The Plant Cell, Vol. 9, 1251–1264, August 1997 

RESEARCH ARTICLE

# Molecular and Cytogenetic Analyses of Stably and Unstably Expressed Transgene Loci in Tobacco

**Victor A. Iglesias,[a,1] Eduardo A. Moscone,[a,2] István Papp,[a] Franz Neuhuber,[a] Susan Michalowski,[b] Thomas Phelan,[b] Steven Spiker,[b] Marjori Matzke,[a] and Antonius J. M. Matzke[a,3]**

[a] Institute of Molecular Biology, Austrian Academy of Sciences, Billrothstrasse 11, A-5020 Salzburg, Austria
[b] Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695-7614

**To study the influence of genomic context on transgene expression, we have determined the T-DNA structure, flanking DNA sequences, and chromosomal location of four independent transgene loci in tobacco. Two of these loci were stably expressed in the homozygous condition over many generations, whereas the other two loci became unstable after several generations of homozygosity. The stably expressed loci comprised relatively simple T-DNA arrangements that were flanked on at least one side by plant DNA containing AT-rich regions that bind to nuclear matrices in vitro. Of the unstably expressed loci, one consisted of multiple incomplete T-DNA copies, and the second contained a single intact T-DNA; in both cases, however, binary vector sequences were directly contiguous to a right T-DNA border. Fluorescence in situ hybridization demonstrated that the two stably expressed inserts were present in the vicinity of telomeres. The two unstably expressed inserts occupied intercalary and paracentromeric locations, respectively. Results on the stability of transgene expression in $F_1$ progeny obtained by intercrossing the four lines and the sensitivity of the four transgene loci to inactivation in the presence of an unlinked "*trans*-silencing" locus are also presented. The findings are discussed in the context of repetitive DNA sequences and the allotetraploid nature of the tobacco genome.**

## INTRODUCTION

The reliable and stable expression of transgenes is a prerequisite for the successful application of gene technology to agriculture. Many investigators, however, have observed significant variability in the expression of the same transgene construct in different transformed lines. Although it is widely assumed that the flanking plant DNA sequences and chromosomal integration site can influence transgene activity, little is actually known about the specific features of plant DNA and chromosomal location that might be favorable for the stable expression of transgenes.

During the past several years, transgene silencing has received considerable attention. A variety of silencing effects, which involve either single transgene loci or interactions between unlinked loci, has been observed. Single loci that appear to be stably expressed initially can become progressively silenced over several generations, particularly when maintained in the homozygous state (Kilby et al., 1992; Neuhuber et al., 1994). Silencing of single loci can be due primarily to the genomic context ("position effects"; Peach and Velten, 1991) and/or to multiple copies of a transgene construct at a given locus (Linn et al., 1990). The latter effect, which has been termed "homology-dependent" (M.A. Matzke et al., 1994) or "repeat-induced" gene silencing (Assaad et al., 1993), can also include interactions between two unlinked loci in which one transgene locus is capable of *trans*-inactivating a second, homologous transgene locus or a homologous endogenous gene (reviewed in Meyer and Saedler, 1996). In several cases, *trans*-silencing loci have been associated with specific arrangements and/or modifications of transgenes. For example, with certain transgene constructs, inverted repeats provoke post-transcriptional silencing of unlinked homologous genes (Hobbs et al., 1993; Van Blokland et al., 1994; Jorgensen et al., 1996; Stam et al., 1997). Transcriptional inactivation resulting from promoter homology has been associated with multicopy transgene inserts that are methylated in the promoter regions (Vaucheret, 1993; A.J.M. Matzke et al., 1994; Park et al., 1996).

When analyzing the effects of the immediate genomic environment on transgene expression, one must also consider the potential broader impact of the genome organization of the host plant with respect to the chromosomal distribution of structural genes, the abundance, diversity, and arrangement of repetitive DNA, and the constitution of subgenomes in

[1] Current address: Friedrich Miescher-Institut, P.O. Box 2543, CH-4002 Basel, Switzerland.
[2] Permanent address: Instituto Multidisciplinario de Biología Vegetal (IMBIV), Casilla de Correo 495, 5000 Córdoba, Argentina.
[3] To whom correspondence should be addressed. E-mail amatzke@oeaw.ac.at; fax 43-662-63961-29.

polyploids. Transgene silencing has been reported for diploids (e.g., Arabidopsis, petunia, tomato, rice, and *Nicotiana sylvestris*), autopolyploids (e.g., potato), and allopolyploids (e.g., tobacco [*N. tabacum*] and wheat [Srivastava et al., 1996]) (reviewed in M.A. Matzke et al., 1994; McElroy and Brettell, 1994; Pawlowski and Somers, 1996). These species all have distinctive genomic attributes that might influence the frequency and types of silencing effects that are observed. For example, where a transgene integrates relative to gene-rich regions or to specific classes of repetitive DNA in a particular species could affect its own expression and/or its ability to interact with other transgenes or endogenous genes that are flanked by a member of the same repeat family. In allopolyploids, integration of transgenes into one or more constituent subgenomes could lead to silencing effects not evident in plants with less complex genomes. The repetitive character of many transgene loci that are subject to silencing might reflect how natural repeats are treated in a given species; that is, the degree to which multiple copies of transgenes or transgene loci comprising repeated elements are tolerated or modified (e.g., by methylation) could depend on the extent to which endogenous repeats had accumulated and become modified in the genome of the host plant.

To encompass all of the aforementioned aspects, analyses of transgene expression should include the determination of (1) transgene copy number and arrangement, (2) the nucleotide sequence and repetitive nature of flanking plant DNA, and (3) the chromosomal location (including, in allopolyploids, the subgenomic allocation) of the integrated transgene. Although progress has been made in assessing the influence of these parameters individually on the expression of different transgene inserts, an integrated study incorporating all of them to compare genetically well-characterized transgenic lines has not been published.

Here, we present results from such an analysis with four transgenic tobacco lines that contained the same construct integrated into distinct chromosomal locations but that exhibited differences in the stability of transgene expression. In two of the lines, the transgene insert was stably expressed in the homozygous condition for many generations, whereas in the other two, marked instability of expression was displayed after several generations of homozygosity. Results are also presented on stability of transgene expression in $F_1$ progeny of intercrosses among the four lines and on the susceptibility of each transgene locus to inactivation in the presence of a *trans*-silencing locus.

## RESULTS

### Transgenic Lines

The H transgene construct used in this study (Figure 1) is composed of two chimeric marker genes: one gene (*hpt*) encodes resistance to hygromycin, and the second gene (*cat*) encodes chloramphenicol acetyltransferase activity. Both coding sequences were under the control of the 35S promoter of cauliflower mosaic virus. The H construct also contained at the left border a β-glucuronidase (*gus*) coding sequence adjacent to a minimal 35S promoter (35Spro). Transgenic plants could be screened for GUS activity to determine whether general transcriptional enhancers were present in flanking plant DNA. A final feature of the construct was an *Escherichia coli* origin of replication from pBR325 that permitted rescue cloning of the transgene inserts together with some flanking plant DNA.

The expression and inheritance of the *hpt* and *cat* genes in four independent transgenic tobacco lines have been described previously (Neuhuber et al., 1994; Matzke and Matzke, 1996). Although initially these lines had been retained for further analysis because they each appeared to contain a single active T-DNA locus, cultivation for several generations as homozygotes revealed differences in the stability of expression of the *hpt* and *cat* genes. In two lines, H9 and H83, stable expression of both genes was observed in sexual progeny for up to eight selfed generations and in all plants regenerated from single cells of leaves. Moreover, the

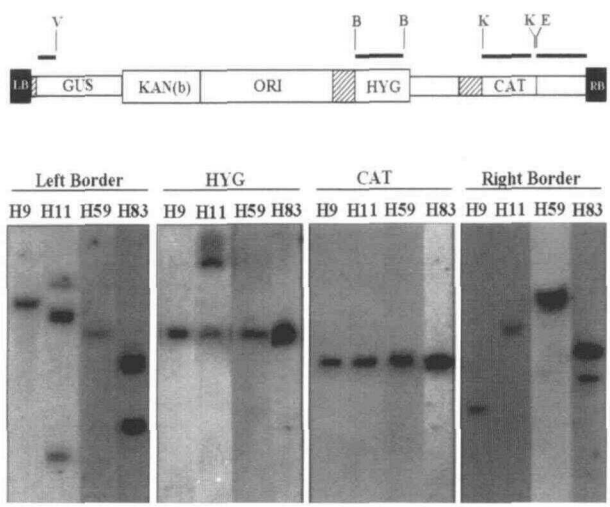


**Figure 1.** H Construct Map and T-DNA Border Analysis.

The H construct consists of a T-DNA left border (LB); a β-glucuronidase (GUS) coding sequence under the control of 35Spro; a bacterial selection marker (KAN[b]); a bacterial origin of replication (ORI); genes encoding resistance to hygromycin (HYG) and chloramphenicol acetyltransferase activity (CAT), both under the control of a 35S promoter (hatched region); a split nopaline synthase gene (unlabeled white boxes); and a T-DNA right border (RB). T-DNA border analysis was performed using EcoRV (V) and EcoRI (E), which cut in the T-DNA (the second sites were in plant flanking DNA). Internal HYG and CAT fragments were detected by using BamHI (B) or KpnI (K), respectively. Probes are indicated by the dark bars above the map.

H9 and H83 parental plants and their progeny contained detectable GUS activity in leaves (data not shown).

In contrast, the expression of the *hpt* and *cat* genes in the other two lines, H11 and H59, became progressively unstable after several generations of homozygosity. In addition, the H59 parental plant and its progeny were GUS-negative (H11 did not have an intact *gus* gene). The nature of the unstable expression differed in the two lines. When homozygous (*HH*), line H59 behaved as if it had an epigenotype of *HH** (i.e., one active [*H*] and one inactive [*H**] allele). The *H** allele was transmitted sexually and was semidominant to the *H* allele. Consequently, strong hygromycin resistance was obtained most frequently in backcrosses to untransformed tobacco when the two "epialleles" segregated away from each other.

