Intrachromosomal recombination between attP regions as a tool to remove selectable marker genes from tobacco transgenes

Elena Zubko[1], Charles Scutt[1,2], and Peter Meyer[1]*

[1]*Leeds Institute for Plant Biotechnology and Agriculture (LIBA), Faculty of Biological Sciences, The University of Leeds, Leeds LS2 9JT, United Kingdom.*
[2]*Present address: Réproduction et Développement des Plantes, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon cedex 07, France.*
**Corresponding author (p.meyer@Leeds.ac.uk).*

Received 13 October 1999; accepted 10 February 2000

Recombinant genes conferring resistance to antibiotics or herbicides are widely used as selectable markers in plant transformation. Once transgenic material has been selected, the marker gene is dispensable. We report a novel strategy to remove undesirable parts of a transgene after integration into the tobacco genome. This approach is based on the transfer of a vector containing a *NPTII* gene flanked by two 352 bp attachment P (attP) regions of bacteriophage λ, and the identification of somatic tissue with deletion events following intrachromosomal recombination between the attP regions. This system was used to delete a 5.9 kb region from a recombinant vector that had been inserted into two different genomic regions. As the attP system does not require the expression of helper proteins to induce deletion events, or a genetic segregation step to remove recombinase genes, it should provide a useful tool to remove undesirable transgene regions, especially in vegetatively propagated species.

Keywords: resistance marker, homologous recombination, transgenic plants

The removal of resistance marker genes from plant transgenes is highly desirable for a number of reasons. Cross-pollination between related species can lead to a transfer of resistance traits into weeds[1,2], jeopardizing the long-term use of transgenic crops and causing potential ecological problems. Consumer groups express concern about a widespread distribution of resistance markers in food products, referring to a theoretical risk of a horizontal transfer of transgenes into gut bacteria. Moreover, because there is only a limited number of selection marker genes that can be used for plant transformation, the combination of multiple transgenic traits through crosses among different transgenic lines will frequently produce plants that contain multiple copies of the same selection marker linked to different effector genes. The presence of multiple homologous sequences in plants enhances the likelihood for homology-dependent gene silencing[3], which could severely limit the reliable long-term use of transgenic crops.

A number of systems have so far been developed to ensure the removal of selectable marker genes. Co-transformation of two different constructs can result in transgenic lines that have integrated both transgenes[4], but the applicability of this approach depends on the efficiency at which both transgenes insert into different genomic regions, as this is a requirement to separate them in genetic crosses. As an alternative to co-transformation, several transposable element systems and site-specific recombination systems have been employed for marker removal[5]. These systems require the expression of a transposase or recombinase that mediates the deletion of regions bracketed between recombination or transposase target sequences, and the subsequent removal of the helper gene by genetic segregation, which makes these systems relatively time consuming, as well as impractical for species that are mainly propagated vegetatively. Moreover, deleted fragments can reinsert into other genomic positions, and recombinase or transposase proteins might cause undesirable secondary effects.

A less complicated approach to induce DNA deletions is based on intrachromosomal homologous recombination (ICR) between two homologous sequences. Although ICR can be enhanced by stimulation of repair systems, ICR frequencies are too low for an efficient application of this system to produce deletions of transgene regions. In tobacco, for example, on average less than 10 ICR events are detectable among all cells of a six-week-old tobacco plant[6]. Such low ICR frequencies might be increased if recombination substrates could be found that provided a more efficient target for the recombination machinery. We tested an ICR strategy based on recombination of the attP region of bacteriophage λ to generate deletions, and found that such a construct allows the identification of ICR products among a relatively small number of transformants, providing a feasible procedure to remove undesirable transgene regions.

