The mechanism of gene targeting in Physcomitrella patens: homologous recombination, concatenation and multiple integration

Yasuko Kamisugi1, Katja Schlink2,4, Stefan A. Rensing2, Gabriele Schween2, Mark von Stackelberg2, Andrew C. Cuming1,*, Ralf Reski2 and David J. Cove1,3

1Centre for Plant Sciences, Faculty of Biological Sciences, Leeds University Leeds, LS2 9JT, UK, 2Plant Biotechnology, Faculty of Biology, University of Freiburg, Schaezenstrasse 1, D-79104 Freiburg, Germany, 3Department of Biology, Washington University in St Louis, St Louis, MO 63130-4899, USA and 4Forest Genetics, Department of Plant Sciences, Life Science Center Weihenstephan, Technische Universität München, Am Hochanger 13, D-85354 Freising, Germany

Received July 21, 2006; Revised September 25, 2006; Accepted October 8, 2006

ABSTRACT

The model bryophyte Physcomitrella patens exhibits high frequencies of gene targeting when transformed with DNA constructs containing sequences homologous with genomic loci. ‘Targeted gene replacement’ (TGR) resulting from homologous recombination (HR) between each end of a targeting construct and the targeted locus occurs when either single or multiple targeting vectors are delivered. In the latter instance simultaneous, multiple, independent integration of different transgenes occurs at the targeted loci. In both single gene and ‘batch’ transformations, DNA can also be found to undergo ‘targeted insertion’ (TI), integrating at one end of the targeted locus by HR with one flanking sequence of the vector accompanied by an apparent non-homologous end-joining (NHEJ) event at the other. Untargeted integration at nonhomologous sites also occurs, but at a lower frequency. Molecular analysis of TI at a single locus shows that this occurs as a consequence of concatenation of the transforming DNA, in planta, prior to integration, followed by HR between a single site in the genomic target and two of its repeated homologues in the concatenated vector. This reinforces the view that HR is the major pathway by which transforming DNA is integrated in Physcomitrella.

INTRODUCTION

Repair of double strand breaks in DNA is essential to ensure the integrity of genomes. The integration of transforming DNA into the genomes of higher eukaryotes occurs through the action of DNA double-strand break repair processes. While prokaryotes and lower eukaryotes favour the use of homologous sequences to repair chromosomal double-strand breaks (1) and to insert transforming sequences, illegitimate integration via microhomologies and non-homologous end-joining (NHEJ) is usual in higher eukaryotes.

Among plants, the moss Physcomitrella patens is an exception to this rule. When transforming DNA contains homology with the Physcomitrella genome, integration at the cognate genetic locus occurs at frequencies up to 100% (2–5), whereas in flowering plants, homologous recombination (HR) occurs at low frequency (10⁻⁴–10⁻⁵) (6). Since higher plants have attracted the most intense investigation of DNA repair processes, the mechanism of illegitimate recombination (IR) in plants is better characterized than that of HR. Unlike HR, IR frequently involves deletions and filler DNA insertion (7–10). Similarly, the random insertion of T-DNA sequences from Agrobacterium tumefaciens into plant genomes is also frequently associated with duplications or translocations of genomic sequences (11).

The prevalent models for double-strand break repair in somatic plant cells are based on single-strand-annealing (12) and the synthesis-dependent single-strand annealing mechanism (13). Both involve sequence homologies and can explain the capture of non-homologous or homologous sequences (1,14). The high rate of HR in Physcomitrella, compared with seed plants, could be due to (i) a different mechanism of recombination, (ii) deviations in proteins participating in recombination complexes or (iii) to a shift in the ratio of HR to IR without differences in the basic mechanisms. Unusually, we have found that whilst the majority of homologous transgene integration involves the accurate substitution of the genomic locus by the transforming DNA (targeted gene replacement-TGR), a significant proportion...
of gene targeting events in Physcomitrella result in insertion of the transgenic sequence adjacent to the target locus (targeted insertion-TI). Typically, TI comprises an HR event at one end of the integrant, accompanied by an apparent NHEJ event at the other (4). To elucidate the nature of DNA integration in Physcomitrella in more detail, we have examined transgene integration patterns for both single constructs and for batches of multiple, independent targeting constructs, and analyzed a number of independent genomic integration sites at the DNA sequence level.

**METHODS AND MATERIALS**

**Plant material**

*P. patens* (Hedw.) B.S.G. was cultured as protonemal tissue, either on agar medium overlain with cellophanes (15) (Leeds) or in liquid culture (16) (Freiburg). The techniques for protoplast isolation, transformation and regeneration of stably transformed plants have been described previously (4,17–19).

