Parameters determining the efficiency of gene targeting in the moss *Physcomitrella patens*

Yasuko Kamisugi, Andrew C. Cuming* and David J. Cove

Centre for Plant Sciences, Faculty of Biological Sciences, Leeds University, Leeds LS2 9JT, UK

Received September 7, 2005; Revised and Accepted October 14, 2005

**ABSTRACT**

In the moss *Physcomitrella patens*, transforming DNA containing homologous sequences integrates predominantly by homologous recombination with its genomic target. A systematic investigation of the parameters that determine gene targeting efficiency shows a direct relationship between homology length and targeting frequency for replacement vectors (a selectable marker flanked by homologous DNA). Overall homology of only 1 kb is sufficient to achieve a 50% yield of targeted transformants. Targeting may occur through homologous recombination in one arm, accompanied by non-homologous end-joining by the other arm of the vector, or by allele replacement following two homologous recombination events. Allele replacement frequency depends on the symmetry of the targeting vector, being proportional to the length of the shorter arm. Allele replacement may involve insertion of multiple copies of the transforming DNA, accompanied by ectopic insertions at non-homologous sites. Single-copy and single insertions at targeted loci (targeted gene replacements, ‘TGR’) occur with a frequency of 7–20% of all transformants when the minimum requirements for allele replacement are met. Homologous recombination in *Physcomitrella* is substantially more efficient than in any multicellular eukaryote, recommending it as the outstanding model for the study of homologous recombination in plants.

**INTRODUCTION**

Plant genetic transformation was first achieved in 1983 (1–3) and has since become a powerful tool for studying all aspects of the regulation of plant development, as well as for the generation of genetically improved crops for agricultural and biotechnological use. In plants, transforming DNA integrates into the nuclear genome at more or less random sites. Despite many attempts to develop gene targeting technology, involving the insertion of exogenous DNA by homologous recombination (HR), this occurs only sporadically, and at low frequency (not >0.1% of transformants) (4,5) in flowering plants. This is unfortunate, since efficient HR is a powerful molecular tool by which precise targeted gene replacement (TGR) can be achieved without additional disruption of the targeted genome. Gene targeting is routine and highly efficient in yeast, and is also a preferred method for the manipulation of mammalian cells, even though the frequency of HR in mammalian cells is lower than that of non-homologous integration. Only very recently have gene targeting rates of the order of 5–20% of transformants been reported in *Arabidopsis* (6).

In contrast with flowering plants, HR appears to be the default pathway for the incorporation of exogenous DNA in the model bryophyte, *Physcomitrella patens* (7). Gene targeting frequencies of up to 100% of transformants have been reported for this organism: the highest for any multicellular eukaryote. Since the first reports of HR-mediated gene targeting in this species (8,9), the potential of *Physcomitrella* as a model for the study of plant development has increased, as the potential for undertaking highly sophisticated reverse genetic functional analysis has become apparent (10,11). Subsequently, a substantial body of molecular resources has been developed to facilitate gene identification, transformation and the analysis of gene function (12–17), culminating in the recent selection of this species for complete genome sequencing: the first non-flowering land plant to be so selected. Notwithstanding these features, our knowledge of the mechanism by which HR occurs preferentially in *Physcomitrella* remains limited (18,19). Little is known of the molecular mechanisms involved, and since only relatively small numbers of gene-targeted transgenic plants have been generated, principally with the aim of producing specific mutations for gene function analysis, there is only a limited appreciation of the parameters of the transformation process that result in targeted, rather than ectopic insertion of transforming DNA.
Consequently, we have set out to determine these parameters, and here we present the first large-scale, systematic analysis of the transformation procedure, with the aim of determining more precisely the experimental parameters that contribute to the high efficiency TGR in *Physcomitrella*. In this article, we focus on the requirement for the length of DNA sequence homology and the geometry of the targeting constructs at several genetic loci, to produce a general set of criteria that define the requirements for efficient gene targeting.

