multiple copies at the same locus with and without MARs were 0.42 and 0.33, respectively, not statistically significantly different (χ^2, $P > 0.05$). For more than two copies, the frequency was 1.00 for both groups. The overall frequency for the integration of multiple copies at the same locus was 0.55 (30 of 55), and direct repeats were the major type of repeat formation (77%). However, there was a strong and statistically significant association between the presence of MARs and the occurrence of DNA truncations when multiple copies were arranged as direct repeats; T-DNA truncation occurred at a frequency of 54% when MARs were present, and 10% without MARs (χ^2, $P < 0.05$). In addition, MARs were associated with reduced formation of inverted repeats at a single locus. When MARs were present, none of the 13 events with repeat formation at a single locus contained inverted repeats. When there were no flanking MARs, seven of the 17 multiple-copy events with T-DNA repeats at one locus contained either inverted repeats or both direct and inverted repeats (χ^2, $P < 0.01$). It was also found that the formation of inverted repeats at a single locus was statistically significantly associated with T-DNA truncations. All seven events containing inverted repeats, arranged in either head-to-head or tail-to-tail configurations, had T-DNA truncations. In contrast, T-DNA truncations occurred in one of the 10 non-MAR events arranged as direct repeats (χ^2, $P < 0.001$).

Discussion

Most previous studies of the effect of MARs on transgene expression have indicated that MARs increase transgene expression to varying degrees, but provide little or no reduction in the variance of expression (reviewed in Allen *et al*., 2000). In contrast, in this study, it was found that MARs from the tobacco root-expressed gene *RB7* did not increase the expression of the two transgenes studied. The mean expression of *GFP* was slightly lower in MAR-containing transgenic events, whereas *BAR* expression was significantly and substantially lower in MAR-containing events. This result contrasts with a previous study in poplar, where highly elevated transgene expression levels (10-fold) caused by *RB7* MARs were reported in a limited sample of transgenic events (Han *et al*., 1997). MAR effects can be strongly influenced by the developmental state and tissue type (Ulker *et al*., 1999; Allen *et al*., 2000). In the study by Han *et al*. (1997), MAR effects were examined in stem tissues from *in vitro*-grown plants, whereas, in the present study, leaf tissues of glasshouse- and field-grown plants were used. In addition, Han *et al*. (1997) reported a much smaller effect of the same MAR on *GUS* expression in transgenic tobacco, where leaf tissues were also assayed. In mouse, a large effect of a MAR on transgene expression was observed in embryonic cells, but a much smaller effect in differentiated adult tissues (Thompson *et al*., 1994). Promoter-dependent effects of MARs have also been reported (Mankin *et al*., 2003; Sidorenko *et al*., 2003).

MARs appeared to reduce significantly the variance in expression of both transgenes, an effect that was largely a result of a decrease in the number of weakly expressing events. The transgenic events with no detectable *GFP* expression all lacked MARs. When these four events were removed, the effect of MARs on variance was not statistically significant. Mankin *et al*. (2003) also reported that the number of low-expressing *GUS* transformants was greatly reduced when MAR-flanked constructs were used in the biolistic transformation of tobacco. Similar results have been obtained in other studies of the biolistic transformation of tobacco and rice (Ulker *et al*., 1999; Vain *et al*., 1999). Likewise, the beneficial effect of the presence of MARs on the correlation of transgene expression was modest and highly variable. In the single-copy transformants, it was very weak and not statistically significant in the first year during glasshouse growth, but it was stronger and statistically significant (or nearly so) in the 2 years of field growth. However, even then, the percentage of variation in expression of one transgene associated with expression of the other was only 4%–10%.

Despite the influence of the four silenced events on the statistical significance of variation in transgene expression, it is reasonable to argue that the decrease in gene silencing in the MAR-containing population may not have strong biotechnological significance. This is because a large number of transgenic events were studied, similar to that in breeding programmes, and these provide a large range of variation in

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd,

expression for selection. However, the number of silenced events in the non-MAR transgenic populations could have been much higher had no selectable marker been used during transformation; our data showed that the MAR elements played a significant role in preventing the truncation of transgenes during T-DNA insertion (discussed below). MARs may also play a larger role when T-DNAs occur at a high frequency in heterochromatic regions; an increase in T-DNA insertion into heterochromatin has been reported in *Arabidopsis* when selectable markers are not employed during the regeneration of transgenic plants (Kim and Gelvin, 2007). However, the use of selectable markers is common practice in transgenic biotechnology programmes. It was also found that MARs appeared to prevent the integration of inverted repeats at the same locus, which may also reduce the number of weakly expressing events as a consequence of induced RNA interference (Kusaba, 2004). The frequency of inverted repeat integration was 0.41 when MARs were not present, whereas no inverted repeat events were observed when MARs were present.