Results

Integration of bacteriophage λ into the *Escherichia coli* genome occurs by recombination between the phage attachment (attP) and the bacterial attachment (attB) regions. The bacterial integration host factor (IHF), and the virally encoded integrase (*Int*) protein are required for attP integration, whereas excision requires in addition the viral excisase (Xis) protein. The three recombination proteins bind to defined DNA regions within a 250 bp attP fragment[7]. The IHF fulfills an accessory function, bending the attP region[8] and possibly assisting *Int* in organizing attP into a nucleosome-like structure

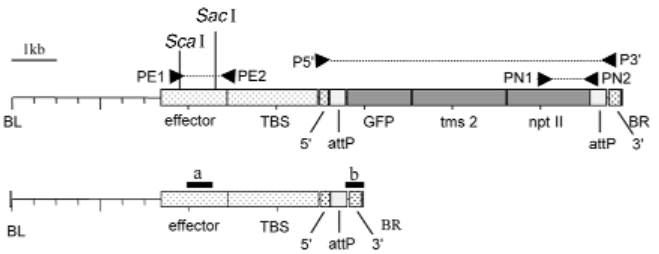

Figure 1 T-DNA region of pattP-ICR. Arrows indicate the regions amplified by the primers specific for detection of the effector gene (PE1 and PE2), the *NPTII* gene (PN1 and PN2), and the attP cassette (P5' and P3'). ICR between the two attP regions will generate a 5.9 kb deletion with only one attP region remaining and the region between the attP regions being removed. Black lines indicate probes used in Southern blot hybridization (Fig. 4).

Figure 2 Selection of marker-free transgenic tobacco plants. (A) Lines 1 and 2 develop green and white shoots on Kan-containing medium. White tissue, which has potentially lost the *NPTII* marker, was further tested for activity of the *tms-2* gene. (B) On NAM-containing medium, shoots with *tms2* activity produce abundant calli instead of roots (left), whereas shoots regenerated from white tissue that have lost the *NPTII* gene produce normal roots (right).

that is required for efficient synapsis with attB[9]. The actual strand exchange, which occurs within a 7 bp homology between the core regions of attP and attB, is mediated by the Int protein that has a topoisomerase I activity. A tyrosine residue of Int covalently binds to the 3′ end of the DNA, resembling a mammalian topoisomerase I type in contrast to *E. coli* topoisomerases I that binds to the 5′ end of the DNA[10].

We designed a plant transformation vector, pattP-ICR, containing two 352 bp attP regions that flank an *NPTII* resistance marker, a GFP gene, and a *tms2* gene (Fig. 1). Next to the left attP site, we positioned the transformation booster sequence (TBS), which enhances homologous and illegitimate recombination[11], and an oryzacystatin-I gene[12] that served as an example for an "effector" gene that would finally be transferred into the genome by means of the attP system. Intrachromosomal homologous recombination between the two attP regions should delete the 5.9 kb region between the attP regions and produce a transgene that retains the effector gene and the TBS sequence. The construct was introduced into tobacco by leaf explant transformation, and resistant calli were selected on kanamycin (Kan) medium. Two months after transformation, 11 Kan-resistant calli, 0.5 cm in diameter, were transferred to Kan-free medium. When calli had grown to 5–6 cm in diameter, they were cut into smaller parts that were transferred onto Kan-containing shoot regeneration medium. After three to five months, when multiple shoots had developed on all 11 calli, we detected two clones that produced a mixture of green and white shoots. These two clones, lines 1 and 2 (Fig. 2), were further studied to evaluate whether white shoots contained tissue that had lost the *NPTII* resistance marker and the *tms2* gene.

White leaves of both lines were placed on Kan-free regeneration medium, and regenerating shoots were placed on medium supplemented by naphthalene acetamide (NAM). As *tms2* gene activity converts NAM into the auxin NAA[13], plants expressing the *tms2* gene produce high auxin levels that prevent root development and induce callus production instead. Tissue that had deleted the region between the two attP fragments should have also lost the *tms2* gene and can therefore be identified by its ability to form roots on NAM-containing medium. Eleven of 20 shoots tested for line 1 produced roots, and 12 out of 32 shoots tested for line 2 did so, demonstrating the chimeric nature of the white shoots. The 23 root-producing plantlets were further analyzed as potential candidates that had lost both the *NPTII* gene and the *tms2* gene. Polymerase chain reaction (PCR) analysis showed that all 23 plantlets had indeed lost the *NPTII* gene and the *tms2* gene. Among the 11 line 1 derivatives, 2 had retained the effector gene, and 1 of the 12 line 2 derivatives had done so, whereas all other lines had lost or rearranged the effector gene as well.