**MATERIALS AND METHODS**

**Plant material**

The ‘Gransden’ strain of *P. patens* (Hedw.) B.S.G. (20) was propagated as protonemal culture on ‘BCD’ agar medium containing 1 mM CaCl$_2$ and 5 mM ammonium tartrate, overlaid with cellophane, as described previously (21). Tissue was vegetatively propagated by homogenisation and subculture every 7 days.

**Construction of transformation vectors**

To facilitate the construction of ‘replacement vectors’ (21), comprising *Physcomitrella* genomic DNA fragments interrupted by a Cauliflower Mosaic Virus 35S promoter-nptII-CaMV poly(A)-addition site selection cassette, we first constructed the intermediate cloning vector ‘pMBlattR’: a Gateway™ (Invitrogen) destination vector (GenBank accession no. DQ228132). This enabled the cloning of PCR-amplified *Physcomitrella* genomic fragments that could be disrupted by the insertion of the selection cassette in selected restriction sites, and the subsequent isolation, for transformation of moss protoplasts, of linear ‘moss DNA-moss DNA’ fragments either by NotI digestion, or by PCR amplification using gene-specific primers. Full details of the construction of this vector, and the sequences of all PCR primers used in this study, are provided in the Supplementary Data.

**Physcomitrella transformation**

Transformation of *Physcomitrella* followed established protocols (20), utilizing direct uptake of DNA in the presence of polyethylene glycol. Linear DNA (10–15 μg) was used for each transformation. The DNA was purified by phenol/chloroform extraction followed by ethanol precipitation. Finally, the DNA was resuspended in up to 30 μl dH$_2$O. Following DNA delivery, protoplasts were embedded in agar overlaying protoplast regeneration medium and incubated for 6 days, by which time protoplasts had regenerated. These were then transferred to selection medium containing G418 (50 mg/l, 10–14 days). The selection protocol (21) involved sequential replica picking of regenerating colonies on to sequential replica picking of regenerating colonies on to 12% regeneration medium and incubated for 6 days, by which time protoplasts had regenerated. These were then transferred to selection medium containing G418 (50 mg/l, 10–14 days). The selection protocol (21) involved sequential replica picking of regenerating colonies on to medium lacking G418 (14 days), followed by a second period of selection (50 mg/l, 14 days) in the presence of the antibiotic, to ensure the recovery of stably transformed plants.

**Analysis of transformed plants**

In order to analyse large numbers of transformed plants for the presence or absence of gene targeting events, PCR-based assays were used. DNA was isolated from transgenic plants which had developed after 14 days incubation from small protonemal inocula using the small-scale DNA isolation procedure described previously (21). Transformants were analysed for gene targeting events by PCR. Primers (35Spro2R and g6termF: Supplementary Data) were designed, allowing amplification outward from the selection cassette when used in conjunction with gene-specific primers that annealed with the sequence of the targeted genes, outside the region of homology incorporated in the targeting vector. Southern blot analysis of genomic DNA was carried out following restriction endonuclease digestion of ~2.5 μg DNA and electrophoresis in a 0.7% agarose gel. The DNA was transferred to Hybond N-Plus (AP Biotech) by high-salt capillary transfer for hybridization with digoxigenin-labelled probe sequences. Labelling, hybridization and detection followed the procedures set out in the ‘DIG Application Manual for Filter Hybridization’ (Roche Diagnostics, Mannheim).

**Statistical analyses**

The effect of genetic locus and of the location of the homologous sequence in relation to the selection cassette on the frequency at which HR occurs was examined by analysis of co-variance (22), using Systat software. The frequency of HR is defined as the proportion of transgenic plants in which HR has occurred between a sequence in a construct and the homologous genomic sequence. Genetic locus, location of the homologous sequence (3' or 3') and locus by location interaction were used as main effects, and length of homology was used as a covariate in the analysis. This covariate was included because of the large effect that length of homology has on frequency of HR. Analysis of co-variance requires that the dependent variable (in this case the frequency of HR) has a Gaussian distribution. The frequency of HR exhibits a binomial distribution (the outcome is defined as whether or not HR has occurred) and so must be transformed to generate a Gaussian distribution. To achieve this, we used the arcsine of the square root of the frequency, a commonly used transform (22).