Frequent truncation of the MAR sequence was observed in transgenic events that contained multiple copies. The truncation effect was unexpected; our initial intent was only to study the rearrangements of T-DNAs in multiple-copy transgenic events. As a result, the rate of truncation in the single-copy transgenic events was not determined. However, there is no reason to believe that the truncation effects observed in multiple-copy transgenics would not also apply in the single-copy transgenic population. The truncation results suggest that the MAR elements may have protected the transgenes at the ends of the T-DNA from partial deletions, effectively reducing gene silencing and increasing the correlation between the expression of genes in the distal and medial portions of the T-DNA. This explanation is supported by our observation that the events responsible for the poor correlation between *BAR* and *GFP* expression in the absence of MARs all showed abnormally low *BAR* expression, which was the gene nearest the right T-DNA border. A similar beneficial effect of MARs has been reported by Mlynarova *et al*. (2003) in tobacco transformed with *Agrobacterium tumefaciens*; the correlation in expression of *GUS* and luciferase (*LUC*) transgenes in the same T-DNA was improved by the presence of a flanking chicken lysozyme A MAR element. A protective effect of MAR elements is also consistent with previous observations that MARs can increase the transformation efficiency (Han *et al*., 1997; Shimizu *et al*., 2001; Petersen *et al*., 2002), as selectable markers are often near T-DNA borders.

Our results showed a very limited or negative effect of MAR elements on several aspects of transgene structure and stability, including expression level, year-to-year variation, position effect, T-DNA configuration and copy number, and correlation in expression among co-integrated transgenes. The lack of a strong beneficial effect of MARs on expression and stability agrees with many other recent observations (for example, Miguel *et al*., 2007), suggesting that MAR elements have a much more limited and variable effect than believed initially. Research on other MARs or other kinds of DNA elements, or on novel transgenic strategies (Miki, 2002; Butaye *et al*., 2005), are badly needed to moderate the still notoriously high variation in the levels and patterns of expression among gene insertion events.

Experimental procedures

Vector construction

A modified *GFP* gene (*mgfp5er*, provided by C.N. Stewart, University of Tennessee, Knoxville, TN, USA) was used, which was altered to increase the sensitivity to blue light, to maximize expression and processing in plants, and to target the encoded protein to the endoplasmic reticulum (Haseloff *et al*., 1997). The *BAR* gene from *Streptomyces hygroscopicus* encodes phosphinothricin acetyltransferase and confers resistance to the herbicide glufosinate, which inhibits glutamine synthetase (Riemenschneider, 1997).

Because gene silencing may be very different with a native vs. a foreign promoter, promoters from the cauliflower mosaic virus *35S* gene and from the poplar gene for the small subunit of ribulose bisphosphate carboxylase (*rbc*S) were used to drive *GFP* and *BAR*, respectively. The binary vector pGreenII (Hellens *et al*., 2000) was used to assemble the neomycin phosphotransferase type II (*NPTII*) selectable marker and *GFP* and *BAR* reporter genes (Figure 1). An *Asc*I linker was inserted at the *Hpa*I and *Stu*I sites of the pGreenII backbone to produce pG3. Two versions of the vector were used: one with flanking MAR elements derived from the tobacco *RB7* gene (provided by S. Spiker, North Carolina State University, Raleigh, NC, USA), and one without MARs. For cloning of MARs, a 1167-bp MAR fragment was removed from the vector pHK10 with the restriction enzymes *Not*I and *Spe*I, blunted and cloned at the *Fsp*I and *Sap*I (blunted) sites of pG3 to produce pG3M.

The selectable marker cassette consisted of the promoter from the nopaline synthase gene (*NOS*), the coding region of the *NPTII* gene, which confers resistance to the antibiotic kanamycin, and the 3′ untranslated region from the *NOS* gene. To assemble this cassette, an 869-bp coding region of the *NPTII* gene was cut from pJIT134 with *Pst*I and *Xba*I, and inserted into the corresponding *Pst*I and *Xba*I sites of the *NOS* promoter cassette from the pGreenII system. The whole cassette (1396 bp) was then excised with *Eco*RV, and cloned into the blunted *Xho*I site of pG3 and pG3M to produce pG3K and pG3MK, respectively.

The *GFP* cassette comprised the *35S* promoter and terminator from cauliflower mosaic virus, and the *mGFP5er* coding sequence. The 818-bp m*GFP*5er coding sequence was cut from pBIN m*GFP*-5-ERer with *Bam*HI and *Sac*I, and ligated into the corresponding sites of the *35S* promoter cassette from the pGreenII system. The whole cassette was then restricted with *Eco*RV, and cloned into the *Eco*RV site of pG3K to produce pG3KG.