One representative plant each from line 1 and line 2 that had retained the effector gene was further characterized to analyze the precise degree of the deletion events. As a control, we used Kan-resistant shoots of lines 1 and 2. Analysis by PCR with primers specific for the effector gene, the *NPTII* gene, and the region flanking the attP dimer cassette showed that the effector gene had been retained, while the *NPTII* gene had been lost and the region between the two attP fragments had been reduced by about 6 kb, as expected if the two attP sites had recombined (Fig. 3). Both clones produced identical PCR patterns, indicating that they had both generated the same deletion product. Sequencing of one of the PCR fragments confirmed that precisely one attP region had been maintained between the 5′ and 3′ region flanking the attP cassette (Fig. 3), a result that is in accordance with the expectation that ICR between the two attP fragments has deleted the 5.9 kb region between them. A Southern blot analysis of green and white tissue of line 1 confirmed the PCR data (Fig. 4). For the Kan sensitive lines 1 and 2, the PCR fragment that was specific for the effector gene (Fig. 3A) was sequenced. Both fragments were identical and corresponded to the sequence on the transferred plasmid. It therefore appears that ICR events between the two attP regions do not alter the effector gene. The PCR and Southern blot data show no indication for the presence of unrecombined tissue in the lines that were regenerated from the chimeric tissue, suggesting that the material selected by the described procedure is homogeneous.

Discussion

The fact that most plantlets that have lost the *NPTII*/*tms2* region also have lost transgene regions outside the attP cassette, shows that ICR is not always associated with precise homologous recombination between the two attP regions but that it can generate larger deletions probably as a result of illegitimate recombination. However, ICR events can clearly be identified where precise intrachromosomal recombination had occurred between the two attP regions. As the same deletion event can be found in two different transformants, it is unlikely that efficient recombination between attP regions only occurs within transgenes that are integrated at specific genomic integration regions. Our data show that, starting with a relatively small number of calli, homogeneous material with precise deletion events

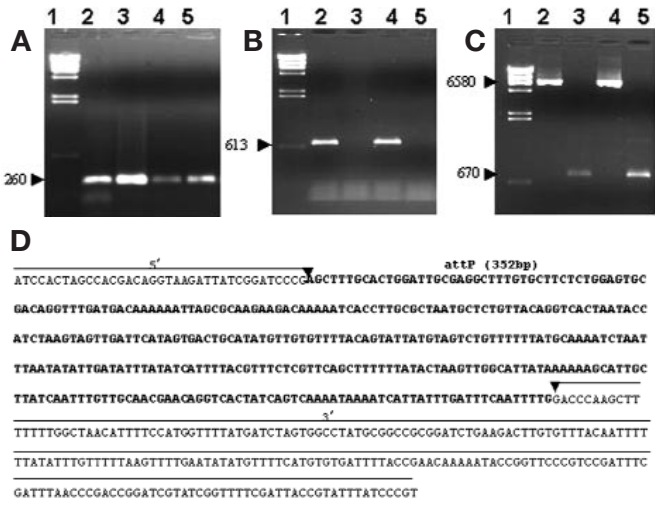

Figure 3. PCR analysis of the Kan-resistant plants derived from line 1 (lane 2) and line 2 (lane 4), and Kan-sensitive plants derived from white tissue of line 1 (lane 3) and line 2 (lane 5). Lane 1 contains a *Hin*dIII-digested λ-DNA as a size marker. Sizes of PCR fragments are indicated in base pairs. (A) PCR with primers PE1 and PE2 shows that all four lines contain the effector gene. (B) PCR with primers PN1 and PN2 shows that the Kan-sensitive plants have lost the *NPTII* gene. (C) PCR with primers P5' and P3' show that the two Kan-resistant lines contain the complete attP cassette, whereas in both Kan-sensitive plants about 6 kb have been deleted from the attP cassette. (D) Sequence of the PCR product shown in (C), lane 3, confirms that recombination leaves precisely one attP region bracketed by the 5′ and 3′ regions.