**Transformation vectors**

Targeted loci analyzed in this study included single-gene disruptions of a 5-fatty acid desaturase gene (GenBank accession AX155059) and the *PpRac-l* gene (GenBank accession AY870928) (4), and batch transformations (codelivery of more than one construct) using cDNA clones in a minimalized plasmid vector, disrupted by an nptII selectable marker introduced by Tn1000 transposition (18). Linearized cDNA constructs were prepared for transformation by restriction enzyme digestion and short (16–43 bp) non-homologous ends. The vector RacKXH (4) was constructed in the Gateway™ vector pMBL6attR (4). A linear fragment for moss transformation was obtained by NotI digestion (‘RacKXHN’), or by PCR (‘RacKXHP’) (Figure 3). These constructs carried 40 bp of terminal non-homology corresponding to the *AtB* sites of the Gateway™ vector.

**Molecular analysis of transgenic plants**

**Southern blot analysis.** Genomic DNA was isolated from 14 day-old transgenic plants developed from small protonemal inocula, using the small-scale DNA isolation procedures previously described (15,20). TGR at the *PpRac-l* locus was identified by PCR using primers 35Spro2R (p6) and 6f-termF (p7) in conjunction with external gene-specific primers 5’rF (p1) and 3’r3R (p5) (4). Untargeted junctions from plants in which 5’-TI had occurred were amplified by PCR with primer gRac3869S (p8) (‘sequence c’, the 3’-arm of the transforming cassette) in combination with primer racPro2R (p3) (‘sequence b’, the central, deleted region of the *PpRac-l* gene). Amplicons purified using a Qiagen Qiaquick PCR cleanup kit (Qiagen GmbH, Hilden) were sequenced using the ABI3130 capillary sequencer in the Faculty of Biological Sciences, Leeds University. Concatemer junctions in plants batch-transformed with cDNA constructs, were analysed using protocols for RACE–PCR (21,22). A 25 µl first-strand synthesis reaction contained 100 ng genomic DNA, 0.2 mM dNTPs, 1 pmol/µl primer and 0.01–0.08 U/µl Taq polymerase (Roche). Two reactions per transgenic line were set up, each with one of two outward primers located at the borders of Tn1000. The same PCR-cycle conditions were used for first-strand and second-strand synthesis (linear amplifications) and PCR amplification: following initial denaturation (5 min at 95°C), amplification was by 30 cycles of 0.5 min at 95°C–1.5 min at 60.0°C with a final extension of 4.5 min at 72°C. First-strand products were purified using the QIAquick PCR purification kit and eluted in 30 µl. These were incubated for 5 min at 65°C and directly cooled on ice. dC-tailing was performed in a 50 µl reaction volume with 31.5 U terminal transferase (MBI Fermentas) containing 1 mM CoCl₂ and 20 µM dCTP (20 min at 37°C) and stopped by heating to 72°C for 10 min. The DNA was again purified with the QIAquick PCR purification kit. First-strand synthesis (40 µl) contained 1 pmol/µl G₁₈-primer (Thermo Electron, Dreieich, Germany), 0.2 mM dNTPs and 0.08 U/µl Taq polymerase. PCR amplification of the products followed immediately in a 45 µl volume by adding the respective first-strand primer to a final concentration of 1.7 pmol/µl, additional 0.03 U/µl Taq polymerase and dNTPs and MgCl₂ to final concentrations of 0.2 and 2.5 mM respectively. Products were cloned in pGEM-T (Promega) and sequenced on an ABI-PriSm 377 sequencer (Applied Biosystems, Foster City, CA, USA) using the DYEEnamic ET terminator cycle sequencing kit from GE Healthcare (Freiburg, Germany).

**RESULTS**

**HR-mediated gene targeting results in accurate allele replacement and targeted insertion**

When *Physcomitrella* is transformed with a targeting construct (a cloned gene disrupted by the insertion of a selectable marker cassette), gene targeting is detected by PCR using ‘outward-pointing’ primers specific to the selection cassette in combination with ‘inward-pointing’ 5’- and 3’-gene-specific primers corresponding to sequences external to the targeting construct. Amplification with both pairs is diagnostic of TGR resulting from accurate HR at each end of the construct; amplification with only one pair is indicative of TI resulting from accurate HR at only one end. Failure to amplify a fragment with both pairs is indicative of transgene
insertion at an illegitimate site (in such cases, the wild-type locus is amplifiable using the external primers). The accuracy of the HR events can be established by sequence analysis of the amplified fragments.