© 2008 The Authors

The *BAR* cassette consisted of the poplar *rbcS* promoter, a transit peptide from the *Arabidopsis rbc*S gene for targeting *BAR*-encoded phosphinothricin acetyl transferase (PAT) proteins to chloroplasts, the *BAR* coding sequence and the terminator sequence from the *Agrobacterium g7* gene. The poplar *rbc*S promoter was cloned from *Populus trichocarpa* genomic DNA using the Universal Genomic Walker kit (Clontech, Mountain View, CA, USA; Cat. # K1807-1) following the manufacturer's instructions. Gene-specific primers were designed from the poplar *rbc*S cDNA sequence (provided by J. Davis, University of Florida, Gainesville, FL, USA). The gene-specific primers for the primary and secondary PCR amplifications were 5′-GCATGCATTGAACTCGTCCACCATTGC-3′ and 5′-ATGTCATT-AGCCTTTCTGGTACTGGCT-3′, respectively. A 1300-bp fragment was amplified from the secondary PCR from the *Pvu*II library at an annealing temperature of 72 °C. The amplified promoter fragment was then cloned into the pGEM Easy vector (Promega, Madison, WI, USA; Cat. # A1360), and sequenced from both ends. An 1100-bp promoter fragment was removed using *Acc*I and *Hin*dIII, and cloned into the corresponding sites of the intermediate vector pBluescript KS (Stratagene, La Jolla, CA, USA). A 971-bp fragment encoding the *Arabidopsis rbc*S transit peptide, the *BAR* coding sequence and the *g7* terminator was amplified from the vector pTTM8 (Li *et al*., 2008) provided by Plant Genetic Systems (Gent, Belgium) with the following primers: 5′-GTCTGCAGGAACAATGGCTTCCTC-TATG-3′ and 5′-AGACTAGTGATGTTAATTCCCATCTTG-3′. The amplified fragment was digested with *Spe*I and *Pst*I, and cloned into the corresponding sites of pBluescript KS containing the poplar *rbc*S promoter. The entire *BAR* cassette was then cut from the intermediate vector with *Xho*I (blunted) and *Spe*I, and cloned into the *Sma*I and *Spe*I sites of pG3KG to produce pG3KGB (Figure 1). The fragment containing the *BAR* and *GFP* cassettes was removed from pG3KGB with *Spe*I and *Sal*I (partial digestion as a result of an internal *Sal*I site in the *BAR* cassette), and cloned into the corresponding sites of pG3MK to produce pG3MKGB (Figure 1).

Plant transformation

Two hybrid poplar clones, 353-38 (*P. tremula* × *P. tremuloides*) and 717-1B4 (*P. tremula* × *P. alba*), were transformed using *Agrobacterium tumefaciens* strain C58/pMP90 (GV3101) harbouring the transgenic constructs pG3KGB and pG3MKGB, following an established protocol based on the *NPTII* gene, as described previously (Filichkin *et al*., 2006). A total of 404 independent transgenic events was produced. Of these, 99 events were for pG3KGB in 353-38, 98 events for pG3KGB in 717-1B4, 108 events for pG3MKGB in 353-38 and 99 events for pG3MKGB in 717-1B4.

Six to eight ramets for all transgenic events, plus 32 non-transgenic plants for each of the two clones, were propagated *in vitro*, potted in soil and grown in the glasshouse for 3–4 months. Together with non-transgenic plants, four ramets from the individual transgenic events were planted at an irrigated field near Corvallis, OR, USA, in October 2003. The plants were distributed in four blocks with a random split-plot design, where the two clones were separated and randomly assigned to subplots of each block, and transgenic events from one of the two clones were randomly distributed within the corresponding subplots. The plants were coppiced in the spring of 2004 to stimulate shoot growth, a common practice for poplar.

Measurement of transgene expression levels

The expression levels of *GFP* and *BAR* in three different years were measured: 2003 in the glasshouse, and 2004 and 2005 in the field. Measurements in the glasshouse were taken before the plants were planted in the field in October 2003. *GFP* expression was quantified on intact leaf samples and surfaces with a hand-held GFP-Meter (Opti-Sciences, Hudson, NH, USA; http://www.optisci.com). The GFP-Meter is a self-contained, hand-held fluorometer that uses a modulated system to compensate for stray light and temperature drift (for a detailed description, see Millwood *et al*., 2003). *GFP* quantification was made on intact leaf samples and surfaces. When powered on, the meter generates excitation light that travels through a bandpass filter to a fibre-optic cable, and is then delivered to the leaf sample through an attached leaf clip. The light emitted from the leaf sample is directed through a bandpass filter into a low-noise pre-amplifier. The fluorescence signal is processed and displayed in units of counts per second. Based on preliminary expression studies (Li, 2006), the fourth and fifth leaves (~plastochrons) from the top of each glasshouse plant, which were fully expanded, were sampled for the measurement of *GFP* expression. For the field plants, the leaf immediately above the first fully expanded leaf from two different shoots was used for each measurement.

Expression of the *BAR* transgene was quantified by ELISA, as described previously (Li *et al*., 2008). Total protein concentration was measured using the Bradford method (Bio-Rad Protein Assay Kit, Hercules, CA, USA; Cat. # 500-0001). Triplicate and duplicate reactions were used for bovine serum albumin (BSA) standards and experimental samples, respectively. The relative concentration of PAT encoded by *BAR* was quantified using the commercial LibertyLink PAT/*BAR* ELISA kit (Envirologix Inc., Portland, ME, USA; Cat. # AP013). Non-transgenic controls gave the same background levels as blanks (protein extraction buffer without any samples added); thus, only blanks were included in each of the assay plates. A reference sample was run in all assay plates for data normalization, and duplicated wells were used for all tested samples. The optical density (OD) was determined at a wavelength of 450 nm, 20 min after adding stop solution to the tested wells. The mean OD from the blank wells was subtracted from all samples. The relative OD readings (i.e. relative to the reference sample) were used for further data analysis.