The homozygous H11 line initially produced 100% hygromycin-resistant progeny, which subsequently turned yellow and died 6 to 8 weeks after germination on selection medium. Regeneration from single cells of leaves yielded plants that had one of three possible epigenotypes, *HH*, *HH**, and *H*H**, suggesting that leaves developed over time into epigenetic mosaics (Neuhuber et al., 1994; Matzke and Matzke, 1996). The demise of seedlings after 6 to 8 weeks on hygromycin-containing medium was presumably the result of a preponderance of *HH** and *H*H** epigenotypes. In contrast to the H59 line, the inactive *H** allele in the H11 line was not transmitted sexually; that is, the *H** allele was only recovered from regenerated plants in which it was then meiotically heritable for at least two generations. The meiotic heritability of the silenced *H** alleles and the increased promoter methylation of the *hpt* gene are consistent with transcriptional silencing (Neuhuber et al., 1994; Matzke and Matzke, 1996).

## T-DNA Structure and Sequences of T-DNA–Plant DNA Junctions and Flanking DNA

Although the initial genetic analysis indicated that each of the four lines contained a single active transgene locus, a T-DNA border analysis revealed one to three more or less complete copies of the T-DNA construct at each locus. The copy number did not correlate with stability of expression: the two stably expressed lines, H9 and H83, appeared to contain one and two copies, respectively; the unstably expressed line H11 appeared to have three left border fragments and one right border fragment; the unstable line H59 had a single left and a single right border fragment (Figure 1, left and right border probes). All four loci contained intact coding sequences for the *hpt* and *cat* genes; only the unstably expressed *H11* locus contained an additional larger HYG-hybridizing sequence that deviated from the expected size (Figure 1, HYG and CAT probes).

Rescue cloning was used to recover the T-DNA inserts and varying amounts of flanking DNA. The results from DNA sequence analysis and restriction enzyme mapping of these clones are summarized in Figure 2.

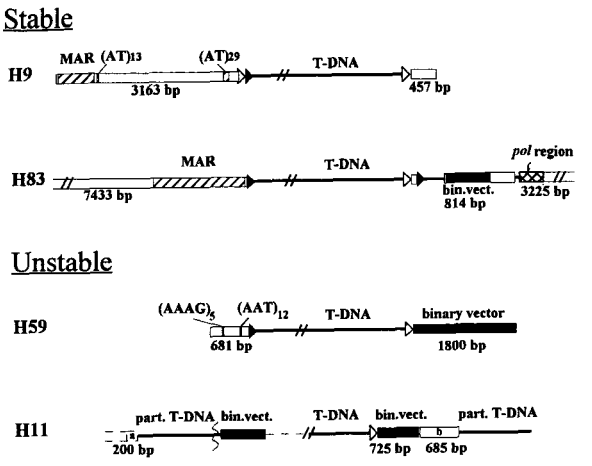


**Figure 2.** T-DNA Structure and Features of Plant Flanking DNA.

T-DNA sequences are indicated by heavy black lines. Left and right T-DNA borders are denoted as black and white arrowheads, respectively. Plant DNA is represented as white bars and binary vector sequences as black bars. Matrix attachment regions (MARs) and AT microsatellites are hatched. The retroelement remnant (*pol* region) is crosshatched. For line H11, nonoverlapping rescue (left) and λ (right) clones were obtained; these could not be fully pieced together (thin dashed line), and the exact relationship to the T-DNA and binary vector sequence in the rescue clone could not be determined (wavy verticle line). The drawing is basically to scale (1 cm = 1.2 kb); diagonal lines indicate regions that have been truncated. If known, the size in base pairs of the plant DNA or binary vector fragments is shown below the bars (except for the *H83* right region, in which sizes of interspersed fragments are provided in the text). The extent of left plant DNA in the rescue clone (a in *H11*) could not be determined accurately (dashed bar). The EMBL accession numbers for the *H59* left flanking DNA and *H83* right flanking DNA, including the *pol* region, are 412534 and 412536, respectively. bin. vect., binary vector; part., partial.

### *H9*

The simplest T-DNA structure was observed with the stably expressed line H9, which contained (as predicted from the border analysis) a single copy of the H construct bounded by nearly complete left (−4 bp) and right (−2 bp) T-DNA border sequences (Figure 3). This relatively normal configuration was slightly complicated by a partial duplication (20 bp) of right border sequences adjacent to the left T-DNA border. Contained within this 20 bp was 6 bp of sequence homologous to plant DNA at the target site (AGTTTA) (Figures 3 and 4). Flanking plant DNA sequences of ~3.16 kb to the left and ~460 bp to the right of the T-DNA were recovered (Figure 4). No extensive similarity was found with any known sequences. Several AT-rich regions were present in the left flanking plant DNA: two microsatellites, $(AT)_{29}$ and $(AT)_{13}$, and ~640 bp of an AT-rich region that bound to tobacco nuclear matrices in vitro (Figure 5). Because the H9 insert was relatively simple, it was possible to use primers

**Theoretical T-DNA Left Border:**

TGG▼CAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACTT

**H9:**

CTTGGTGAGTTTAAACTATCAGTGTTTAATATATTGTGGTGTAAACAAATTGACGCTTAGACAACTT
(right border)

**H83:**

CCGGCTAGTTGTGTAAAGATCCCTTTACTTTTCTCCACTGACGTACAAATTGACGCTTAGACAACTT
left border #1

**H59:**

TCCAAGGACTGTTTCCATACTTACGCCGTTGAATTGCATTCCAAGAAAGTCACCTATTAGACAACTT

**Theoretical T-DNA Right Border:**

TTTCCCGCCTTCAGTTTAAACTATCAGTGTTTGA▼CAGGATATATTGGCGGGTAAA

**H9:**

TTTCCCGCCTTCAGTTTAAACTATCAGTGTTTaaactgaagTTTGATTCGGCCACGAACTTTTACCTG
9 bp filler

**H83:**

TTTCCCGCCTTCAGTTTAAACTATCAGTGTTattaataacacgtattaccgcctttgagt

gctacagtgttattgtcaatttgtttatcagtgtaaagtgataaaCAAATTGACGCTTAGACAACTT
74 bp filler left border #2

**Figure 3.** Nucleotide Sequences of Plant DNA–T-DNA Junctions.

Theoretical left and right T-DNA borders based on the VirD2 nicking site (arrows) (Tinland and Hohn, 1995) are shown. T-DNA sequences are underlined. Plant DNA sequences are not underlined. Filler DNAs of unknown origin are shown in lowercase letters. Right border sequences found to the left of the *H9* left border are shown in italics. The boxed regions show the short region of homology found between the T-DNA right border and plant DNA at the *H9* insertion site (Figure 4).

homologous to the left and right flanking plant DNA to clone the target site from untransformed tobacco. A comparison of the target site sequence with the sequence of the T-DNA–plant DNA junctions revealed that 32 bp of plant DNA had been deleted upon T-DNA integration; a 9-bp filler DNA sequence was interposed between the right T-DNA border and plant DNA, and no filler sequence was found between the fused left–right border and plant DNA (Figures 3 and 4).

### *H83*

The stably expressed H83 line, which had two left and two right border fragments, according to the DNA gel blots, was found to contain one intact copy of the H construct together with a largely truncated second copy and three borderless fragments from the T-DNA or binary vector that were interspersed with putative plant DNA (Figure 2). The intact copy of the H construct was bounded by partial T-DNA border sequences: the first left border lacked 20 bp relative to theoretical sequence, and the right border lacked 3 bp (Figure 3).