Initially, we performed PCR-screens in 258 transgenic lines transformed with single disruption constructs (Table 1) and analyzed the sequences of 32 recombination junctions at three loci where integration was by TGR. Five single-copy TGRs of the PpRac-1 gene (GenBank accession AY870928) (4) were sequenced at each end following PCR-amplification of the HR-integration sites (Supplementary Figure 1A and B), as were five TGRs of a 5-fatty acid desaturase gene (GenBank accession AX155059). Additionally, twelve single-copy TGRs of the PpLea-1 gene (GenBank accession AY870926) (4,23) were sequenced at one end (Supplementary Figure 1C). In each case, the recombination junctions showed no deviations from the sequence of the native gene, confirming that HR-mediated replacement of the wild-type gene by the disrupted construct occurred in an entirely precise manner.

### Accurate gene targeting occurs following batch transformation using multiple constructs

It is also possible to achieve simultaneous targeting of multiple genes by cotransformation with batches of independent targeting constructs (17), however, the extent and accuracy of multiple targeting has not yet been systematically characterized. We therefore undertook a detailed molecular analysis of transgene integration following delivery of multiple constructs.

Efficient simultaneous delivery of multiple transforming DNAs has previously been used to create a large population of mutant plants (>50,000) by transformation with cDNA clones disrupted by the random insertion of a transposon-borne selection cassette (18,24). For efficient generation of this collection either normalized cDNA libraries, or batches of 7 or 20 disrupted cDNA clones, were delivered simultaneously to protoplasts. Since cDNA sequences typically exhibit shorter lengths of contiguous sequence homology with their genomic targets (due to the presence of introns within the genomic DNA), it was clearly important to verify that these knock-out constructs integrated into their target loci with a high degree of accuracy. We therefore estimated the numbers of integrations, and the nature of the integration events at targeted loci. The number of integration events was analysed by Southern blot analysis (Figure 1) in 146 independently regenerated transgenic lines (91 in this study and 55 in a previous study (18)). On average, 11 copies of the selection cassette were integrated per plant. A subset of these transgenic plants [n = 34: 18 in this study and 16 previously (18)] were further analyzed to determine the number of integration sites. The number of integration sites per plant showed a high degree of variation, from 1 to 16 loci, with the majority of plants (59%) exhibiting transcript gene integration at between 3 and 7 loci (Figure 1).

In these experiments, the DNA used for transformation was linearized by restriction enzyme digestion to cut the cDNA insert out of the vector. Consequently, both nptII-disrupted cDNAs and vector backbone were delivered to protoplasts. The adventitious integration of vector sequence was demonstrated both by Southern blotting (Figure 1) and by the isolation and sequencing of concatemer structures that comprised cDNA and vector sequences from three ‘batch mutants’.

In mammalian cells, the copy number of integrated transgenes is related to the amount of DNA delivered per transformation (25). We tested this for Physcomitrella. Protoplasts were transformed with 5–50 µg DNA/transformation and the numbers of nptII copies per plant were estimated by Southern hybridization. The efficiency of transformation increased dramatically with the quantity of DNA, reaching an optimum at 20 µg, but the number of integrated transgenes per plant was not influenced by the quantity of DNA per transformation (Table 2).

The nature of the targeting events (TGR versus TI) was determined in plants following the batchwise delivery of

### Table 1. Outcome of transformation with single and multiple gene fragments.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Transforms</th>
<th>HR/HR</th>
<th>5'-TI</th>
<th>3'-TI</th>
<th>Illegitimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>RacKXHN</td>
<td>130</td>
<td>57</td>
<td>39</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>RacKXHP</td>
<td>71</td>
<td>36</td>
<td>26</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>PpLEA-1</td>
<td>45</td>
<td>22</td>
<td>0</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>tKO25.2</td>
<td>12</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>7 cDNAs</td>
<td>140 loci</td>
<td>40</td>
<td>60</td>
<td>40</td>
<td>nd</td>
</tr>
</tbody>
</table>

The numbers of transformants of each class are provided. HR/HR indicates the number of targeted gene replacements. 5'-TI refers to targeted insertions resulting from HR only occurring in the 5'-arm of the targeting construct. 3'-TI refers to targeted insertions occurring only in the 3'-arm. Illegitimate events are transfectants in which the targeting construct did not integrate at the targeted locus.
Table 2. Transformation efficiency as a function of quantity of transforming DNA. Protoplasts were transformed with 5–50 μg DNA in 38 independent transformation experiments.