Comparative real-time PCR for transgene copy number

To estimate the transgene copy number in transgenic plants, comparative real-time PCR was performed on 396 transgenic events. The endogenous, single-copy gene *PTLF* (poplar *LEAFY/FLORICAULA*) was used as the reference gene (Rottmann *et al*., 2000). The transgene *GFP* was chosen for copy number analysis as it was closer to the last-transferred left border of the T-DNA than was the *BAR* gene. Real-time PCR was performed using dual-labelled TaqMan® probes (Biosearch Technologies, Novato, CA, USA). All primers and probes were designed using the program Primer 3 (Rozen and Skaletsky, 2000; http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). For the *GFP* transgene, the primers used for amplification were 5′-TTAAGGGAATCGATTTCAAG-3′ and 5′-ACGTTGTGGGAGTTG-TAGTT-3′. The hybridization probe was dual labelled with 6-hexachlorofluorescein (HEX) and Black Hole Quencher 1 (BHQ-1): 5′ HEX d(CCTCGGCCACAAGTTGGAATAC) BHQ-1 3′. For the *PTLF* gene, the

© 2008 The Authors

PCR amplification primers were 5′-GGTTTCTCTGAGGAGCCAGTA-CAG-3′ and 5′-GCCTCCCATGTCCCTCTTC-3′. The hybridization probe was labelled with carboxyfluorescein (FAM) and 6-carboxy tetraethyl-rhodamine (TAMARA): 5′ FAM d(CAAGGAGGCAGCAG-GGAGCGGT) TAMARA 3′. The primer and probe sets were tested for their amplification efficiency using two-fold serial standard dilutions of each transgenic sample, and were optimized to provide comparable efficiency (data not shown). For the *PTLF* gene, the optimal concentration for both forward and reverse primers was 0.4 μM, and for the probe was 0.2 μM. For the *GFP* gene, the optimal concentration for both primers was 0.4 μM, and for the probe was 0.3 μM.

All PCRs were carried out using QuantiTect Multiplex PCR buffer (Qiagen, Valencia, CA, USA; Cat. # 204545) in a volume of 25 μL. For each sample, 100 ng of genomic DNA purified with the Dneasy Plant Mini Kit (Qiagen; Cat. # 69106) was used in a duplexed reaction to amplify both *PTLF* and *GFP* targets. Each sample was duplicated in two wells. The amplification was performed in an Mx3000 P real-time PCR machine (Stratagene) with the following cycles: 95 °C for 15 min, and 45 cycles of 94 °C for 30 s and 60 °C for 1 min. The threshold cycle (Ct) was determined using Mx3000 P System software (version 2). The transgene copy number of the *GFP* transgene was determined using the formula: $2^{[1 - \Delta Ct\,(GFP - PTLF)]}$.

Inverse PCR for T-DNA structure

To investigate the T-DNA structure for inserts with multiple copies, five different primers located along the T-DNA sequence were used for PCR amplification (Figure 4). The presence and size of the amplified bands were used to infer the T-DNA structure (Table 1). The five primers used were as follows: 1, 5′-TTTCTGGCAGCTGGACTTCAG-3′; 2, 5′-TAGAAAAGGAAGGTGGCTCCTACA-3′; 3, 5′-CCAAGCTCT-TCAGCAATATCAC-3′; 4, 5′-AGAAGGACCATGTGGTCTCT-3′; 5, 5′-AACCCACGTCATGCCAGTTCC-3′. PCR was performed using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA, USA; Cat. # 00304=029) in a volume of 25 μL with the following concentrations of reaction reagents: primer, 0.4 μM; deoxynucleoside triphosphates (dNTPs), 0.2 mM; $MgSO_4$, 2 mM; DNA, 20 ng; Taq, 2 U; and 1 × high-fidelity buffer. The PCR programme was as follows: 95 °C for 2 min; 30 cycles of 94 °C for 20 s, 59 °C for 30 s and 72 °C for 30 s (per kb of amplicon); and 72 °C for 4 min.

Statistical analysis

GFP fluorescence levels measured in the field in 2004 and 2005 were normalized using the ratio of the mean background levels of non-transgenic plants from each of the 2 years. As a result of the different optical cables used for the glasshouse and field measurements, the measured *GFP* fluorescence levels in the glasshouse could not be normalized in the same way; the sample values and background levels responded differently to different optical cables (data not shown). Thus, the study of the expression levels involving glasshouse measurements used non-normalized data. Quantitative values were $\log_2$-transformed for statistical tests.

To account for the variation between different ELISA plates, the measured OD of the samples run on a given plate was divided by the OD of the reference sample on the same plate. The normalized OD was then square root-transformed before performing the statistical analyses. Tests of variance in the mean expression levels among clones, constructs and their interactions were performed with the general linear model (GLM) procedure using the statistical program SAS version 8. Levene's test (Levene, 1960) on equality of variances was used to examine the effect of MARs on the variance of expression levels among events. Associations between the presence of MARs and the frequency of T-DNA truncation or formation of inverted repeats were examined using a χ^2 test. Differences in transgene expression distributions were studied using the Kolmogorov–Smirnov test.

Acknowledgements

We thank Dr C. Neal Stewart for providing the *mgfp5er* gene, Dr Steven Spiker for providing the MAR element and Dr John Davis for the use of poplar *rbc*S cDNA. We also thank Dr Christopher Bayne for allowing us to use the microtitre plate reader in his laboratory. The research project was supported by the Initiative for Future Agriculture and Food Systems, Grant no. 00-52100-9623, from the USDA Cooperative State Research, Education, and Extension Service, industrial members of the TBGRC/TGERC Research Cooperative and OSU Foundation Forest Biotechnology Fund.