**RESULTS**

**Gene targeting constructs**

Five *Physcomitrella* genes were selected for targeting experiments. These were either members of small multigene families, or single-copy genes which were chosen because preliminary experiments had shown that the creation of null alleles by homologous recombination did not adversely affect the regeneration of protoplasts and the subsequent viability of the protonemal stage of plant growth. *PpPum-1* (GenBank accession no. AY870927) is a member of the ‘PUF-domain’ RNA-binding protein family, exemplified by the *Drosophila* polarity determinant ‘pumilio’. Southern blotting experiments (data not shown) indicate this to be a member of a small family of highly similar genes in *Physcomitrella*, of which two have been cloned and sequenced in our laboratory. *PpRac-1* (GenBank accession no. AY870928) is a member of a small gene family encoding Rho-1 small GTPase proteins (‘Rac’ in yeast). Three highly similar cDNA clones have been characterized from *Physcomitrella* (23), and we have isolated their corresponding genomic sequences. Fragments corresponding to the *PpLEA-1* and *PpRac-1* genes were PCR-amplified and
cloned in the Gateway™ destination vector pMBL6atrR (see Supplementary Data for details) to create replacement vectors in which the coding sequence was disrupted by the insertion of an nptII selection cassette. The structures of the PpPum-1 and PpRac-1 genes, and the replacement vectors derived from them, are illustrated in Figure 1A and B. These two genes were used for the majority of transformation experiments directed towards determining the requirement for homology length in relation to gene targeting frequency. Over 5 kb sequence information is available for each gene, including the complete protein coding sequences, 5'- and 3'-UTRs and 5'- and 3'-flanking sequences. PpPum1 comprises 11 exons and 10 introns. PpRac-1 comprises 7 exons and 6 introns. In each case, introns and non-coding sequences comprised a significant proportion of the constructs, these sequences being almost entirely divergent between the individual members of each gene family, thus differentiating each family member. For PpPum-1, a 3396 bp fragment was amplified and cloned in pMBL6atrR. The nptII selection cassette was cloned into restriction sites within the coding sequence, as indicated in Figure 1A, to create three targeting vectors with different lengths of coding sequence homology flanking the selection cassette, and with varying degrees of symmetry with respect to the position of the selection cassette within the coding sequence. Digestion with NotI was used to liberate the targeting constructs from the vector for transformation of Physcomitrella. Additional targeting vectors with shorter homology lengths were obtained by PCR amplification of the ‘PumKH’ construct using gene-specific primers (Table 1 and Figure 1A).

Using a similar strategy, a 3681 bp fragment of the PpRac-1 gene was cloned and used for the construction of a second set of targeting vectors, again varying in the length and symmetry of the coding sequence homology flanking the nptII selection cassette. Again, PCR primers were designed to enable the amplification of a targeting construct with equal homology lengths on either side of the selection cassette (Table 1 and Figure 1B).

In order to investigate the efficiency with which other loci were targeted, three further constructs were prepared, derived from PpMago—a homologue of the Drosophila ‘magos-nashi’ gene (GenBank accession no. AY878246) (Figure 1C), a single-copy gene in Physcomitrella, PpWD-1, encoding a WD-40 protein similar to members of the Arabidopsis ‘MSI’ gene family (GenBank accession no. AY870929) (Figure 1D, and PpLEA-1 encoding a group 1 LEA protein (GenBank accession no. AY870926) (24) (Figure 1E).