CCTAGGAGTATCACGGCTAATCTTTGGGCCTTTAGTTGGCAATTGTGGTC
AATTATAAATTTAGGTGCCTCAGTCCTACGTATAATTTAATTTTAAATTT
TTAATTTTAAATTCAACATCAAGTAAGTAATAGCAATAATATTTTTAATG
GAACATTGAAATTTTAATCAAAATTTTTAAATCGGCAACTTTTTTAAA
AAGGGATTAACGTTTTCTTATACGGGATCTTAAAACTTTTTTAATAAATT
ATAAAAAATAGCATAATAAAATAATATAGAAAGCACATTAAAAATTTAAT
AAAATTAAAAAAGATTAAAGAAACTCATAGAAAAAATAAAATATCTCGAAT
TCTTTTATTAGTATAAATAACACAATAATGTTAAGTTTATTATTATAAAT
CAACTATAACGTTTGAACTTTCAAATAATTAAAATATTAAATTTTTAAAA
ATATATTAACAAATTACAAATTGAATTACCTCTCAAAAAGTAAATAATCA
ATAAACTTTATAAGTACTATAATTTAAAATATTACTCCTTCATAAATAAG
TAAAATTGTAGTAAAAGAATAAGTCTCATAATTTTGGAAGACATAAAGCT
AGGAAATAAAGAGCAATAGCAGGTGAAGATATAAATTTCGGCTATTCATT
TTCAAAAGAAATATTCAAATGATATGGCCTCACGAGCTCTCACTCAAATA
GTAAATATGCAATATTTTGACTCGATATCTTAATTAATCTCACATCTTAT
ATATATATATATATATATATATGTGTGTGTGTGTGTGTGTGAATTGAA
CTTTTATCATTCATTTGTTAAAATTATTTTGAGGGTCAATGTTACTAATT
AGAAAATTTAACATATGATTTATGTAGTAACTCCCGAACACTAAGGATAG
GCATGTTCAGGATGCACAATTAGACATTTGGGGAGTAAGCCTCACACATG
AGCCATCGCAACATGTGCCTCGCCCAGGAGTGAGTCATGGACGAGTCGAG
AAAGACTTTTAKAGACATTAATCGACATCTTATATCAAATGTTATTTTTT
TGCAACATGTAGTATCAATCGTTGTAGTTTTCACAACATTTCTTCATAAA
TTTGACTGTTTTTACGATATTTCTTCATAAATGTTATTGTTTTTACGATA
TTTAGTATCAAATGTCATTTTTTATAACATTTAATATCAAATGTCATTAT
TTTTATTACATTTTGTATCAAATGTCGTAATTTTGACGATATTTCTTCAT
TAATGTCGTTATTATTACGACATTTTGTATCAAATGTCATATTTTGATGA
CACTTAGTATCAAATGTCGTAATGTTTGCGACACTTTGTATCAAATGTCG
TAGTATTTACGACATTTATTCATAAATGACATGAATTTCATATTTTTCTG
AGATTTATTTCAAATGTCACTAAATAAATGTCGCCGTATCCCTCACTTTC
TTGGTAGTGGGTTGGACGGGAGGTCATGAATTTTGTACTACACTTTTGCA
TCTTGCATGCAAATGTTGGATGTTGATGTTGCTGTGATCGACGAGAGTTG
GATTTGAAGACGTACCTGTATTCCGGTTGTAGCTGCATCTTGTTCATGGT
AGCCTTAGATTTATAAAAAATCTGTTTATGTACTTTTCGAACAGATGATG
TATTTTTTTTTATATTAGTTTTGTAAATTCTAAATCTTAGAAGCTCATGA
TTTGTACTACCAATCCTTGAGAAATTTTTGTATAAAAACAGTTTTATTAT
CATTTCTTTTCTTAATAAATCTCATTTAAATTGGATAGTTGTTAATTGAC
TTATCTGACGGGTTAAGTTATGTGCCATCACGACTAGGTGATTTTGGATC
GTAACATTAATCATGAACTAACGGACTATATCCTAATATCCTTAAACATT
AGGATAAGAGGTGAAAAAAGACACTTGAGAGCCTACCACAAAAATCTGGT
ACAATTAAATCCATCTCTGAATACCTAATACTATATAGTTATCCCAACAC
CATCCCCCGCTTGTTATAGAAATATATTGAAAATAAATCAATGGAGTTAA
CTCAAGCAAGATAATCTAATAAGAAATCTGAAAAAGTAGACAAGAGTTTG
AGGAAATTATCTTTAGGATAGTAATTTCCCCCACACGATGAAATGGTTAA
AGGAAAATCTATCCCCAAAGATACAATGCCTACTCTAGAAATCAAAGTCG
CACTCTTTGATTTCCCAACCACAAGAACTCAACTCGAAAGAATGAAGGGT
TTGGAAGAAAGTATTACGAAGATTTGAGTATATAAATAGATCTAATGATG
GGACAAGAGAGATGTTGTATATAAGATTGCCTAGAAGTTGTTTAAGACTT
CTGTAGAATACCCCTTTATATAGATAAAGTTTATGACCTTTGCCAATAGA
TGCTTCATTTAAATTTAAACAAACAGTTAACAAAGAGTTAGTTACTCATA
GAAAAACACATCAGTTAGTTCCACAGAAGGCAACGGTTGAATTCAAATTT
AAGATAGAATTCTTTTTCCTGTTTTTTAAAATAAATATTATTATTAAATA
ATACATATATATTATTTTATATATAACACCGCTATCTTTCCCTCTTTGGA
ACATTATTGCCAAGAACATGAAACTTTCTGCTGTCAGAGGCGGACGTAGC
CATATAGGTGCGGGTTCGGTCGAATTCAGTCATTTTTGTTTAAACATTAT
ATTTATACGAAGAAATTTATTAAATATGTACAAATATTTAATTGCGAACC
CCGTAACTAAGACGATCTATGGGTTCCGATGCAAATTCAGAACCCATAAA
CTTTAAATCCTGAATTCGTCTCTGTCTACAGTCTCATATATATATATATA
TATATATATATATATATATATATATATATATATATATATATACACACA
CGTAATGCTATACAACATAAGCTAGTATTTTAAGATTGTATTGCTTGTGA
AAAGTCATACAGTTGATGCTGTTTGAACAGCTTCTATTTCATGCTTAACT
ATGATAGGCAAGTTTTACCTAGTGCACAGCCAAATGCACCCAAGGGATTG
TCAACCGACGTCCTTTTGTCAAAAAATTACACTATTTAGTTCGATAATAT
TTTAAAAAAAATATGTATATATACTATATAATGGATACTTTTTTTGACTT
CTTGGTGAGTTTATtttcttcatattttgacttaccttaatgaaaaTTTTG
ATTCGGCCACGAACTTTTACCTGCAATTATTTCAGTTCATCCATATCAAT
CAATGATGTGACAGATTTGTAAGAGTTAAGATAACGAGAAAGATCCAAAC
ATATATATGTCTCACACAACATGGAAAAGGCATTTTTTTGTGTGTCTAAT
ATCAAACTCCAACCTAACTTGCTAGGATCATTGACATTGTCACATGGCTT
ATCATTTTAGTTATCAGCTTTTGATAATTGAAGAAACAACTTTCACTGGC
GAGTGGTAACCCAAACACAATGCCAAGAGCCAAGGGAAGAGACATCAGGA
TTCCGAAGAGACAGCATTAAGGGGTGAGTGCGAGTATACAAGAGTCTGAA
ATGATATACTCTGCTATCTTTTCATGTCACTGGTTCTTTTTTGTTGTCAC
TACTCTTCGAGAAGCCTAGG

**Figure 4.** Tobacco DNA Sequence at the *H9* Target Site.

The arrowhead adjacent to the boxed region indicates the site where the *H9* T-DNA integrated into the tobacco genome, resulting in a 32-bp deletion of plant DNA (lowercase letters). The boxed region shows the short region of homology between plant DNA and the T-DNA right border sequence (Figure 3). AT-rich regions are indicated by white letters on a black background: these include two AT microsatellites and an ~640-bp MunI-EcoRV fragment (CAATTG and GATATC recognition sequences are underlined) that binds to tobacco nuclear matrices in vitro (Figure 5). The EMBL accession number of the *H9* target site is Y12535.

To the right of the intact T-DNA copy was 74 bp of filler DNA of unknown origin, followed by a second left T-DNA border (also missing 20 bp, as was found with the first left border) and a T-DNA sequence extending only 160 bp into the *gus* coding sequence, which then broke off and led into a 24-bp DNA fragment of unknown origin, 814 bp of binary vector sequence, 492 bp of (presumably) plant DNA, and 30 bp of the 35Spro plus 96 bp of the *gus* coding sequence (Figure 2 and data not shown). These fragments of the H construct, which were not associated with T-DNA borders, were followed by ~3.2 kb of uninterrupted plant DNA. As shown in Figure 6, this sequence contains a putative open reading frame that encodes a polypeptide with significant amino acid similarity to the integrase region of several retroelements.

In addition to the ~3.2 kb of plant DNA to the right of the entire complex T-DNA insert, ~7.4 kb of plant DNA to the left was sequenced. Figure 7 shows an ~1.8-kb flanking sequence directly adjacent to the T-DNA left border; this sequence contains three AT-rich regions. Two restriction enzyme fragments (2.2 and 2.9 kb) comprising these regions bound to tobacco nuclear matrices in vitro (Figure 5). Extensive similarity to known sequences was not found in left flanking DNA.

Although several different primers were tested, the T-DNA target site in the H83 line could not be recovered from untransformed tobacco. This was possibly due to postintegration scrambling of multiple T-DNA copies and associated plant DNA, as suggested by the presence of borderless T-DNA fragments to the right of the intact copy.

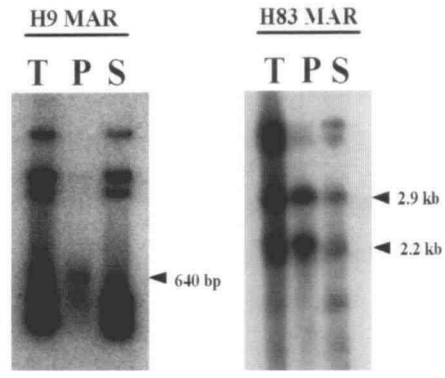


**Figure 5.** Nuclear Matrix Binding Assays.

T indicates total end-labeled fragments derived from rescue clones containing the *H9* and *H83* loci (doubly digested with MunI-EcoRV and XbaI-HindIII, respectively; see Methods) before incubating with tobacco nuclear matrices. After incubation and centrifugation, specific fragments of ~640 bp (*H9*) and 2.2 and 2.9 kb (*H83*) partition more completely with the matrix preparation in the pellet (P) (arrowheads), identifying these fragments as MARs. Their positions relative to the respective T-DNAs are shown in Figure 2 and in the DNA sequences shown in Figures 4 (*H9*) and 7 (*H83*). S, supernatant containing unbound fragments.

### *H59*

As expected from the border analysis, the unstable H59 line contained a single intact copy of T-DNA. Although the left border (shortened by 32 bp from the expected sequence; Figure 3) was fused to plant DNA (681 bp of which were recovered and sequenced; data not shown), the right T-DNA border led directly into binary vector sequences (of which ~1800 bp were present in the rescue clone) (Figure 2). The plant DNA to the left, which consisted of previously unreported sequences, comprised two microsatellites: $(AAAG)_5$ and $(AAT)_{12}$ (Figure 2). We have not yet been able to obtain a rescue clone of this locus that contains plant DNA to the right.

### *H11*

As suggested by the border analysis, the second unstable line (H11) harbored a complex transgene locus that comprised at least three truncated T-DNA copies (Figure 2). Although one rescue clone and one λ clone containing partial T-DNA copies were recovered from this line, we were unable to piece together the complete structure of the *H11* locus. Two of the three truncated T-DNAs were adjacent to binary vector sequences, although in only one case were these vector sequences directly contiguous with a standard right T-DNA border sequence (Figure 2). Extensive flanking plant DNA was not obtained in either the rescue clone or λ clone, although 685 bp of putative plant DNA that was present between the binary vector sequence to the left and pBR325 (i.e., T-DNA) sequences to the right (Figure 2, *H11*-b) was recovered and sequenced; noteworthy features were absent (data not shown). An unambigous sequence from the left flanking plant DNA (Figure 2, *H11*-a) could not be obtained with the T-DNA primers tested, presumably because of sequence duplications elsewhere in the clone.

## T-DNA Chromosomal Location, Subgenomic Allocation, and Species Specificity of Flanking Plant DNA

To characterize further the genomic environment of the four transgene loci, we examined two aspects. First, fluorescent and genomic in situ hybridization (FISH/GISH) were used to establish, respectively, the chromosomal location of the transgene insert and its subgenomic allocation (*N. tabacum* [tobacco] is a natural allotetraploid [2n = 4x = 48] derived from two diploid [2n = 24] progenitors: *N. sylvestris* [S genome] and *N. tomentosiformis, N. otophora,* or *N. tomentosa*, all of which are very closely related [T genome]; Okamuro and Goldberg, 1985; Parokonny and Kenton, 1995). Second, to investigate the possible repetitiveness and species specificity of the flanking plant DNA sequences, the rescue clones containing left and right flanking sequences were

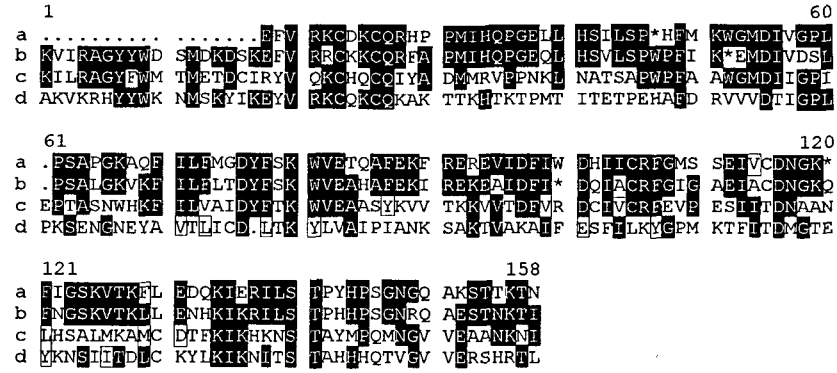

**Figure 6.** Comparison of Deduced Amino Acid Sequences of the Putative Retroelement *pol* Region in Plant DNA Flanking Two T-DNA Inserts.