References

Allen, G.C., Hall, G.E. Jr, Childs, L.C., Weissinger, A.K., Spiker, S. and Thompson, W.F. (1993) Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. *Plant Cell*, **5**, 603–613.

Allen, G.C., Hall, G. Jr, Michalowski, S., Newman, W., Spiker, S., Weissinger, A.K. and Thompson, W.F. (1996) High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco. *Plant Cell*, **8**, 899–913.

Allen, G.C., Spiker, S. and Thompson, W.F. (2000) Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Mol. Biol*. **43**, 361–376.

Brouwer, C., Bruce, W., Maddock, S., Avramova, Z. and Bowen, B. (2002) Suppression of transgene silencing by matrix attachment regions in maize: a dual role for the maize 5′ ADH1 matrix attachment region. *Plant Cell*, **14**, 1–14.

Brunner, A.M., Li, J., DiFazio, S.P., Shevchenko, O., Montgomery, B.E., Mohamed, R., Wei, H., Ma, C., Elias, A., Wormer, K.V. and Strauss, S.H. (2007) Genetic containment of forest plantations. *Tree Genet. Genomics*, **3**, 75–100.

Butaye, K.M.J., Cammue, B.P.A., Dularé, S.L. and De Bolle, M.F.C. (2005) Approaches to minimize variation of transgene expression in plants. *Mol. Breed*. **16**, 79–91.

Conner, A.J., Mlynárová, L., Stiekema, W.J. and Nap, J.-P. (1998) Meiotic stability of transgene expression is unaffected by flanking matrix-associated regions. *Mol. Breed*. **4**, 47–58.

Filichkin, S.A., Meilan, R., Busov, V.B., Ma, C., Brunner, A.M. and Strauss, S.H. (2006) Alcohol-inducible gene expression in transgenic *Populus*. *Plant Cell Rep*. **25**, 660–667.

Han, K.-H., Ma, C. and Strauss, S.H. (1997) Matrix attachment regions (MARs) enhance transformation frequency and transgene expression in poplar. *Transgenic Res*. **6**, 415–420.

© 2008 The Authors

Haseloff, J., Siemering, K.R., Prasher, D.C. and Hodge, S. (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proc. Natl. Acad. Sci. USA*, **94**, 2122–2127.

Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S. and Mullineaux, P.M. (2000) pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol. Biol*. **42**, 819–832.

Kim, S.-I. and Gelvin, S. (2007) Genome-wide analysis of Agrobacterium integration sites in the Arabiopsis genome generated under non-selective conditions. *Plant J*. **51**, 779–791.

Kohli, A., Twyman, R.M., Abranches, R., Wegel, E., Stoger, E. and Christou, P. (2003) Transgene integration, organization and interaction in plants. *Plant Mol. Biol*. **52**, 247–258.

Kumar, S. and Fladung, M. (2000) Transgene repeats in aspen: molecular characterisation suggests simultaneous integration of independent T-DNAs into receptive hotspots in the host genome. *Mol. Gen. Genet*. **264**, 20–28.

Kusaba, M. (2004) RNA interference in crop plants. *Curr. Opin. Biotechnol*. **15**, 139–143.

Levene, M. (1960) Robust tests for equality of variance. In: *Contributions to Probability and Statistics* (Olkin, I., ed.), pp. 58–62. Palo Alto, CA: Stanford University Press.

Li, J. (2006) Stability of reporter gene expression and rnai in transgenic poplars over multiple years in the field under vegetative propagation. Dissertation. Corvallis, OR: Oregon State University.

Li, J., Meilan, R., Ma, C., Barish, M. and Strauss, S.H. (2008) Stability of herbicide resistance over eight years of coppice in field-grown, genetically engineered poplars. *West. J. Appl. For*. **23**, 89–93.

Mankin, S.L., Allen, G.C., Phelan, T., Spiker, S. and Thompson, W.F. (2003) Elevation of transgene expression level by flanking matrix attachment regions (MAR) is promoter dependent: a study of the interactions of six promoters with the RB7 3′ MAR. *Transgenic Res*. **12**, 3–12.

Maximova, S., Miller, C., Antúnez de Mayolo, G., Pishak, S., Young, A. and Guiltinan, M.J. (2003) Stable transformation of *Theobroma cacao* L. and influence of matrix attachment regions on GFP expression. *Plant Cell Rep*. **21**, 872–883.

Miguel, B., Butaye, K., Goderis, I., Wouters, P., Jacobs, A., Delauré, S., Depicker, A. and Cammue, B. (2007) The influence of matrix attachment regions on transgene expression in *Arabidopsis thaliana* wild type and gene silencing mutants. *Plant Mol. Biol*. **3**, 533–543.

Miki, B. (2002) Transgene expression and control. *In Vitro Cell. Dev.-Plant*, **38**, 139–145.

Millwood, R.J., Halfhill, M.D., Harkins, D., Russotti, R. and Stewart, C.N. (2003) Instrumentation and methodology for quantifying GFP fluorescence in intact plant organs. *Biotechniques*, **34**, 638–643.