<table>
<thead>
<tr>
<th>Quantity of DNA (μg)</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>NptII copies/plant</td>
<td>11.9</td>
<td>11.3</td>
<td>nd</td>
<td>12.1</td>
<td>nd</td>
<td>nd</td>
<td>11.5</td>
</tr>
<tr>
<td>Transforms/ transformation</td>
<td>2.2</td>
<td>6.6</td>
<td>14.3</td>
<td>19.1</td>
<td>18</td>
<td>19.1</td>
<td>20.1</td>
</tr>
<tr>
<td>Transforms/μg DNA</td>
<td>0.44</td>
<td>0.66</td>
<td>0.95</td>
<td>0.96</td>
<td>0.72</td>
<td>0.62</td>
<td>0.4</td>
</tr>
</tbody>
</table>

By PCR amplification and sequencing of the untargeted junctions, the average number of stable transformatns obtained, per transformation was shown, as are the average numbers of transgenes integrated in 22–24 representative transgenic lines regenerated following transformation with 5, 10, 20 and 50 μg DNA, as estimated by Southern blot hybridization using an nptII probe.

Figure 2. Possible models for integration of transforming DNA. (A) The transforming fragment integrates by HR in both arms of the vector, and a perfect TGR occurs. (B) The transforming fragment integrates by HR in one arm of the vector: the other arm invades the breakpoint to integrate by NHEJ. (C) The transforming fragment undergoes head-to-tail concatenation prior to integration. Integration occurs in region 'a' of the targeted locus by two HR events occurring in separated regions 'a' in the concatemer. 'S' and 'sel' indicate the nptII selection cassette. 'B1' and 'B2' indicate the attB1 and attB2 sites that terminate the transforming DNA fragment. (D) The transforming fragment is circularized by NHEJ. Recombination at a single point in region 'a' generates a transgenic locus indistinguishable from that in 'C'.

How do targeted insertion events arise?

We envisage three ways in which a targeting construct might integrate via TI. These alternatives are illustrated in Figure 2B–D. Consider a vector constructed from a cloned gene 'a-b-c' in which the central segment ('b') is replaced by a selectable marker cassette 'sel'. This vector additionally is terminated by two distinct, non-homologous, cloning vector-derived sequences 'B1' and 'B2'. TGR will result in a precise replacement of the endogenous locus 'a-b-c' by the sequence 'a-sel-c' (Figure 2A). TI might occur through an HR event between one arm of the vector and its target, accompanied by an NHEJ event between the other end of the vector and the DNA double-strand break in the targeted locus. This will generate a transgenic locus 'a-sel-c-B2-a'-b-c'. (Figure 2B), where 'a' represents a terminally truncated sequence 'a'. Alternatively, the vector may form a concatemer within the cell, and integrate by an HR event between one arm of the vector and its target, and a second HR event between the DNA double-strand break in the target and another identical homologous sequence available in the concatenated vector, as illustrated in Figure 2C. Here, the resulting transgenic locus will comprise 'a-sel-c-B2-(B1-a-sel-c-B2)-a-B1-a-b-c'. Finally, if the vector becomes circularized, a single HR event between this and the target locus would generate the transgenic locus 'a-sel-c-B2-B1-a-B1-a-b-c', in which only a single copy of the vector becomes inserted (Figure 2D). These outcomes are predictable and testable by PCR amplification and sequencing of the untargeted junction sequence in TI events.

We analyzed transgenic plants in which the PpRac-1 locus (Figure 3A) was targeted. Plants were transformed with the targeting vector 'RacKXH' (4). This comprises 1119 bp from the 5'-end of the PpRac-1 gene (sequence 'a') and 529 bp from the 3'-end of this gene (sequence 'c') interrupted by an nptII selectable marker cassette ('sel') which replaced seven different constructs, in which the selection cassette was flanked by at least 500 bp of genomic DNA on either side. The nature of targeted integration events at each locus was analyzed by PCR in a number of independently derived transgenic plants (in total, 140 targeted loci were analyzed). At 40 of these loci, accurate TGR had occurred. The remaining 100 gene targeting events comprised TIs, with HR occurring either in the 5'-flanking sequence (n = 60) or the 3'-flanking sequence (n = 40) of the targeted locus (Table 1, row 5). Nine independently regenerated transgenic plants obtained in this way were further analyzed by PCR, using primers specific to the selection cassette and to each cDNA used in the targeting constructs, in order to detect the numbers of transgenes integrated. On average, 86% of the targeting constructs were found to have become integrated into the genomes of these plants, either by HR or IR.