![Figure 1. Genes and targeting constructs. The five genes and the targeting constructs derived from them are illustrated.](image-url)

- **A:** PpPum-1
  - **PumKH:** 1097 bp
  - **PumKE:** 1647 bp
  - **PumKB:** 2574 bp

- **B:** PpRac-1
  - **RacKBgH:** 293 bp
  - **RacKBg:** 988 bp
  - **RacKXH:** 119 bp
  - **RacKSSP:** 940 bp
  - **RacKSS:** 2149 bp
  - **RacKE:** 242 bp
  - **RacKH:** 3155 bp
  - **RacKX:** 1119 bp

- **C:** PpMago
  - **Mag:** 870 bp

- **D:** PpWD-1
  - **YWD:** 1216 bp

- **E:** PpLEA-1
  - **Lea:** 932 bp
The molecular outcomes of transformation were determined in 1811 transgenic plants, derived following transformation with 22 different constructs, representing five genetic loci. The table indicates the lengths of the homologous sequences in each arm of the targeting construct, the number of homologous recombination events occurring in the 5′-arm and the 3′-arm alone, as well as the number of double-recombination (‘replacement’) and untargeted transformation events. The number of targeting events as a percentage of total transformants is given in brackets (%). Constructs indicated with an asterisk represent fragments generated by PCR that were terminally homologous with their genomic targets. All other fragments used for transformation were generated by NotI digestion of cloned sequences and contained 40 bp terminal non-homology at either end derived from the AttB sites in the cloning vector.

### Identifying gene targeting

In order to generate statistically-analysable results, it was necessary to obtain large numbers of stably transformed plants with each targeting construct. Each transgenic plant was first analysed by PCR to determine whether or not the transforming DNA had integrated at the homologous locus. This was necessary also, to distinguish between different outcomes of homologous recombination. We identify several types of gene targeting event: (i) ‘one-end gene targeting’ that may result from an homologous recombination event at one end of the construct accompanied by a non-homologous end-joining event at the other (HR/NHEJ), (ii) true allele replacement that occurs by two homologous recombination events (HR/HR): this may involve insertion of multiple copies of the construct (NHEJ/NHEJ), no gene-specific PCR product was evidently derived from an ectopic insertion of the targeting construct, accompanied by NHEJ at the 5′ end. Where a double-recombination event has created an allele replacement (HR/HR), both 5′- and 3′-specific PCR products are generated (transformants [2], [5], [6], [9], [10], [13]). In this analysis, transformants [1] and [12] evidently derived from an ectopic insertion of the targeting construct (NHEJ/NHEJ), no gene-specific PCR product was being generated. Such negative results do not represent a failure of the PCR, since amplification using the 5′- and 3′-external primers results in amplification of the wild-type locus (as seen in Figure 4A).

### Analyses of the relationship between homologous recombination and homology length

Two methods were used to generate the transforming constructs: PCR or restriction enzyme digestion. Those constructs generated by PCR are homologous with their cognate genomic sequences throughout their length, whereas those constructs generated by NotI digestion of clones in pMBL6attR have 40 bp of terminal sequence that has no genomic counterpart, corresponding to the vector-derived attB sites (indicated by asterisk in Table 1).

To examine the relationship between the frequency of homologous recombination and length of homologous DNA sequence we have carried out analysis of co-variance and regression analysis on the the 1670 plants transformed with PpRac-L and PpPum-I constructs (22). Constructs with