Mlynarova, L., Hricova, A., Loonen, A. and Nap, J.-P. (2003) The presence of a chromatin boundary appears to shield a transgene in tobacco from RNA silencing. *Plant Cell*, **15**, 2203–2217.

Petersen, K., Leah, R., Knudsen, S. and Cameron-Mills, V. (2002) Matrix attachment regions (MARs) enhance transformation frequencies and reduce variance of transgene expression in barley. *Plant Mol. Biol*. **49**, 45–58.

Riemenschneider, D.E. (1997) Genetic engineering of horticultural and forestry crops for herbicide tolerance. In: *Biotechnology of Ornamental Plants* (Geneve, R.L., Preece, J.E. and Merkle, S.A., eds), pp. 367–384. New York: CAB International.

Rottmann, W.H., Meilan, R., Sheppard, L.A., Brunner, A.M., Skinner, J.S., Ma, C., Cheng, S., Jouanin, L., Pilate, G. and Strauss, S.H. (2000) Diverse effects of overexpression of LEAFY and PTLF, a poplar (*Populus*) homolog of LEAFY/FLORICAULA, in transgenic poplar and Arabidopsis. *Plant J*. **22**, 235–245.

Rozen, S. and Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz, S. and Misener, S., eds), pp. 365–386. Totowa, NJ: Humana Press.

Shimizu, K., Takahashi, M., Goshima, N., Kawakami, S., Irifune, K. and Morikawa, H. (2001) Presence of an SAR-like sequence in junction regions between an introduced transgene and genomic DNA of cultured tobacco cells: its effect on transformation frequency. *Plant J*. **26**, 375–384.

Sidorenko, L., Bruce, W., Maddock, S., Tagliani, L., Li, X., Daniels, M. and Peterson, T. (2003) Functional analysis of two matrix attachment region (MAR) elements in transgenic maize plants. *Transgenic Res*. **12**, 137–154.

Thompson, E.M., Christians, E., Stinnakre, M.G. and Renard, J.P. (1994) Scaffold attachment regions stimulate HSP70.1 expression in mouse preimplantation embryos but not in differentiated tissues. *Mol. Cell. Biol.* **14**, 4694–4703.

Ulker, B., Allen, G.C., Thompson, W.F., Spiker, S. and Weissinger, A.K. (1999) A tobacco matrix attachment region reduces the loss of transgene expression in the progeny of transgenic tobacco plants. *Plant J*. **18**, 253–263.

Vain, P., Worland, B., Kohli, A., Snape, J.W., Christou, P., Allen, G.C. and Thompson, W.F. (1999) Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. *Plant J*. **18**, 233–242.

Supporting information

Additional supporting information may be found in the online version of this article:

Figure S1 Normalized green fluorescent protein (*GFP*) gene expression distribution of transgenic events with and without matrix attachment regions (MARs) in three consecutive years.

Figure S2 Normalized Basta resistance (*BAR*) gene expression distribution of transgenic events with and without matrix attachment regions (MARs) in three consecutive years.

Figure S3 Polymerase chain reaction (PCR)-amplified bands of transgenic events with two or more copies of inserted T-DNA.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Supplemental figures