The accuracy of targeting events in batch transformants was determined by sequencing PCR-amplified junctions from 12 loci at which TGR had occurred. This analysis showed precise integration of the targeting cassette at the cognate locus, but also that the recombination event could occur at any point within the homologous sequence present in the targeting construct, as evidenced by loss of intron sequences occurring as a consequence of recombination between cDNA and genomic sequence (Supplementary Figure 2). In one plant only, we observed an anomalous event: the duplication and translocation of a genomic sequence of at least ca. 1.5 kb lying adjacent to the transgene integration site (Supplementary Figure 3). The genomic integration site of the duplicated sequence could not be determined.

How do targeted insertion events arise?

We envisage three ways in which a targeting construct might integrate via TI. These alternatives are illustrated in Figure 2B–D. Consider a vector constructed from a cloned gene 'a-b-c' in which the central segment ('b') is replaced by a selectable marker cassette 'sel'. This vector additionally is terminated by two distinct, non-homologous, cloning vector-derived sequences 'B1' and 'B2'. TGR will result in a precise replacement of the endogenous locus 'a-b-c' by the sequence 'a-sel-c' (Figure 2A). TI might occur through an HR event between one arm of the vector and its target, accompanied by an NHEJ event between the other end of the vector and the DNA double-strand break in the targeted locus. This will generate a transgenic locus 'a-sel-c-B2-a'-b-c'.
Concatenation of transforming DNA: multiple targeting vectors

Although tandem, head-to-tail concatemers are the only structures detected when a single transforming fragment is delivered to protoplasts, this is not the case when multiple independent fragments are delivered in a batch transformation. Isolation and sequence analysis of single transgenic loci at which concatenated cDNA-based vectors had integrated by Irr revealed concatemers between different cDNA constructs in all possible orientations (Figure 4). We found six loci at which concatemers arose from HR between identical constructs (data not shown). Eight loci contained equimolar quantities of both the disrupted PpRac-1 gene and the plasmid vector backbone. Additionally, the fragments terminated in NotI cohesive ends. In the latter case, protoplasts were transformed with a highly pure DNA fragment comprising the disrupted PpRac-1 gene that retained the terminal attB sites, but with non-cohesive ends.

We obtained 201 transgenic plants using these fragments, of which 93 exhibited accurate TGR. Among the remainder, 65 had undergone TI in the S'-arm of the targeting construct, 13 had undergone TI in the 3'-arm of the targeting construct and 30 contained DNA integrated ectopically at unknown sites in the genome (4). We analyzed the DNA in these plants by PCR to determine two features. (i) The nature of concatenation between transforming fragments and (ii) the nature of the non-targeted junction between the transforming DNA and its target locus, in plants exhibiting S'-TI.

Concatenation of transforming DNA: single targeting vectors

Using the 'outward-pointing' primers 'p6' and 'p7', corresponding to the nptII selection cassette (Figure 3B), it is possible to determine whether concatemers form, and their relative orientation. 'Head-to-tail' concatemers will be amplified using both primers in conjunction (Figure 3H). 'Head-to-head' and 'tail-to-tail' concatemers will be amplified using a single primer only (e.g. Figure 3I).

We first analyzed the 30 plants in which the DNA had integrated ectopically. Use of a single primer failed to generate an amplification product (Figure 3D), whereas the combination of primers 'p6' and 'p7' resulted in amplification of a product from 23 plants, and in 20 of these, the fragment was of the size predicted for a 'head-to-tail' concatemer (Figure 3E and H). Three plants yielded larger amplicons, probably resulting from DNA rearrangement in the generation of the transgenic locus. Similarly, when we analyzed plants in which S'-TI had occurred, 13 out of 14 plants tested exhibited the characteristic amplification pattern of head-to-tail concatemers (Figure 3G and H). Head-to-head and tail-to-tail concatemers were not detected (Figure 3F and I).

We therefore conclude that transgenic loci containing multiple copies of transforming DNA comprise head-to-tail concatemers of the transforming fragment.

Using primers external to the targeted locus ('p1' and 'p5' in Figure 3A), we were able to amplify single-copy integrants that resulted from accurate TGR (4). However, no single-copy integrants resulting from TI could be identified in the PpRac-1 locus.