### Table 1. Summary of gene targeting experiments

<table>
<thead>
<tr>
<th>Gene</th>
<th>Construct code</th>
<th>Homology (bp)</th>
<th>No. of targeting events (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5′</td>
<td>3′</td>
<td>5′ end</td>
</tr>
<tr>
<td>PpRac-L</td>
<td>KE</td>
<td>242</td>
<td>343</td>
<td>1 (1.1)</td>
</tr>
<tr>
<td></td>
<td>KG</td>
<td>1119</td>
<td>2562</td>
<td>5 (4.3)</td>
</tr>
<tr>
<td></td>
<td>KSS</td>
<td>2149</td>
<td>908</td>
<td>16 (25.8)</td>
</tr>
<tr>
<td></td>
<td>KSSF*</td>
<td>940</td>
<td>908</td>
<td>1 (1.8)</td>
</tr>
<tr>
<td></td>
<td>KKh</td>
<td>1119</td>
<td>526</td>
<td>65 (32.3)</td>
</tr>
<tr>
<td></td>
<td>KB</td>
<td>293</td>
<td>988</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>KBG</td>
<td>293</td>
<td>526</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>KH</td>
<td>3155</td>
<td>526</td>
<td>16 (27.1)</td>
</tr>
<tr>
<td>PpPum-I</td>
<td>KB</td>
<td>2547</td>
<td>840</td>
<td>8 (21.1)</td>
</tr>
<tr>
<td></td>
<td>KE</td>
<td>1647</td>
<td>1740</td>
<td>13 (19.7)</td>
</tr>
<tr>
<td></td>
<td>KB</td>
<td>1097</td>
<td>1200</td>
<td>24 (25.3)</td>
</tr>
<tr>
<td></td>
<td>KB-S1*</td>
<td>531</td>
<td>422</td>
<td>24 (20.2)</td>
</tr>
<tr>
<td></td>
<td>KB-S2*</td>
<td>531</td>
<td>199</td>
<td>15 (19.5)</td>
</tr>
<tr>
<td></td>
<td>KB-S3*</td>
<td>277</td>
<td>422</td>
<td>13 (20.3)</td>
</tr>
<tr>
<td></td>
<td>KB-S4*</td>
<td>277</td>
<td>199</td>
<td>3 (4.2)</td>
</tr>
<tr>
<td></td>
<td>KB-S5*</td>
<td>165</td>
<td>422</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>KB-S6*</td>
<td>165</td>
<td>199</td>
<td>3 (4.1)</td>
</tr>
<tr>
<td></td>
<td>KB-S7*</td>
<td>58</td>
<td>422</td>
<td>2 (2.6)</td>
</tr>
<tr>
<td></td>
<td>KB-S8*</td>
<td>58</td>
<td>199</td>
<td>0</td>
</tr>
<tr>
<td>PpMag-1</td>
<td>Mag*</td>
<td>870</td>
<td>540</td>
<td>6 (15.0)</td>
</tr>
<tr>
<td>PpWd-1</td>
<td>YWD</td>
<td>1216</td>
<td>1534</td>
<td>12 (9.4)</td>
</tr>
<tr>
<td>PpLea-1</td>
<td>Lea*</td>
<td>932</td>
<td>495</td>
<td>4 (8.9)</td>
</tr>
</tbody>
</table>

The table indicates the lengths of the homologous sequences in each arm of the targeting construct, the number of homologous recombination events occurring in the 5′-arm and the 3′-arm alone, as well as the number of double-recombination (‘replacement’) and untargeted transformation events. The number of targeting events as a percentage of total transformants is given in brackets (%). Constructs indicated with an asterisk represent fragments generated by PCR that were terminally homologous with their genomic targets. All other fragments used for transformation were generated by NotI digestion of cloned sequences and contained 40 bp terminal non-homology at either end derived from the AttB sites in the cloning vector.
homologous termini give a better fit for a linear relationship between targeting frequency and homology length, than for a logarithmic relationship (F-value for linear regression: 59.4; for logarithmic: 26.1. The higher the F-value, the better the data are fitted to the regression). By contrast, for constructs with non-homologous termini, the data fit better to a logarithmic relationship (F-value for linear regression: 14.5; for logarithmic: 42.4) (see Figure 3A and B). Additionally, the frequencies of recombination observed for the 5' and 3' sequences of the single constructs derived from PpWD-1, PpMago and PpLea-I fell within the 95% confidence limits of these regressions.