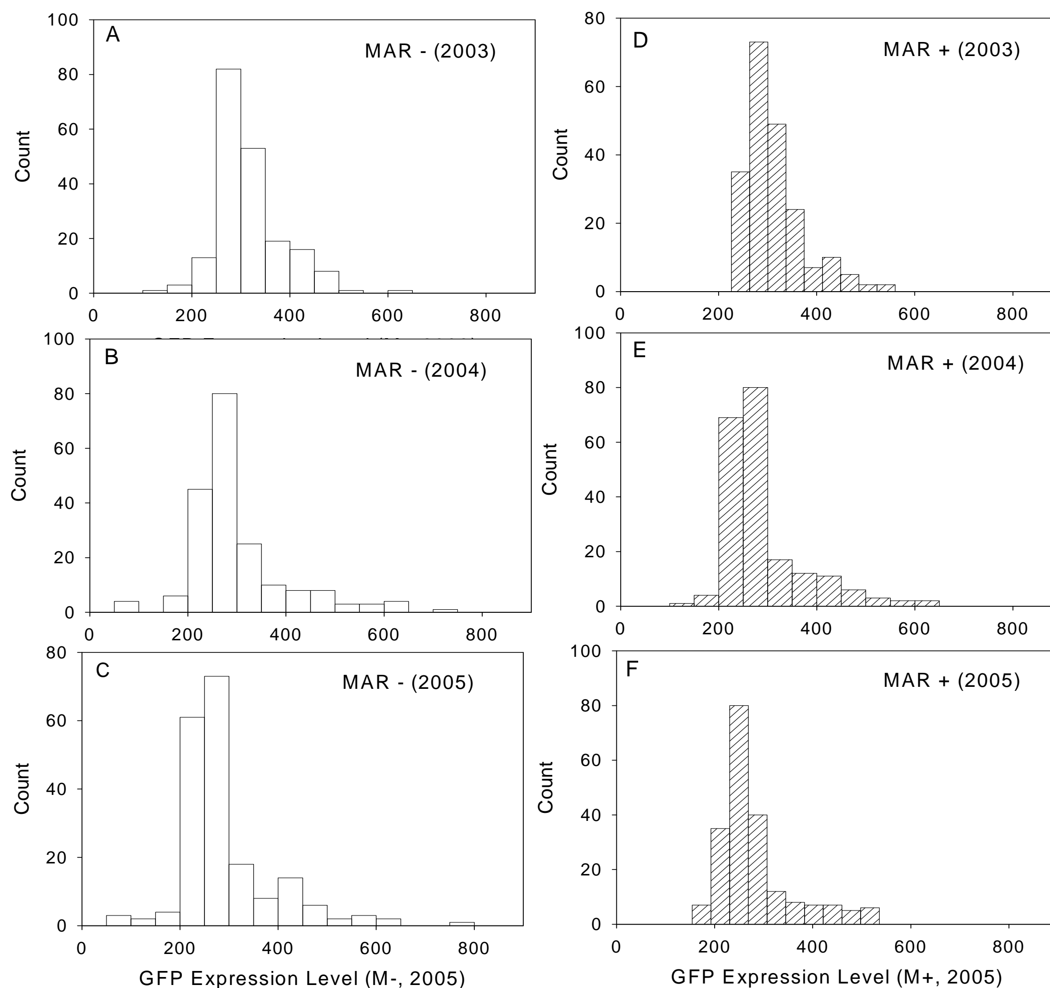

Figure S1 Normalized *GFP* expression distribution of transgenic events with and without MARs in three consecutive years. *GFP* expression was normalized to the median expression values for each of the three years. A−C: *GFP* expression without MARs. D−F: *GFP* expression with MARs.

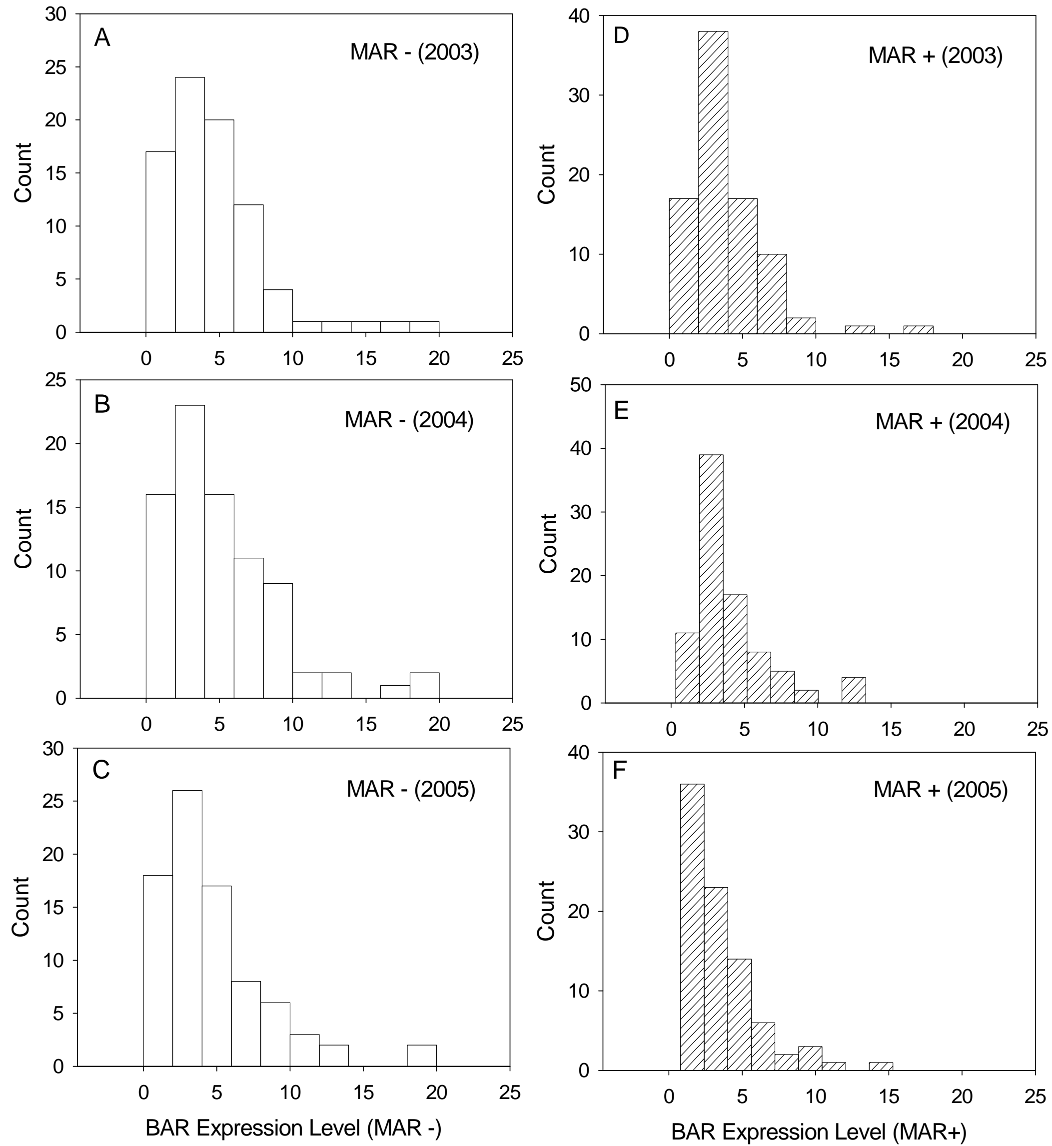

Figure S2 Normalized *BAR* expression distribution of transgenic events with and without MARs in three consecutive years. *BAR* expression was normalized to the median expression values of each of the three years. A−C: *BAR* expression distributions without MARs. D−F: *BAR* expression distributions with MARs.