Concatenation of transforming DNA: multiple targeting vectors

A: Structure of the PpRac-1 gene. Exons shown as shaded boxes, and non-coding DNA (5'- and 3'-flanking sequences and introns) as a solid line. Restriction sites for BglII (B), EcoRI (E), HindIII (H), SalI (S), SphI (Sp) and XbaI (X) are indicated. PCR primers p1 and p3 are external primers used to confirm gene targeting in conjunction with the ‘outward’ primers p6 and p7 (B). Primers p2 and p4 were used to amplify the PpRac-1 gene and create the targeting fragment RacKXH. (B) The targeting construct RacKXH. The selection cassette replaces the central XbaI–HindIII fragment. (C) The outcome of a TI event of the type shown in Figure 2C. ‘B1’ and ‘B2’ denote the AttB1 and AttB2 sites flanking the transforming fragment. Primer p3 is specific for the region of the PpRac-1 gene deleted in the construction of RacKXH (‘region b’ in Figure 2), and is used in conjunction with primer p8, corresponding to ‘region c’, for amplification of 3'-NHEJ junctions, indicated by the grey bar (D and E). DNA from transgenic plants in which integration occurred at non-targeted sites. (D) Amplification using primer p6. (E) Amplification using primers p6 and p7 together. (F and G) DNA from transgenic plants in which S'-TI at the PpRac-1 locus occurred. (F) Amplification with primer p6. (G) Amplification with primers p6 and p7 together. (H) Predicted structure of concatenated transforming DNA by ‘head-to-tail’ ligation. (I) Predicted structure of concatenated transforming DNA by ‘head-to-head’ ligation.
concatemers that were derived from end-to-end ligation, of which three were in head-to-head (Figure 4A), three in tail-to-tail (Figure 4B) and two in head-to-tail orientation, respectively (Figure 4C). Additionally 3 loci contained concatemers that were probably formed by recombination between microhomologous sequences in different cDNAs (Figure 4D). We routinely observed short deletions at the ends of each concatenated transgene with an average length of 10 bp ($n=10$). We therefore conclude that concatenation of transforming DNA by NHEJ is a frequent occurrence during protoplast transformation.

It is possible that concatemers could have formed by annealing between complementary cohesive ends of DNA molecules in the highly concentrated DNA mixture prior to transformation, rather than in planta, following DNA delivery. This was tested in cDNA batch transformation experiments. To discriminate between the two possibilities, the DNA was either used directly for transformation or heated at 60°C for 10 min and chilled on ice before transformation to prevent concatenation, in vitro, via annealing of cohesive ends. Transformation efficiency was determined as the mean number of transgenic plants per transformation. The mean numbers of integrated nptII cassettes per plant and the numbers of integration loci per plant were estimated by Southern blot analysis of transgenic lines ($n=9$).

DNA was used directly, or heated at 60°C for 10 min and chilled on ice before transformation to prevent concatenation, in vitro, via annealing of cohesive ends. Transformation efficiency was determined as the mean number of transgenic plants per transformation. The mean numbers of integrated nptII cassettes per plant and the numbers of integration loci per plant were estimated by Southern blot analysis of transgenic lines ($n=9$).

Table 3. Transformation with heat-denatured DNA protoplasts were transformed with batches of disrupted cDNA clones (50 μg DNA) linearised by digestion with SdaI.

<table>
<thead>
<tr>
<th>Transformation efficiency</th>
<th>nptII cassettes</th>
<th>Integration loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA (untreated)</td>
<td>30.4</td>
<td>14.8</td>
</tr>
<tr>
<td>Heat-treated:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-sided t-test: $t =$</td>
<td>0.55</td>
<td>0.204</td>
</tr>
<tr>
<td>$N = 9$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Plants exhibiting TI in the 5′-arm of the PpRac-1 gene (Figure 3C) were selected for analysis of the untargeted (3′) integration junction, the shorter length of the 3′ targeting sequence enabling the PCR amplification and subsequent sequencing of this junction using the primers ‘p3’ and ‘p8’ (Figure 3A and C). Because primer ‘p3’ corresponds to the central region of the PpRac-1 gene (sequence ‘b’), deleted from the transforming construct, only the 3′-untargeted junction will be amplified using this primer combination.

When 19 transgenic lines carrying 5′-TIs of the RacKXHP fragment were subjected to amplification with these primers, 10 yielded a PCR product. Similarly, of the 26 transgenic lines carrying 5′-TIs of the RacKXHP fragment, 21 yielded a PCR product. All but one of these latter amplicons were electrophoretically identical, the remaining fragment being slightly longer. This fragment, together with nine other RacKXHP amplicons and all ten RacKXHN amplicons were sequenced.