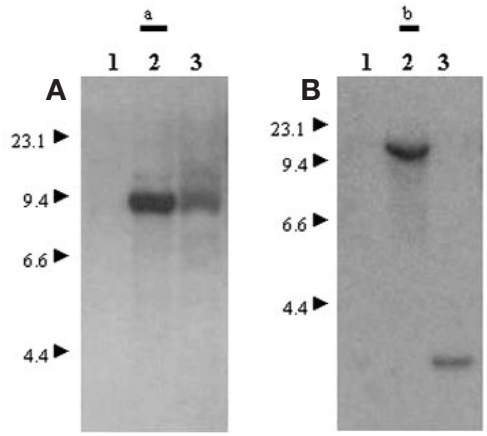

Figure 4. Southern blot analysis of DNA isolated from Kan-resistant (lane 2) and Kan-sensitive tissue of line 1 (lane 3). Lane 1 contains SR1 DNA as a negative control. (A) Genomic DNA digested with *Sac*I was probed with region "a" (see Fig.1) to label the left junction fragment of the integrated T-DNA. (B) Genomic DNA digested with *Sca*I was probed with region "b" (see Fig. 1) to label the right junction fragment of the integrated T-DNA. In Kan-sensitive tissue, the left end fragment remains unaltered (A), while the right end fragment is shortened by about 6 kb because of the deletion of the region between the two attP regions (B).

can be selected, and that, because of the availability of PCR technology, such events can easily be distinguished from undesirable deletion events that have lost more than the region between the two attP sites. As at least 35% of the shoots regenerated from white tissue represent deletion events, the use of the *tms2* gene for preselection of deletion events appears dispensable and can probably also be substituted by PCR.

Our data do not allow us to determine the precise ICR frequency. So far, nobody has used the loss of a selectable marker gene, like the Kan-resistance gene, for ICR studies. Instead, the visible *GUS* gene has been used[14], which is not suitable for selection of ICR events but has the advantage of monitoring ICR events more accurately, as it detects ICR within very small groups of cells within a plant. With a selectable marker, individual late ICR events in somatic tissue would not be detectable unless they happen very frequently, giving rise to branches with a mixed cell population of white ICR products and green tissue that has not undergone ICR. This implies that a comparison of ICR data between *GUS* and *Kan* constructs is very difficult, but in any case, a *GUS* reporter system would detect a much higher number of ICR events.

ICR experiments with *GUS* constructs show that one ICR event occurs in about 10^6–10^7 genomes if the construct contains 566 bp of duplicated homology, within which ICR can occur[6]. The efficiency of homologous recombination is directly correlated with the size of the homologous regions, and a reduction of homology for the *GUS* system to 456 bp results in a 2.5-fold reduction in ICR efficiency[14]. The attP homology in our construct is only 352 bp, which suggests that the expected ICR frequency for our vector is about one ICR event per 10^7 genomes at most. Considering that our system allows the detection of ICR events only if they occur very early, giving rise to a large clonal sectors of ICR tissue, or if they occur at a much higher frequency than the expected one event per 10^7 genomes, producing chimeric tissue with multiple ICR events, it appears that the attP construct shows a relatively high recombination frequency, at least for certain transformants. The ICR lines that were isolated, derived from calli that produced multiple white shoots representing independent deletion events. These shoots were chimeric, most likely also representing multiple deletion events within each shoot, with 44% of the cells of the chimeric tissue having lost the *NPTII* gene. 13% of these deletion derivatives were ICR events, whereas the majority represented larger deletions due to illegitimate recombination.

That the majority of deletion events were due to illegitimate recombination suggests that there has been no significant alteration in the known dominance of illegitimate recombination over homologous recombination[15]. It rather appears that both illegitimate and homologous recombination events occur with a relatively high efficiency at certain attP transgene loci, suggesting an improved accessibility of these transgene regions. A similar effect is also experienced for extrachromosomal recombination constructs that undergo recombination at much higher frequencies than recombination substrates that are integrated into the genome[16].

Out of 11 lines tested, 9 did not produce any detectable deletion events, whereas multiple deletion events were detectable in different parts of the other two lines, implying that the integration region of the individual transgenes influences the efficiency of recombination, as was suggested in other ICR studies[6,17]. The complexity of the transgene locus may also play a role in ICR efficiency. Considering that ICR can occur among any two homologous regions, transgene loci with multiple homologous vector fragments may be less suitable for ICR between the two attP regions, as ICR could occur among other homologous vector fragments.