Sequence a is right flanking plant DNA of *H83* (*H83R*), nucleotides 1 to 400; sequence b is flanking plant DNA of a *K* locus, nucleotides 3456 to 3926 (Papp et al., 1996; EMBL accession number Z71319). The most significant similarity was found to sequence c, the upstream region of the tobacco defense-related pseudogene 246N1 (Froissard et al., 1994; PIR protein database accession number S47444). These sequences are also highly homologous to the integrase region of the Pol polyprotein of a number of different retrotransposon sequences, the most similar being the Osvaldo retrotransposon and Tn*412*, both from Drosophila. Tn*412* is shown in sequence d (Swiss-Prot accession number P10394). Sequences a to c are not highly homologous to tobacco retrotransposon *Tnt1*. Identical amino acid residues are shown as white letters on a black background.

used as probes on blots containing DNA isolated from tobacco, *N. sylvestris*, and *N. tomentosiformis*.

As shown in Figure 8, many of the flanking plant DNA regions produced hybridization patterns characteristic of dispersed, moderately repetitive sequences, that is, smears occasionally overlaid by some bands. With respect to species specificity of the dispersed repeats, several variations were observed. When a rescue clone containing ~500 bp of left flanking DNA from the stably expressed *H9* locus was used as a probe, comparable signals were observed in all three species (Figure 8, *H9L*). When a second rescue clone containing an additional 2.6 kb of plant DNA to the left and ~460 bp to the right of the *H9* insert was used as a probe, signals of similar intensity were again observed in all three species, although this probe also hybridized with a tandem repeat (ladderlike pattern) that was enriched in the S subgenome but not found in any of our *H9* rescue clone sequences (Figure 8, *H9R+L*). FISH/GISH demonstrated that the *H9* locus is near the long arm telomere of a subtelocentric chromosome from the T subgenome, T3 (Figure 9, *H9*, and Figure 10). Moderately repetitive dispersed DNA that was common to all three species ("species mixed") thus appeared to surround the *H9* locus; some copies of this sequence were apparently connected to a tandemly repeated region that was present primarily in the S subgenome.

A different pattern was observed with *H59*. Unlike the species-mixed dispersed repeats that surrounded the stable *H9* locus, the 681 bp of left flanking plant DNA of the unstably expressed *H59* locus hybridized with moderately repetitive dispersed sequences that were enriched in the *N. tomentosiformis* subgenome (Figure 8, *H59L*). The FISH/GISH analysis, however, indicated that the *H59* locus was present in an intercalary region of the long arm of a subtelocentric chromosome from the S subgenome, S11/t2 (Figure 9, *H59*, and Figure 10). The *H59* insert was thus flanked on the left by "species-incompatible" moderately repetitive plant DNA (in this study, we define a T subgenome–enriched repeat that is present on an ancestral *N. sylvestris* chromosome as species incompatible) and directly contiguous binary vector sequences on the right (these produced weak hybridization signals on the plant DNA blots; Figure 8, *H59R*).

One of the few examples from the four lines of flanking plant DNA that appeared to be of relatively low copy number was in the left flanking region of the stably expressed *H83* locus (Figure 8, *H83L*). The plant DNA to the right of the *H83* insert was moderately repetitive. Although a signal was observed in all three species, two bands were enriched in the *N. tomentosiformis* genome (Figure 8, *H83R*). This DNA gel blot pattern was virtually identical to that obtained with the plant DNA adjacent to a T-DNA insert analyzed in a separate study (Papp et al., 1996; I. Papp and A.J.M. Matzke, unpublished results) and was almost certainly due to the retroelement remnant present in both flanking DNA sequences (Figure 6), which otherwise shared no additional similarity over ~2 kb. The FISH/GISH analysis localized the *H83* insert in the vicinity of a telomere on a small metacentric chromosome from the T subgenome, T9/t3 (Figure 9, *H83*, and Figure 10). The stably expressed *H83* insert was thus adjacent to a "species-compatible" repeat (this term is used here to refer to a T subgenome–enriched sequence that is present on a chromosome from the T subgenome).

The unstably expressed locus *H11* was present in a paracentromeric location on a small metacentric or submetacentric chromosome from the T subgenome (Figure 9, *H11*).

Because this chromosome was one of three from the tobacco (cultivar SR1) karyotype that contained no additional markers for identification (Moscone et al., 1996), we could not ascertain whether it was T5, T8, or T12 (Figure 10). DNA gel blots were probed with the rescue clone (*H11*-a) and the λ clone (*H11*-b) (Figure 2). Both probes produced a species-mixed pattern, with *H11*-a appearing to be moderately repetitive and *H11*-b of relatively low copy number (Figure 8).

### Intercrosses

To determine whether the unstable loci (*H11* and *H59*) could influence the expression of the stably expressed loci (*H9* and *H83*) and whether the stably expressed loci would remain so when combined in a single genome, intercrosses between all four homozygous lines were made, and $F_1$

AGGCCTAATTAGTAAAAGAAGAATTTACCTAAATAGTCGCTCATTTAGTA
GAGCTGGCAGATGTATAATATATGTAAAATTCATATATAAAATGAGTATA
AAGAAGGCAAAAATATATAAAGACCCCATTAAACTTGTTCGAAGTGAATT
AAATGCTACCCTAAAGGCTTATATGAAAGAAAAGGAAAAAAAAAAAAAAA
AGAGCGTAAAATTTTTTAAAAAATGGGCCCGCTGATAATATGTAATAAAA
TAAAAATAAAAAATAAAATTGGGCTCCTCTCATTTTTTTAACTACCACC
TGCCTTCTTTCTCCTTCTTCTAACCCCCCCCCCCCCCTCTATTCGCCCAG
CTATTCATCTTCTTCCTTTCCCACTATCTGCTCGATCTTTTTCTTTTAA
TTTTTTTCTTTCTCGCCATCTGCCTCATCTTCTTTCTTTTTCTTTTTTAT
TCAGTCCCAACATATTAAGAACATGATAATGAAATTTTCTTTGTTCTAAA
AATTAAGAAATTAATATAGCATCTTTTCGTATTTTCTCTGTTGCGGTTAA
TTAAATAAGAAAAAAATCATAGTGGATGTTGAAGATTGAAATAAAAAATTA
AAAATAATTAATATGATAAGATAACGAAACAAAAAAAAGAGGAGAGTAAAC
CAAGATCGAAGGGATTTAAATTTGTAGGATATACGGATGAGAATTCGATG
AAAAAATATAATTAAGAGGAATTTATTAATTAATTAGAGATTTATTAGTG
TATTAAATAAAGAGCAAGGAAAAAAATATGATAGACAAAAAAGAAACGAA
AAGGTTAAAAAAATTAAAATTTAGATAATTTCTGATGTGGCTCTAACGTG
GTGCTGATATGGCTCTGATATGGCAATGATGTGGCGCGTGTGTAAAGCAC
ATCACATTGTGAGGTTGGTATTATGTGTTAGGGTAGCATTTAATTCACTC
CGAACAAGAATAGGGGTAAAAGACACCCGTTTAATTAAGAGTTTAAACCT
ATTTTTTGAACAAGTTTAATGGGGTCTTTATGTATTTTGCCTGTAGAGAC
GGATCCCAGGACTTTTATAAGGTTCAAATAAGTTCCCAACATGACTGTTG
TAGTCACAGTCATTTCATGGGAGCTAGTAGTTACACATAAGAGAAGGGGG
GAAAAGGCAAGTGTGCAATTGAATATTTGAAACATGTCGTTTATGATTGA
CGTACAACGGAGTTTTTAAATGATTTTTAAATGATCAAGATGATGCAAAT
GATATGTATATAATTAGCGTATAATTTATATTAATTTTCTAGAAGGTCAA
ATGAAGCTATGTGGGCGGGCCTGATTAAAGTTTTCATTTCATGAATAACC
TATACAAATTTCACTAATTATTTTCACTCTAAACCAAATTGTCAAAATGC
AATCTATTGCTACCTTAATTTCATTTTAGTTGAAATTGGATATACTTTCA
CTTTCATTATTTATATAACCGTGTCCAAATCAAATGACACCGCATGTGGC
TAAAAACTCATAAGTTCTTTTTTCATCATTTACTTTTGAAAAATTAACCC
AAAATGTTGCTCACCTAACCGCTTAAACTAAAAATAGCCGGTGGAGGTAT
AATATATACATAATTAATGTATTATATATGTATAATTGTGTATAATCAAT
ATATAAACTATGTATATAGCTAGAAAAAGTAAATAATCAATATGGCCGGC
TAGTTGTGTAAAGATCCCTTTACTTTTCTCCACTGACGTacaaattgacg
cttagacaacttaataacacattgcggacgtttttaatgtactgacggtc
gac

**Figure 7.** Nucleotide Sequence of the *H83* MAR.

This fragment of plant DNA, which is directly adjacent to the *H83* left T-DNA border (white arrowhead, third line from bottom; see also Figure 2) contains three AT-rich (>75%) regions (white letters on black background). The XbaI site used in the matrix binding assay is underlined (12 lines from bottom); HindIII sites also used in the assay are farther upstream and downstream in plant DNA and T-DNA, respectively (data not shown). An XbaI-HindIII double digest of the *H83* rescue clone generated fragments of 2.9 and 2.2 kb that bound to tobacco nuclear matrices in vitro (Figure 5). T-DNA right border sequences are in lowercase letters. The EMBL accession number for the *H83* MAR is Y12533.