All but 2 of the 20 amplified junction sequences contained the ‘cAtB2-AnB1-a-b’ configuration consistent with
formation of the untargeted junction by HR as illustrated in Figure 2C and D. However, all of the junction sequences were characterized by local rearrangements between the AttB sites.

In the lines transformed by RacKXHN, one transformant contained a near-perfect match with the predicted concatemer junction sequence (Figure 5A: line T28/3–56). Three lines contained a small deletion in the microhomologous XbaI–NotI–XbaI sequence located between the AttB sites (Figure 5A: lines T28/3–20, 3–63, 3–64), whilst a further three exhibited a deletion in a region of microhomology between the two AttB sites themselves, but still carried the recognizable signatures of the AttB1 and AttB2 motifs (Figure 5A: lines T8/3–61, 3–62 and 3–89). Two lines contained additional sequences derived from the vector backbone. In one case, this comprised the entire vector (Figure 5A: line T8/3–45). In the other line, however, the structure of the NHEJ junction was more complex (Figure 5A: line T8/3–44). The amplified fragment contained a 455 bp insertion between the AttB2 and AttB1 sites, made up of two sequences of 212 and 243 bp derived from the plasmid vector backbone. These were not contiguous sequences within the source plasmid. Instead, they appear to result from a 1790 bp deletion. We suggest that this arises from a recombination event between two microhomologous CA dinucleotides that flank the deleted intervening sequence. Because in the generation of these transformants, no effort was made to purify the targeting fragment free of its vector, it is likely that T8/3–44 and T8/3–45 derived from an initial extrachromosomal concatenation event between the targeting fragment and the free vector sequence. The tenth line analyzed (Figure 5A: line T28/3–29) was deleted for the AttB2 and AttB1 sites, but still retained the 'c-a-b' sequences in the orientation predicted in Figure 2C and D.

In the lines transformed with RacKXHP, 9 of the 10 fragments contained the 'c-AttB2-AttB1-a-b' configuration. The tenth (Figure 5B: line T15/5–24) was deleted for the AttB sites between a flanking microhomology, but retained the characteristic 'c-a-b' sequence arrangement. Of the nine fragments containing AttB sites, eight displayed a small deletion between regions of microhomology on the AttB2 and AttB1 sequences (Figure 5B: lines T15/5–1, 5–19, 5–22, 5–23, 5–31, 5–48, 5–49 and 5–52) whilst the ninth (Figure 5B: line T15/5–55b, corresponding to the larger, amplified fragment) contained a small insertion of 'filler DNA' of unknown origin, characteristic of NHEJ (7–10).

These data are consistent with a model for integration by TI in which two HR events occur between the target locus and a tandemly repeated targeting construct formed by concatenation of the transforming DNA (Figure 2C).
adventitious integration of vector DNA, elsewhere within the genome, nor from an ectopic insertion of a targeting cassette. As with all analysis of mutant strains, the standard precaution of genetic analysis of mutants should be routinely undertaken, by crossing to a wild-type strain in order to demonstrate Mendelian segregation of the mutant phenotype with a targeted locus. Complementation of the mutant strain by further transformation with a wild-type allele should provide additional confirmation that a particular targeted gene is responsible for the phenotype in question.

Integration of targeting DNA into the homologous genomic locus frequently occurs via HR at both ends. However, transforming DNA can integrate at multiple genomic loci (from 3 to 7 for batch transformation) due to additional IR. Additionally, multiple copies of transgenes are integrated (ave = 11). We ascribe this to the concatenation of transforming DNA being the rule, rather than the exception: we have shown previously that between 40 and 85% of TGR loci generated by single transformation constructs contain multiple copies of the targeting cassette (4).

Concatenation is most frequent in Tl loci. Significantly, it appears that such events do not arise by a combination of HR and NHEJ, as first postulated (4), but arise wholly by HR-mediated integration of concatenated DNA. It is known that concatenated copies of plasmid DNA insert at one or a small number of loci when circular molecules are delivered to eukaryotic cells (25–28), and linear DNA delivered to mammalian cells can undergo end-to-end ligation, by NHEJ, all possible orientations of ligated fragments being detectable (25). Such a process is the likely cause of the incorporation of nonspecific ‘carrier’ DNA in transgenic loci in plant cells, following its co-delivery with a circular transforming plasmid (29). Thus, after entering the cell, transforming DNA can undergo both fragmentation and religation events prior to its insertion into the genome.