It remains to be determined whether the attP vector provides transgenes with an improved accessibility to repair and recombination enzymes, and whether different genomic positions or transgene complexities interfere with such a hypothetical role of the attP vector. For practical purposes, however, the attP vector should provide a useful tool, especially for vegetatively propagated species, to select transformants that have lost the resistance markers or any other undesirable transgene regions from transgenic plants.

Experimental protocol

Plant transformation. The pattP-ICR construct was introduced into *Agrobacterium tumefaciens* strain GV3101 (pMP90RK) by conjugation as described[18].Leaf disk transformation of *Nicotiana tabacum* cv. Petit Havana SR1 was performed according to the method of Koncz and coworkers[19].

Vector construction. To construct the transformation vector pattP-ICR, two 352 bp regions of the attP region (located between position 27,492 and 27,844 of the phage λ genome) were inserted at both sides of the *NPTII* gene of pPCV002[20]. Into the attP cassette, we inserted the *tms-2* coding region and polyA region[13] and the GFP coding region[21] linked to the nopaline synthase polyA region. Both genes are transcribed by the dual 1'-2' promoter[22]. The attP cassette was embedded into a 0.6 kb Dissociation element that would allow removal of the complete cassette in the presence of an active Ac transposase. 5′ to the attP cassette a TBS fragment was inserted that enhances illegitimate and homologous recombination and an oryzacystatin-I gene as an example for an effector gene.

DNA preparation and analysis. After DNA mini-preparation[23], PCRs were performed in a reaction volume of 50 µl containing 10 mM Tris-HCl pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 0.1% (wt/vol) gelatin, 0.2 mM of each nucleotide, 25 pmol of each primer, and 1 U of *Taq* polymerase (Promega, Madison, WI). The PCR cycles used were, for detection of the effector gene, 4 min at 94°C, 30 cycles of 1 min at 94°C, 30 s at 55°C, 30 s at 72°C followed by 10 min at 72°C; and to amplify the sequence of *NPTII* gene, 4 min at 94°C, 30 cycles of 1 min at 94°C, 1 min at 60°C, 1 min at 72°C followed by 10 min at 72°C. For amplification of the attP cassette, a long-template PCR was performed using the Expand Long Template PCR System according to manufacturer's protocol (Boehringer Mannheim, Mannheim, Germany). The sequences of the primers used in PCR were (5′ to 3′):

PE1: TCA TCA GAC GGA GGA CCA GTT TTG G
PE2: ATC CAT GGT TTT TCC CAA ACT TTA G
PN1: CCA TGA TCA TGT CGA TTG AAC AAG ATG
PN2: CCA TTT TCC ACC ATG ATA TTC GGC AAG
P5': GAA TTC TAA TTC GGG ATG ACT GCA ATA TGG
P3': GGA TCC AAC GGG ATA TAC CGG TAA CGA AAA CG

For the Southern blot analysis, genomic DNA was isolated as described[24]. A 15 µg sample of genomic DNA was digested with *Sca*I and fractionated by electrophoresis in 0.7% agarose gel. After electrophoresis the DNA was blotted on Nybond N filter (Amersham Pharmacia Biotech, Buckinghamshire, UK), crosslinked by UV irradiation, and hybridized with a ^{32}P-labeled 0.3 kb *Eco*RI/*Sac*I DNA fragment of the effector gene at 65°C according to the method of Koes and colleagues[25].

Acknowledgment

We would like to thank Dr Peter Urwin for providing the effector gene. This work was supported by a Scottish Office grant, "Transgenic control of nematodes."