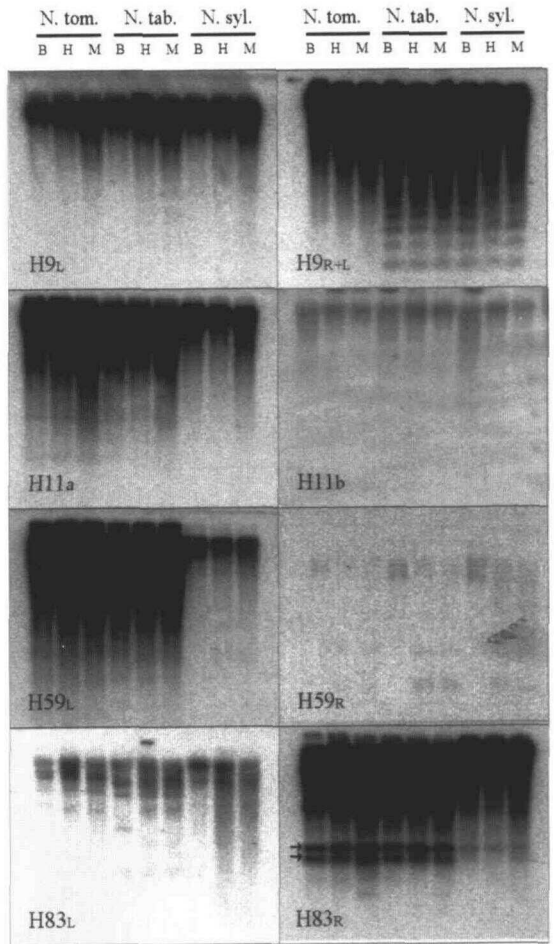


**Figure 8.** Gel Blot Analysis of DNA Flanking the T-DNA Inserts.

Each blot contains three lanes each of DNA isolated from *N. tomentosiformis* (N. tom.), *N. tabacum* (N. tab.), and *N. sylvestris* (N. syl.). DNA preparations were digested with BamHI (B) alone or with BamHI plus HpaII (H) or MspI (M) to test for possible methylation. The probes used are shown in the lower lefthand corner of each blot; L and R denote left and right flanking DNA, respectively, of the indicated T-DNA insert (Figure 2). *H9L* consists of ~500 bp directly adjacent to the left T-DNA border; *H9R + L* contains this region plus an additional ~2.5 kb of the left flanking plant DNA and 457 bp of the right flanking DNA (i.e., the entire *H9* target site; Figure 4). The positions of *H11*-a and *H11*-b relative to the T-DNA copies at the *H11* locus are shown in Figure 2. Arrows in *H83R* indicate bands enriched in the *N. tomentosiformis* fraction of the genome: these arise from a retroelement *pol* remnant, because a virtually identical pattern was seen when the probe was either flanking DNA isolated by Papp et al. (1996) or a tobacco repeat isolated by Kuhrová et al. (1991)—both of which also contain the *pol* sequence (Figure 6 and A.J.M. Matzke, unpublished results). This retroelement remnant is present in ~1000 copies per haploid tobacco genome (Kuhrová et al., 1991). Based on similar hybridization intensities with comparably labeled flanking plant DNA probes, we estimate that the other repeats shown here are also present in ~1000 copies per haploid tobacco genome.

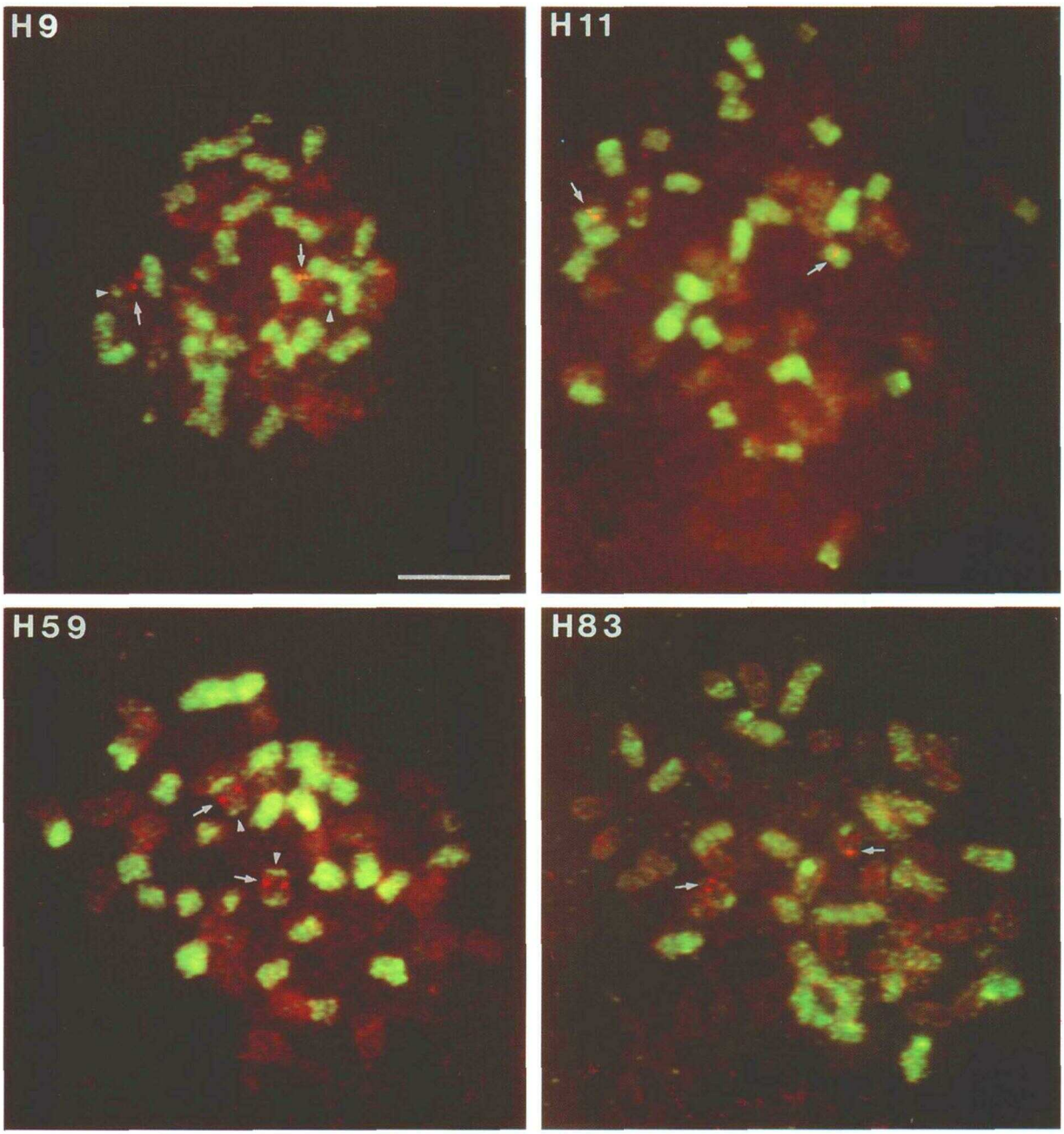


**Figure 9.** FISH/GISH of Somatic Chromosomes of Transformed *N. tabacum* cv Petit Havana SR1 (2n = 48) Carrying the H Construct in the Homozygous Condition.

Orange-red fluorescent spots indicate hybridization to the biotin-labeled H construct probe specific for the transgene insert, and green fluorescence indicates hybridization to the digoxigenin-labeled genomic DNA probes of *N. sylvestris* (S subgenome) (H9 and H83 plants) or *N. otophora* (T subgenome) (H11 and H59 plants). Unlabeled chromatin, which does not fluoresce, appears brown. Arrows indicate the location of the transgene insert. Cross-hybridization of the *N. sylvestris* or *N. otophora* genomic probes to T3 (in H9) or S11/t2 (in H59) nucleolus organizing regions, respectively, is shown by arrowheads. Green segments not indicated in the chromosomes carrying the transgene inserts S11/t2 (in H59) and T9/t3 (in H83) are intergenomic translocations (Figure 10). Bar = 10 μm.

seeds were sown on medium containing increasing concentrations of hygromycin. Although some weakening of the stable loci in the presence of one of the unstable loci was observed in $F_1$ progeny, this silencing was not dramatic; $F_1$ progeny from these intercrosses were still moderately resistant to hygromycin. When hybrid lines were selfed, the silencing was slightly enhanced (presumably in seedlings homozygous for the transgene loci); however, complete silencing of *H9* and *H83* in the presence of either *H11* or *H59* was never observed. Progeny of H9 × H83 crosses continued to stably express the *hpt* gene, as indicated by the undiminished strength of hygromycin resistance of $F_1$ seedlings (data not shown).

### Crosses to the *271* Silencing Locus

The four *H* loci were also tested for activity in the presence of the multipurpose silencing locus *271*. This locus is able to transcriptionally inactivate genes under the control of the 35S and 19S promoters of cauliflower mosaic virus as well as silence endogenous nitrite reductase genes of tobacco via a post-transcriptional process (Vaucheret, 1993; Park et al., 1996). It has previously been localized to the long arm telomere of a tobacco chromosome from the T subgenome, T3 (Moscone et al., 1996; Figure 10). The poor hygromycin resistance conferred by the 35Spro–*hpt* gene at the two unstable loci, *H11* and *H59*, was further weakened in the presence of *271* in newly germinated $F_1$ seedlings. In contrast, the *H9* and *H83* loci displayed reductions in *hpt* gene activity only after 6 to 8 weeks, manifested by mottling and yellowing of $F_1$ seedlings, which eventually died if maintained further on hygromycin-containing medium (data not shown).

## DISCUSSION

To study the influence of genomic context on transgene expression, we have determined the T-DNA structure, sequence of flanking plant DNA, and chromosomal integration site of four independent transgene loci in tobacco. Although similar analyses on additional transgenic lines are required before we can draw solid conclusions, the data obtained thus far provide a remarkably consistent picture. As summarized in Table 1, the two stably expressed transgene loci (*H9* and *H83*) contained one intact copy of the T-DNA construct with no contiguous binary vector sequences and were present in the vicinity of telomeres. They were also flanked on at least one side by plant DNA that (1) could act as a transcriptional enhancer (as indicated by a GUS-positive phenotype), (2) contained long (~0.5- to 1-kb) AT-rich regions that behaved as matrix atttachment regions (MARs) in vitro, and (3) was either of low copy or related to species-compatible or species-mixed moderately repetitive dispersed sequences.

In contrast, although the unstably expressed loci (*H59* and *H11*) were found to contain either a single intact copy or multiple truncated copies of the H construct, respectively, both comprised binary vector sequences that were directly contiguous with a right T-DNA border. In addition, both inserts occupied chromosomal sites that were distant from the ends of chromosome arms (intercalary in the case of *H59* and paracentromeric in the case of *H11*). Finally, the GUS-negative phenotype of line H59 indicated the absence of enhancer elements in left flanking plant DNA, which was of a species-incompatible moderately repetitive type (Table 1).