In Physcomitrella, a typical transformation experiment generates, in addition to stable (concatemers) transformants, a larger number of unstable transformants, that maintain transgenes only so long as selection is maintained. These greatly outnumber stable transformants, when circular DNA is delivered to protoplasts. Unstable transformants contain extrachromosomal concatemers of the transforming DNA (30) and the generation of such concatemers may occur prior to and preferentially over transgene integration for circular molecules.

Delivery of linear DNA reduces the frequency of unstable transformants (17), but evidently concatenation still occurs. This must occur in planta, since prior heat-denaturation of potential cohesive ends prior to DNA delivery did not prevent the recovery of transformants containing concatenated transgenes (Table 2). Recombination between a linear concatemer and the target locus would generate TI events by the model indicated in Figure 2C. Although recombination between a circularized fragment of transforming DNA would also have the same outcome (Figure 2D), rates of transformation by circular DNA are orders of magnitude lower than those resulting from delivery of linear DNA in Physcomitrella (17) and consequently we do not favour a model involving circular intermediates.

The occurrence of small deletions between the terminal sequences (‘c-a’) that are found at the untargeted junction in TI events are characteristic of NHEJ events, as is the occurrence of insertions via DNA sequence microhomologies (10,31). We observe just such insertions and deletions in nearly all the untargeted junctions analyzed, strongly supporting the view that TI involves ligation of fragments by NHEJ prior to homologous integration into the genome.

We propose that TI events involve the following series of events. Following uptake of massive numbers of linear fragments by the cell, their ends would be perceived by the DNA damage repair machinery as the apparently catastrophic occurrence of double-strand breaks. The most rapid mechanism by which these could be ‘repaired’ is by NHEJ between these fragments. A priori, we would expect the ends to be joined at random, so that concatemers containing head-to-tail, head-to-head and tail-to-tail junctions should be formed in equal number. Such randomly joined concatemers are indeed found integrated in plants batch-transformed with multiple, heterogeneous cDNA targeting constructs. However, only head-to-tail concatemers of the transforming DNA are detected in plants transformed with single genomic targeting constructs. End-to-end ligation of identical sequences to form a large inverted repeat probably generates a transforming molecule either (i) not readily detectable by PCR, or (ii) not suitable for TI or (iii) a transgenic locus in which the inverted repeats are unstable, and are excised.

The model presented in Figure 2C requires a concatemer to be the substrate for TI in the target locus. The mechanism we suggest for such insertion is elaborated in Figure 6, and is based on current models for HR-mediated DNA repair (32). One end of the concatemer (region ‘a’) initiates HR (Figure 6B), following recruitment of DNA repair proteins resulting in resection of the targeting construct to generate 3’-ssDNA (presumably complexed with PpRad51 proteins (33) and Rad52 epistasis group homologues) that can initiate invasion of the homologous genomic target (32). Strand scission of the targeted locus (Figure 6C) would then generate a genomic 3’-ssDNA that could, in turn, invade the concatemer targeting construct (Figure 6D). For a long concatemer, invasion by genomic DNA of a nearby repeat of region ‘a’ in the transforming DNA might be sterically more favourable than invasion of the genomic region ‘c’ by the distant termini of the targeting construct. We note that the length of the concatemer is not fixed, but if it were only a dimer then only a single copy of the targeting construct would be inserted, and consequently no PCR products would be obtained using the outward-facing primers ‘p6’ and ‘p7’ (Figure 3H). Since in 13 of the 14 plants we analyzed, a PCR product was obtained (Figure 3G), the insertion substrate must have been at least trimeric. This may reflect either (i) that dimers are structurally unsuitable for insertion, (ii) that dimers are preferentially integrated as multicycle TGRs (4), or (iii) may be a stochastic reflection of the population of available concatemers. If multiple subunit concatemers were very common, the chance of the second recombination event occurring in the concatemer subunit adjacent to the first would be low, and our sample may be too small to identify such an event.

Overall, the observation that the overwhelming majority of targeted transformation events in Physcomitrella involve only HR, supports the contention that HR represents the predominant route for transgene insertion in this species.
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Tanja Egerner, José Granado, Annette Hohe and Christina Reinhard for their contribution to this work and Antje Bakker, Sabine Glatzel, Joachim Lonien, Hans-Jürgen Schwarz and Jasmin Weise for excellent technical assistance. This work was supported by the EU FPV ‘PREGENE’ programme (DJC), the UK BBSRC Physcomitrella EST Programme (ACC) and BASF Plant Science, Germany (RR). Funding to pay the Open Access publication charges for this article was provided by the Universities of Leeds and Freiburg.

Conflict of interest statement. None declared.

REFERENCES