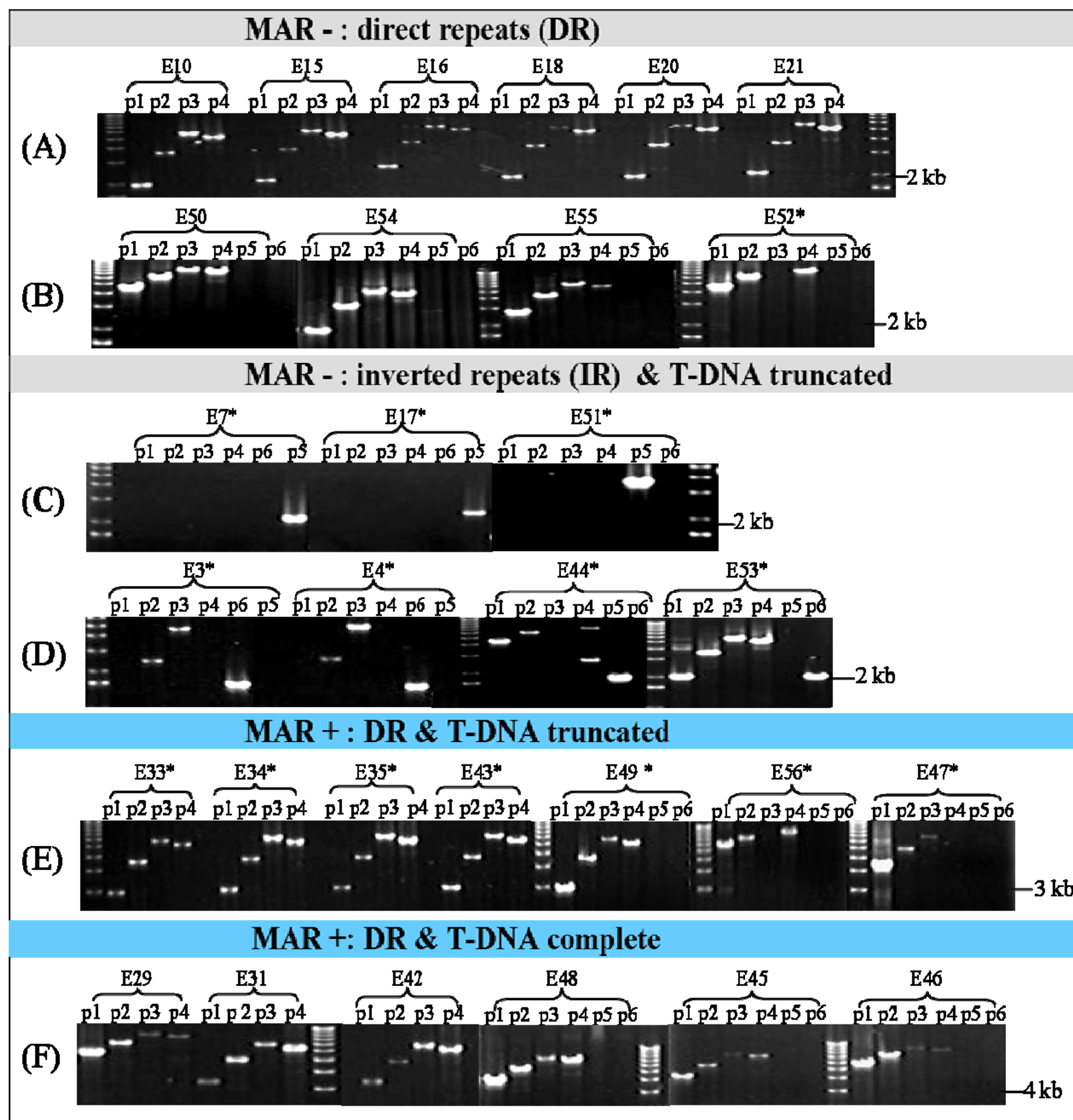

Figure S3 PCR amplified bands of transgenic events with two or more copies of inserted T-DNAs. E10 ~ E21: two-copy transgenic events without MARs. E29 ~ E43: two-copy transgenic events with MARs. E45 ~ E 49: three-copy transgenic events with MARs. E50 ~ E55: three-copy transgenic events without MARs. E56: four-copy transgenic event. p1 ~ p6: various primer pairs. Events with T-DNA truncated are denoted with "*". A: Two-copy non-MAR events with DR; B: three-copy non-MAR events with DR; C: multi-copy non-MAR events with IR; D: multi-copy non-MAR events with both DR and IR; E: multi-copy MAR events with truncated DR; F: multi-copy MAR events with complete DR. 1-kb DNA ladders are shown.