### T-DNA Copy Number and Configuration

Previous studies have established that single copies of transgenes tend to be more stably expressed than are multicopy or scrambled inserts (Meyer and Saedler, 1996). Although our data generally support these former results, factors other than T-DNA copy number are clearly involved. The unstable expression of the single T-DNA copy in the H59 line might have resulted from several features of the genomic context: flanking sequences that could be considered foreign (contiguous binary vector to the right and a species-incompatible repeat to the left) and the intercalary chromosomal location, which was distant from the gene-rich ends of chromosome arms (see below). The stable expression of the H83 insert, despite the presence of several T-DNA and binary

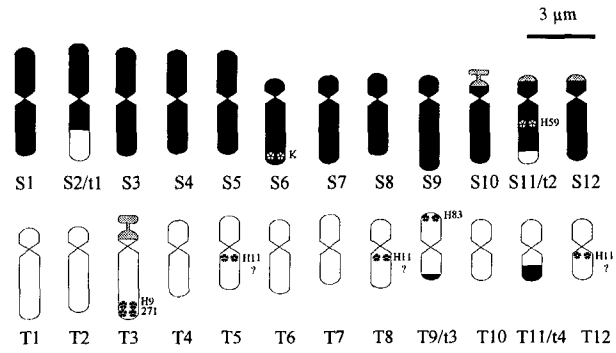


**Figure 10.** Summary of T-DNA Inserts Localized to Specific *N. tabacum* Chromosomes by FISH/GISH.

Black and white areas indicate S and T subgenomes, respectively. Four intergenomic translocations are present (S2/t1, S11/t2, T9/t3, and T11/t4). Dotted regions show rDNA loci (S10, S11/t2, S12, and T3), two of them active in nucleolus formation (S10 and T3). Transgene inserts localized so far by FISH/GISH include the four *H* loci described in this article (*H9*, *H11*, *H59*, and *H83*), the *271 trans*-silencing locus (Moscone et al., 1996), and a *K* locus (Papp et al., 1996). *H11* could be on T5, T8, or T12 (indicated by H11?), which are the only three *N. tabacum* (cv Petit Havana SR1) chromosomes that do not have a distinctive morphology or other physical marker (Moscone et al., 1996).

**Table 1.** Genomic Context of Stably and Unstably Expressed *H* Loci

| Line | HYG–CAT Expression | T-DNA Copy Number | Chromosomal Location | | Flanking Plant DNA[a] | | DNA Motifs[b] | | Binary Vector[c] | GUS Expression[d] |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Chrom. No.[e] | Position | Left | Right | Left | Right | | |
| H9 | Stable | 1 | T3 | Telomeric | Mod-rep mixed | Mod-rep mixed | MAR, $(AT)_{13}$, $(AT)_{29}$ | None | – | + |
| H11 | Unstable | 2 or 3 fragments | T5, T8, or T12 | Paracentromeric | Mod-rep mixed | ? | ? | ? | + | No intact *gus* gene |
| H59 | Unstable | 1 | S11/t2 | Intercalary | Mod-rep T on S chromosome | BV | $(AAAG)_5$, $(AAT)_{12}$ | ? | + | – |
| H83 | Stable | 1 intact (+fragments) | T9/t3 | Telomeric | Low copy | Mod-rep T on T chromosome | MAR | *pol* region | – | + |

[a] Mixed refers to a moderately repetitive (mod-rep) sequence that is present in comparable amounts in both the S and T subgenomes of *N. tabacum*. A species-incompatible repeat as defined in this article refers to a copy of a T-enriched repeat present on a chromosome from the S subgenome (e.g., H59); a species-compatible repeat refers to a copy of a T-enriched repeat on a chromosome from the T subgenome (e.g., H83). BV, binary vector sequences; ?, not known.

[b] MAR, matrix attachment region; ?, presence of motifs not known because flanking plant DNA was not recovered in rescue clones; *pol* region, retroelement remnant.

[c] The plus and minus signs refer, respectively, to the presence or absence of binary vector sequences that are contiguous with a right T-DNA border (see also Figure 2).

[d] A plus sign indicates that a plant and its progeny contained GUS activity in leaves; a minus sign indicates no GUS activity.

[e] Chromosome numbers refer to those from the T and S subgenomes, as indicated in the karyotypic diagram of *N. tabacum* chromosomes shown in Figure 10.

vector fragments, was presumably due to the single intact T-DNA copy that was flanked by relatively complete border sequences leading directly on both sides into plant DNA.

The sequences of T-DNA–plant DNA junction fragments and the target site sequence from line H9 confirm features of T-DNA fine structure that have been obtained from previous studies of T-DNA integration events (reviewed in Koncz et al., 1994; Tinland, 1996). The right T-DNA borders were better conserved (compared with the theoretical expectation based on the VirD2 nicking site) than were the left ones (Tinland and Hohn,1995). In line H9, the 6 bp of homologous sequence between plant DNA directly at the site of insertion and a short region outside of the left T-DNA border (consisting, in this case, of right T-DNA border sequences that were fused to the left border) is consistent with the short patches of plant DNA and T-DNA homology (7 to 14 bp) found for a number of other T-DNA integration events (Matsumoto et al., 1990; Ohba et al., 1995; Tinland, 1996).

The presence of binary vector (non–T-DNA) sequences directly contiguous with a right T-DNA border was found in both of the unstably expressed lines H11 and H59. Binary vector sequences appear to be transferred relatively frequently into plant genomes, at least as assessed by DNA gel blot experiments (Martineau et al., 1994; Ramanathan and Veluthambi, 1995; Cluster et al., 1996). These previous studies have not addressed whether these binary vector sequences are actually continuous with T-DNA or whether they affect the expression of adjacent transgenes. A recent cloning and sequencing study revealed T-DNA inserts that led continuously from the right border into a complete copy of the binary vector (Van der Graaff et al., 1996).

## Chromosomal Location

The stably active inserts *H9* and *H83* were found adjacent to telomeres. T-DNA might preferentially integrate close to telomeres, as other recent FISH studies have suggested (Wang et al., 1995; ten Hoopen et al., 1996). The distal ends of chromosome arms contain high concentrations of genes in wheat (Gill et al., 1993), maize (Bernardi, 1995), and humans (Saccone et al., 1992). Although a (sub)telomeric location of stably expressed transgene loci might seem inconsistent with previous data showing that methylated *trans*-silencing loci were present at chromosome ends (A.J.M. Matzke et al., 1994; Park et al., 1996), a distinction must be made between cytogenetic and molecular resolution. The *H9* and *H83* loci were obviously not integrated into telomeres, as indicated by the absence of recognizable telomeric repeats in the flanking plant DNA. In contrast, a preliminary analysis of plant DNA in a cosmid clone comprising the *H2 trans*-silencing locus (Park et al., 1996) has revealed the presence of a high-copy-number tandemly repeated sequence (J. Jakowitsch, I. Papp, and A.J.M. Matzke, unpublished results). Such tandem ar-

rays are found predominantly in subtelomeric and centromeric blocks of heterochromatin.

The unstably expressed loci *H59* and *H11* occupied intercalary and paracentromeric sites, respectively, and thus were remote from the gene-rich domains close to telomeres. Mosaically expressed transgenes have been localized close to centromeres in mice (Dobie et al., 1996, 1997) and in fission yeast (Allshire et al., 1994). These results are consistent with the mosaic expression of the paracentromeric *H11* locus (Neuhuber et al., 1994).

Taking into account two other transgene loci that we have localized previously by FISH/GISH on tobacco chromosomes (summarized in Figure 10), it is clear that both the S and T subgenomes are equally susceptible targets for T-DNA integration. At least some ability to *trans*-silence across constituent subgenomes is suggested by the observation that the *trans*-silencing *271* locus on chromosome T3 further weakens the activity of the unstable *H59* insert on chromosome S11/t2.

### Flanking Plant DNA

As demonstrated by a GUS-positive phenotype of H9 and H83 plants, the left flanking plant DNA of the respective T-DNA inserts contained transcriptional enhancers; in contrast, the H59 plant was GUS-negative, indicating the lack of enhancer elements in left flanking plant DNA at this locus. T-DNA often integrates adjacent to enhancer elements—as evidenced by the fact that transcriptional fusions between promoterless marker genes and plant promoters occur with a frequency of 30%—or in AT-rich domains (reviewed in Tinland and Hohn, 1995). Both of the stably expressed *H9* and *H83* inserts were flanked on the left by AT-rich regions. In the case of *H9*, this included two runs of the dinucleotide AT. Sequences rich in AT have been shown previously to activate reporter genes from a 35Spro (Bustos et al., 1989). The GUS-negative phenotype of the unstably expressed *H59* locus, despite the presence of an adjacent $(AAT)_{12}$ microsatellite, might have resulted from the presence of a second microsatellite sequence, $(AAAG)_5$, because GA/CT-rich sequences in Drosophila have been associated with heterochromatic regions (Raff et al., 1994).

Significantly longer (0.5- to 1-kb) AT-rich regions, which behaved as MARs in vitro, were found in left flanking DNA of both of the stably expressed loci, *H9* and *H83*. Although the *H83* MAR was directly adjacent to the T-DNA, the *H9* MAR was separated from the T-DNA left border by ~2.5 kb, which comprised, however, two AT microsatellites. In only one other study (from petunia) has a T-DNA been found integrated adjacent to a genuine endogenous MAR (Dietz et al., 1994). Although no published data are yet available for the number or frequency of MARs in the tobacco genome, it is noteworthy that the proximity of MARs correlated completely with the two stably expressed loci. Despite their possible role in promoting the stable expression of the individual *H* loci, the adjacent MARs did not completely protect *H9* and *H83* from some degree of *trans*-silencing by the *271* locus. We cannot rule out the possibility that MARs would have been found relatively close to the unstably expressed *H11* and *H59* loci if more flanking plant DNA had been recovered and analyzed, although the GUS-negative phenotype of H59 argues against this possibility.