1. Dale, P.J. Spread of engineered genes to wild relatives. *Plant Physiol.* **100**, 13–15 (1992).
2. Gressel, J. Indiscriminate use of selectable markers—sowing wild oats? *TIBTECH* **10**, 382 (1992).
3. Matzke, M.A. & Matzke, A.J.M. Differential inactivation and methylation of a transgene in plants by two suppressor loci containing homologous sequences. *Plant Mol. Biol.* **16**, 821–830 (1991).
4. De Block, M. & Debrouwer, D. Two T-DNAs co-transformed into *Brassica napus* by a double *Agrobacterium* infection are mainly integrated at the same locus. *Theor. Appl. Genet.* **82**, 257–263 (1991).
5. Yoder, J.I. & Goldsbrough, A.P. Transformation systems for generating marker-free transgenic plants. *Bio/Technology* **12**, 263–267 (1994).
6. Puchta, H., Swoboda, P., Gal, S., Blot, M. & Hohn, B. Somatic intrachromosomal homologous recombination events in populations of plant siblings. *Plant Mol. Biol.* **28**, 281–292 (1995).
7. Gardner, J.F. & Nash, H.A. Role of *Escherichia coli* IHF protein in lambda site-specific recombination. *J. Mol. Biol.* **191**, 181–189 (1986).
8. Yang, C. & Nash, H.A., The interaction of *E. coli* IHF protein with its specific binding sites. *Cell* **57**, 869–880 (1989).
9. Craig, N.L. & Nash, H.A. The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. *Cell* **39**, 707–716 (1984).
10. Weisberg, R.A. & Landy, A. In *Lambda II* (eds Hendrix, R.W., Roberts, J.W., Stahl, F.W. & Weisberg, R.A.) 211–250 (Cold Spring Harbor Press, Cold Spring Harbor, NY; 1983).
11. Galliano, H., Müller, A.E., Lucht, J.M. & Meyer, P. The transformation booster sequence is a retrotransposon derivative that binds to the nuclear scaffold. *Mol. Gen. Genet.* **247**, 614–622 (1995).
12. Urwin, P.E., Atkinson, H.J., Waller, D.A. & McPherson, M.J. Engineered oryzacystatin-I expressed in transgenic hairy roots confers resistance to *Globodera pallida*. *Plant J.* **8**, 121–31 (1995).
13. Depicker, A.G., Jacobs, A.M. & Van Montagu, M.C. A negative selection scheme for tobacco protoplast-derived cells expressing the T-DNA gene 2. *Plant Cell Rep.* **7**, 63–66 (1988).
14. Puchta, H. & Hohn, B.A transient assay in plant cells reveals a positive correlation between extrachromosomal recombination rates and length of homologous overlap. *Nucleic Acids Res.* **19**, 2693–2700 (1991).
15. Roth, D. & Wilson, J. In *Genetic recombination* (eds Kucherlapati, R. & Smith, G.R.) 621–653 (American Society for Microbiology, Washington, DC; 1988).
16. Puchta, H. & Meyer, P. In *Homologous recombination and gene silencing in plants* (ed. Paszkowski, J.) 123–155 (Kluwer Academic Publishers, Dordrecht, The Netherlands; 1994).
17. Swoboda, P., Hohn, B. & Gal, S. Somatic homologous recombination *in planta*: the recombination frequency is dependent on the allelic state of the recombining sequences and may be influenced by genomic position effects. *Mol. Gen. Genet.* **237**, 33–40 (1993).
18. Koncz, C. et al. Specialized vectors for gene tagging and expression studies. *Plant Mol. Biol. Manual* **B2**, 1–22 (1994).
19. Horsch, R.B. et al. A simple and general method for transfering genes into plants. *Science* **227**, 1229–1231 (1985).
20. Konz, C. & Schell, J. The promoter of T_L-DNA gene 5 controls the tissue-specific expression of chimeric genes caried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* **204**, 383–396 (1986).
21. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
22. Velten, J. & Schell, J. Isolation of a dual plant promoter fragment from Ti plasmid of *Agrobacterium tumefaciens*. *EMBO J.* **3**, 2723–2730 (1984).
23. Dellaporta, S.J., Wood, J. & Hicks, J.B. A plant DNA minipreparation. *Plant Mol. Biol. Rep.* **1**, 19–21 (1983).
24. Van Blokland, R., Ross, S., Corrado, G., Scollan, C. & Meyer, P. Developmental abnormalities associated with deoxyadenosine methylation in transgenic tobacco. *Plant J.* **15**, 543–551 (1998).
25. Koes, R.E., Spelt, C.E., Mol, J.N.M. & Gerats, G.M. The chalcone synthase multigene family of *Petunia hybrida* (V30): sequence homology, chromosomal localization and evolutionary aspects. *Plant Mol. Biol.* **10**, 159–169 (1987).