Much of the flanking plant DNA that we recovered appeared—from the DNA gel blot analysis—to contain moderately repetitive dispersed sequences, which presumably consist of retrotransposons or their degenerate remains (Bennetzen, 1996). In only one case (*H83R*), however, were we able to obtain sequence data confirming the presence of a retroelement remnant. Sequences related to this particular retroelement remnant, which was not highly homologous to the best characterized tobacco retrotransposon *Tnt1* (Grandbastien, 1992), have also been found in the flanking plant DNA of another T-DNA insert analyzed previously (this study; Papp et al., 1996), in the 5′ flanking region of a tobacco defense-related pseudogene (Froissard et al. 1994), and as part of the moderately repetitive sequence R8.3, reported to be present in the *N. tomentosiformis* component of the tobacco genome (Kuhrová et al., 1991; A.J.M. Matzke, unpublished results). This retroelement remnant appears to be relatively abundant in the tobacco genome; a previous copy number estimate is ~1000 copies per haploid tobacco genome (Kuhrová et al., 1991). The apparent enrichment of this element in the T subgenome could have been due to either preferential amplification of the ancestral transposon in *N. tomentosiformis* or more rapid sequence divergence in *N. sylvestris*. A third possibility is that a T-specific retroelement infiltrated the S subgenome after polyploid formation but did not amplify there to the same extent as in the T subgenome. Although molecular evidence for local S–T interchange has been reported previously for a tobacco glucanase gene (Sperisen et al., 1991), the different base compositions of the two diploid ancestral genomes of tobacco indicate that extensive recombination between the two genomes has not occurred (Matassi et al., 1991). Although more data are needed, it is conceivable that the stability of transgene expression in allopolyploids can be differentially influenced by the proximity of moderately repetitive sequences that are preferentially enriched in one subgenome.

The pervasiveness of diverse repeats in the tobacco DNA flanking the *H* transgene inserts is reminiscent of the genomic environment of a maize alcohol dehydrogenase gene, which is embedded in a mosaic of highly and moderately repetitive sequences with MARs at the boundary of the repeated regions (Avramova et al., 1995; SanMiguel et al., 1996). Additional work is required to determine how transgenes that are integrated into such complex genomes are influenced by different types of repeats.

Several T-DNA inserts have been found in highly repetitive tobacco DNA specific to the T subgenome (Suter-Crazzolara et al., 1995), although no cytogenetic data or information on the stability of T-DNA expression were presented. The plant

DNA sequences at the integration sites of three T-DNA inserts in petunia have also been studied: one inactive T-DNA was present in a highly repetitive region, whereas two expressed loci were integrated into low-copy AT-rich regions (Pröls and Meyer, 1992). The chromosomal locations of these T-DNAs were not reported.

## Strategies for Avoiding Transgene Silencing

Intact copies of T-DNAs that are not associated with foreign DNA sequences appear to be favorable for preventing the type of silencing that we have studied (transcriptional inactivation associated with increased methylation). Sequences recognized as foreign might include prokaryotic vector DNA that joins T-DNA and plant DNA and, in allopolyploids, "alien" repeats that have infiltrated a different subgenome. Flanking transgene constructs with MAR sequences probably improves chances for stable expression (Mlynárová et al., 1996; Spiker and Thompson, 1996), but they will not necessarily protect fully from *trans*-silencing. Although it is not yet possible to target genes to subtelomeric or gene-rich compartments, including an enhancerless *gus* gene at a T-DNA border would allow screening for integration into actively transcribed regions. Intercalary and centromeric chromosomal locations do not appear optimal for obtaining stable expression. It should also be emphasized that silencing resulting from post-transcriptional RNA turnover induced by high levels of transcription (Elmayan and Vaucheret, 1996) presumably requires different strategies for avoidance.

Apart from the issue of minimizing unwanted silencing, transgene technology provides a powerful approach for probing the influence of genome organization and repetitive DNA on gene expression and for studying the factors that alter gene activity in the course of genome evolution, such as after polyploid formation. This is also illustrated by a recent report describing enhanced epigenetic silencing of a transgene locus in newly formed triploid Arabidopsis (Mittelsten Scheid et al., 1996).

## METHODS

### Recovery of T-DNA Inserts and Flanking DNA

For rescue cloning, 2 μg of genomic DNA isolated from each plant line was digested with an appropriate restriction enzyme, ligated, and used to transform electroporation-competent cells (Epicurian Coli XL1-Blue; Stratagene) by using an *Escherichia coli* Pulsar unit (Bio-Rad), following the manufacturer's instructions. The restriction enzymes that allowed recovery of the following left and right flanking DNAs from each line are as follows: *H9L*, EcoRI; *H9L+R*, BlnI; *H11*-a, BlnI-SpeI; *H59L*, EcoRI; *H59R*, MluI; *H83L*, ClaI; and *H83R*, BlnI.

In addition to the BlnI rescue clone obtained from line H11, a λ clone containing plant DNA (region *H11*-b) and surrounding T-DNA fragments (Figure 2) was isolated from a λ phage library, which was constructed using a Lambda FIX II/XhoI partial fill-in vector kit and a Gigapack III gold packaging extract (both from Stratagene).

### DNA Gel Blots

DNA gel blot analysis was performed as described by Papp et al. (1996). As probes, plasmids containing the appropriate left or right flanking DNA recovered in the rescue or λ clones (in the case of *H11*-b) were labeled with phosphorus-32 by using an Amersham Multiprime kit (model RPN 1600Z). Hybridization and washing were performed under moderately stringent conditions (3 × SSC [1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate] at 65°C).

### Nucleotide Sequence Analysis

Nucleotide sequences were determined using either a Sequenase kit (version 2.0; United States Biochemical Corp.) or a ThermoSequenase cycle sequencing kit (Amersham) using infrared-labeled primers obtained from MWG-Biotech (Ebersberg, Germany) and run on a LI-COR (Omaha, NE) DNA sequencing system.

Programs from the GCG and EGCG program packages of the Genetics Computer Group (Madison, WI) were used for pairwise and multiple sequence alignments (GAP, PILEUP, and PRETTY BOX). Database searches were performed by the BLAST algorithm (Altschul et al., 1990).

### Tobacco Nuclear Matrix Isolation and Binding Assay

Nuclear matrices were isolated from tobacco NT-1 cells, and matrix binding assays were performed as described by Hall and Spiker (1994). The DNAs tested consisted of the entire *H9* BlnI rescue clone digested with EcoRV and MunI and the entire *H83* BlnI rescue clone digested with XbaI and HindIII. These enzyme combinations were chosen so that the resulting 5′ ends could be labeled efficiently with phosphorus-32. The *H9* and *H83* rescue clones contained right and left flanking plant DNA in addition to the entire T-DNA region. These rescue clones were analyzed because sequencing revealed that the plant DNA flanking the T-DNA inserts contained extended AT-rich regions (Figures 4 and 7). (Such regions were not found in the plant DNA that was recovered from the unstably expressing lines; therefore, this DNA was not tested for matrix binding ability.) Any fragments from the *H9* and *H83* rescue clones that did not partition more completely in the pellet fractions served as internal controls.

### Cytogenetics

Combined fluorescent and genomic in situ hybridization (FISH/GISH) was performed exactly as described previously (Moscone et al., 1996). For probes to detect the transgene inserts (FISH), a binary vector containing the H construct was used; for the GISH analysis, total genomic DNA isolated from either *Nicotiana sylvestris* or *N. otophora* was used to identify the S or T subgenome, respectively. *N. otophora* DNA was used for the T subgenome because it produced a more uniform pattern of fluorescence than did *N. tomentosiformis* DNA.

## ACKNOWLEDGMENTS

We thank Barbara Hohn for helpful discussions. This work was supported by grants to A.J.M.M. and M.M. from the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (Grant Nos. S6006BIO and P10021-MOB), the Austrian National Bank (Grant No. J-5221), and the European Community Human Capital and Mobility Network, Contract No. CHRX-CT94-0503. E.A.M. thanks the Consejo Nacional de Investigaciones Científicas y Técnicas (Argentina) for permission to take a study leave.

Received February 20, 1997; accepted May 1, 1997.

## REFERENCES

**Allshire, R.C., Javerzat, J.-P. Redhead, N.J., and Cranston, G.** (1994). Position effect variegation at fission yeast centromeres. Cell **76,** 157–169.

**Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J.** (1990). Basic local alignment search tool. J. Mol. Biol. **215,** 403–410.

**Assaad, F.F., Tucker, K.L., and Signer, E.R.** (1993). Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. Plant Mol. Biol. **22,** 1067–1085.

**Avramova, Z., SanMiguel, P., Georgieva, E., and Bennetzen, J.L.** (1995). Matrix attachment regions and transcribed sequences within a long chromosomal continuum containing maize *Adh*1. Plant Cell **7,** 1667–1680.

**Bennetzen, J.L.** (1996). The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol. **4,** 347–353.

**Bernardi, G.** (1995). The human genome: Organization and evolutionary history. Annu. Rev. Genet. **29,** 445–476.

**Bustos, M.M., Guiltinan, M.J., Jordano, J., Begum, D., Kalkan, F.A., and Hall, T.C.** (1989). Regulation of β-glucuronidase expression in transgenic tobacco plants by an A/T-rich, *cis*-acting sequence found upstream of a French bean β-phaseolin gene. Plant Cell **1,** 839–853.

**Cluster, P.D., O'Dell, M., Metzlaff, M., and Flavell, R.B.** (1996). Details of T-DNA structural organization from a transgenic petunia population exhibiting co-suppression. Plant Mol. Biol. **32,** 1197–1203.

**Dietz, A., Kay, V., Schlake, T., Landsmann, J., and Bode, J.** (1994). A plant scaffold attached region detected close to a T-DNA integration site is active in mammalian cells. Nucleic Acids Res. **22,** 2744–2751.

**Dobie, K.W., Lee, M., Fantes, J.A., Graham, E., Clark, A.J., Springbett, A., Lathe, R., and McClenaghan, M.** (1996). Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. Proc. Natl. Acad. Sci. USA **93,** 6659–6664.

**Dobie, K.W., Mehtali, M., McClenaghan, M., and Lathe, R.** (1997). Variegated gene expression in mice. Trends Genet. **13,** 127–130.

**Elmayan, T., and Vaucheret, H.** (1996). Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally. Plant J. **9,** 787–797.

**Froissard, D., Gough, C., Czernic, P., Schneider, M., Toppan, A., Roby, D., and Marco, Y.** (1994). Structural organization of *str* 246C and *str* 3246N, plant defense–related genes from *Nicotiana tabacum*. Plant Mol. Biol. **26,** 515–521.

**Gill, K.S., Gill, B.S., and Endo, T.R.** (1993). A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. Chromosoma **102,** 374–381.

**Grandbastien, M.-A.** (1992). Retroelements in higher plants. Trends Genet. **8,** 103–108.

**Hall, G.E., and Spiker, S.** (1994). Isolation and characterization of nuclear scaffolds. In Plant Molecular Biology Manual, S.B. Gelvin and R.A. Schilperoort, eds (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 1–12.

**Hobbs, S.L.A., Warkentin, T.D., and DeLong, C.M.O.** (1993). Transgene copy number can be positively or negatively associated with transgene expression. Plant Mol. Biol. **21,** 17–26.

**Jorgensen, R.A., Cluster, P.D., English, J., Que, Q., and Napoli, C.** (1996). Chalcone synthase cosuppression phenotypes in petunia flowers: Comparison of sense vs. antisense constructs and single-copy vs. complex T-DNA sequences. Plant Mol. Biol. **31,** 957–973.

**Kilby, N.J., Leyser, H.M.O., and Furner, I.J.** (1992). Promoter methylation and progressive transgene inactivation in *Arabidopsis*. Plant Mol. Biol. **20,** 103–112.

**Koncz, C., Németh, K., Rédei, G., and Schell, J.** (1994). Homology recognition during T-DNA integration into the plant genome. In Homologous Recombination and Gene Silencing in Plants, J. Paszkowski, ed (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 167–189.

**Kuhrová, V., Bezdek, M., Vyskot, B., Koukalová, B., and Fajkus, J.** (1991). Isolation and characterization of two middle repetitive DNA sequences of nuclear tobacco genome. Theor. Appl. Genet. **81,** 740–744.

**Linn, F., Heidmann, I., Saedler, H., and Meyer, P.** (1990). Epigenetic changes in the expression of the maize A1 gene in *Petunia hybrida*: Role of numbers of integrated gene copies and state of methylation. Mol. Gen. Genet. **222,** 329–336.

**Martineau, B., Voelker, T.A., and Sanders, R.A.** (1994). On defining T-DNA. Plant Cell **6,** 1032–1033.

**Matassi, G., Melis, R., Macaya, G., and Bernardi, G.** (1991). Compositional bimodality of the nuclear genome of tobacco. Nucleic Acids Res. **19,** 5561–5567.

**Matsumoto, S., Ito, Y., Hosoi, T., Takahashi, Y., and Machida, Y.** (1990). Integration of *Agrobacterium* T-DNA into a tobacco chromosome: Possible involvement of DNA homology between T-DNA and plant DNA. Mol. Gen. Genet. **224,** 309–316.

**Matzke, A.J.M., Neuhuber, F., Park, Y.-D., Ambros, P.F., and Matzke, M.A.** (1994). Homology-dependent gene silencing in transgenic plants: Epistatic silencing loci contain multiple copies of methylated transgenes. Mol. Gen. Genet. **244,** 219–229.

**Matzke, M.A., and Matzke, A.J.M.** (1996). Stable epigenetic states in differentiated plant cells: Implications for somaclonal variation and gene silencing in transgenic plants. In Epigenetic Mechanisms of Gene Regulation, V.E.A. Russo, R.A. Martienssen, and A.D. Riggs, eds (Cold Spring Harbor, NY; Cold Spring Harbor Laboratory Press), pp. 377–392.

**Matzke, M.A., Matzke, A.J.M., and Mittelsten Scheid, O.** (1994). Inactivation of repeated genes—DNA–DNA interaction? In Homologous Recombination and Gene Silencing in Plants, J. Paszkowski, ed (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 271–307.

**McElroy, D., and Brettell, R.I.S.** (1994). Foreign gene expression in transgenic cereals. Trends Biotechnol. **12,** 62–68.

**Meyer, P., and Saedler, H.** (1996). Homology-dependent gene silencing in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47,** 23–48.

**Mittelsten Scheid, O., Jakovleva, L., Afsar, K., Maluszynska, J., and Paszkowski, J.** (1996). A change of ploidy can modify epigenetic silencing. Proc. Natl. Acad. Sci. USA **93,** 7114–7119.

**Mlynárová, L., Keizer, L.C.P., Stiekema, W.J., and Nap, J.-P.** (1996). Approaching the lower limits of transgene variability. Plant Cell **8,** 1589–1599.

**Moscone, E.A., Matzke, M.A., and Matzke, A.J.M.** (1996). The use of combined FISH/GISH in conjunction with DAPI counterstaining to identify chromosomes containing transgene inserts in amphidiploid tobacco. Chromosoma **105,** 231–236.

**Neuhuber, F., Park, Y.-D., Matzke, A.J.M., and Matzke, M.A.** (1994). Susceptibility of transgene loci to homology-dependent gene silencing. Mol. Gen. Genet. **244,** 230–241.

**Ohba, T., Yoshioka, Y., Machida, C., and Machida, Y.** (1995). DNA rearrangement associated with the integration of T-DNA in tobacco: An example of multiple duplications of DNA around the integration target. Plant J. **7,** 157–164.

**Okamuro, J.K., and Goldberg, R.B.** (1985). Tobacco single-copy DNA is highly homologous to sequences present in the genomes of its diploid progenitors. Mol. Gen. Genet. **198,** 290–298.

**Papp, I., Iglesias, V.A., Moscone, E.A., Michalowski, S., Spiker, S., Park, Y.-D., Matzke, M.A., and Matzke, A.J.M.** (1996). Structural instability of a transgene locus in tobacco is associated with aneuploidy. Plant J. **10,** 469–478.

**Park, Y.-D., Papp, I., Moscone, E.A., Iglesias, V.A., Vaucheret, H., Matzke, A.J.M., and Matzke, M.A.** (1996). Gene silencing mediated by promoter homology occurs at the level of transcription and results in meiotically heritable alterations in methylation and gene activity. Plant J. **9,** 183–194.

**Parokonny, A.S., and Kenton, A.Y.** (1995). Comparative physical mapping and evolution of the *Nicotiana tabacum* L. karyotype. In Kew Chromosome Conference IV, P.E. Brandham and M.D. Bennett, eds (Kew, UK: Royal Botanic Gardens), pp. 301–320.

**Pawloski, W.P., and Somers, D.A.** (1996). Transgene inheritance in plants genetically engineered by microprojectile bombardment. Mol. Biotechnol. **6,** 17–29.

**Peach, C., and Velten, J.** (1991). Transgene expression variability (position effect) of *CAT* and *GUS* reporter genes driven by linked divergent T-DNA promoters. Plant Mol. Biol. **17,** 49–60.

**Pröls, F., and Meyer, P.** (1992). The methylation patterns of chromosomal integration regions influence gene activity of transferred DNA in *Petunia hybrida.* Plant J. **2,** 465–475.

**Raff, J.W., Kellum, R., and Alberts, B.** (1994). The Drosophila GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. EMBO J. **13,** 5977–5983.

**Ramanathan, V., and Veluthambi, K.** (1995). Transfer of non–T-DNA portions of the *Agrobacterium tumefaciens* Ti plasmid pTiA6 from the left terminus of $T_L$-DNA. Plant Mol. Biol. **28,** 1149–1154.

**Saccone, S., De Sario, A., Della Valle, G., and Bernardi, G.** (1992). The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. Proc. Natl. Acad. Sci. USA **89,** 4913–4917.

**SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L.** (1996). Nested retrotransposons in the intergenomic regions of the maize genome. Science **274,** 765–768.

**Sperisen, C., Ryals, J., and Meins, F.** (1991). Comparison of cloned genes provides evidence for intergenomic exchange of DNA in the evolution of a tobacco glucan endo-1,3-β-glucosidase gene family. Proc. Natl. Acad. Sci. USA **88,** 1820–1824.

**Spiker, S., and Thompson, W.F.** (1996). Nuclear matrix attachment regions and transgene expression in plants. Plant Physiol. **110,** 15–21.

**Srivastava, V., Vasil, V., and Vasil, I.K.** (1996). Molecular characterization of the fate of transgenes in transformed wheat (*Triticum aestivum* L.). Theor. Appl. Genet. **92,** 1031–1037.

**Stam, M., de Bruin, R., Kenter, S., Van der Hoorn, R.A.L., Van Blokland, R., Mol, J.N.M., and Kooter, J.M.** (1997). Post-transcriptional silencing of chalcone synthase in *Petunia* by inverted transgene repeats. Plant J. **12,** 63–82.

**Suter-Crazzolara, C., Brzobohaty, B., Gazdova, B., Schell, J., and Reiss, B.** (1995). T-DNA integrations in a new family of repetitive elements of *Nicotiana tabacum.* J. Mol. Evol. **41,** 498–504.

**ten Hoopen, R., Robbins, T.P., Fransz, P.F., Montijn, B.M., Oud, O., Gerats, A.G.M., and Nanniga, N.** (1996). Localization of T-DNA insertions in petunia by fluorescence in situ hybridization: Physical evidence for suppression of recombination. Plant Cell **8,** 823–830.

**Tinland, B.** (1996). The integration of T-DNA into plant genomes. Trends Plant Sci. **1,** 178–183.

**Tinland, B., and Hohn, B.** (1995). Recombination between prokaryotic and eukaroytic DNA: Integration of *Agrobacterium tumefaciens* T-DNA into the plant genome. In Genetic Engineering, Vol. 17, J.K. Setlow, ed (New York: Plenum Press), pp. 209–229.

**Van Blokland, R., Van der Geest, N., Mol, J.N.M., and Kooter, J.M.** (1994). Transgene-mediated suppression of chalcone synthase expression in *Petunia hybrida* results from an increase in RNA turnover. Plant J. **6,** 861–877.

**Van der Graaff, E., den Dulk-Ras, A., and Hooykaas, P.J.J.** (1996). Deviating T-DNA transfer from *Agrobacterium tumefaciens* to plants. Plant Mol. Biol. **31,** 677–681.

**Vaucheret, H.** (1993). Identification of a general silencer for 19S and 35S promoters in a transgenic tobacco plant: 90 bp of homology in the promoter sequence are sufficient for trans-inactivation. C.R. Acad. Sci. Ser. III Sci. Vie **316,** 1471–1483.

**Wang, J., Lewis, M.E., Whallon, J.H., and Sink, K.C.** (1995). Chromosomal mapping of T-DNA inserts in transgenic *Petunia* by *in situ* hybridization. Transgenic Res. **4,** 241–246.

## NOTE ADDED IN PROOF

Kononov et al. have recently reported frequent integration of T-DNA binary vector "backbone" sequences into the tobacco genome (**Kononov, M.E., Bassuner, B., and Gelvin, S.B.** [1997]. Integration of T-DNA binary vector "backbone" sequences into the tobacco genome: Evidence for multiple complex patterns of integration. Plant J. **11,** 945–957).