Analysis of co-variance requires that the dependent variable and the covariate (in this case length of homology) have a linear relationship (22). For the analysis of co-variance of the data for constructs having homologous termini (Table 2, homologous termini), the relationship between length of homology and frequency of recombination is linear so no mathematical transformation of the length was necessary. For the analysis of the constructs where the termini were non-homologous, the relationship between length and frequency is log/linear and so the logarithm of the length has been used for the analysis of co-variance (Table 2, non-homologous termini).

Results of the two analyses show (i) that there is a significant dependency of the frequency of homologous recombination on length of homology, (ii) that the location of the homologous sequence, 5' or 3' to the selection cassette does not affect this relationship and (iii) there is no significant difference in the behaviour of the two loci. (iv) The interaction term (locus by location) is also not significant, confirming that the effect of location is the same for both loci (PpRac-1 and PpPum-1).

Given the dependency of recombination on sequence length, it is to be expected that, where asymmetric targeting constructs are used, the frequency with which allele replacement (HR/HR) occurs will be critically related to the length of the shorter arm of the construct. Regression analysis confirms this. With constructs whose homology extends to the termini, a 50% allele replacement frequency is achieved with a shorter homology length of ~400 bp. Where the terminal sequences of the construct are non-homologous, longer lengths of homologous flanking sequence (~600 bp) are required to achieve a 50% allele replacement frequency (Figure 4C).

Gene targeting, copy number and ectopic insertion

Transformation of plant protoplasts by direct DNA uptake is commonly associated with the insertion into the genome of multiple copies of the transforming DNA, often in the form of long tandem repeats. Insertion at multiple loci has also been observed to occur (25). We have analysed the outcome of transformation in terms of both copy number and insertion site. In order to determine the frequency with which allele replacement occurs by a single transgene insertion (TGR), transgenic lines in which allele replacement was identified were subjected to PCR analysis using the external primer pairs: single-copy replacement events were identified by the amplification of a fragment corresponding to the transforming cassette (Figure 4A). Because of the problems inherent in amplifying longer fragments, we focused on loci targeted by constructs having an overall length of no more than 5 kb. Three loci were examined: PpRac-1, PpWD-1 and PpLEA-I, and a total of 312 transgenic plants were analysed. Forty single-copy replacement lines were identified, of which 35 were then analysed by Southern blot hybridization, to determine whether additional copies of the transgene had been inserted elsewhere in the genome. Figure 4B shows the results for the PpWD-1 locus. The results of this analysis, for all three loci, are summarized in Table 3, and demonstrate that TGR events occur as a high proportion (70–85%) of the single-copy allele replacements, and with an overall frequency of 7–20% of the population of stable transfectants.

DISCUSSION

TGR by homologous recombination provides a uniquely powerful tool for the sophisticated manipulation of genomes. Among eukaryotes, this approach has been most dramatically exploited in the manipulation of the yeast, Saccharomyces cerevisiae, in which organism transforming DNA is
preferentially integrated at homologous sites (26). However, the relationship of the mechanisms responsible for TGR in *S.cerevisiae* to other processes involving HR, remains unclear (27).

Short lengths of sequence homology (30–45 bp) are sufficient to potentiate homologous recombination-mediated transformation in yeast (28), although greater lengths of homology are required to target some genes (29). Other simple eukaryotes, including filamentous fungi (30) and single-celled protozoan parasites such as *Trypanosoma brucei* (31) and *Leishmania* (32,33) also exhibit relatively high rates of HR-mediated TGR when transformed with constructs containing relatively short lengths (~1 kb) of sequence homology.

Within the plant kingdom, there has been very little success in achieving TGR. Despite optimistic reports, most observations of gene targeting in the model angiosperm *Arabidopsis thaliana* had been incidental (5,34–36) and a series of investigations into the occurrence of intrachromosomal recombination in flowering plants indicate that, although this phenomenon can be observed to occur, its frequency is low and sporadically distributed among plant organ types (36,37). Similarly, although it has been reported that TGR could be achieved in rice, through a stringent positive–negative selection procedure (38), the absolute rate of TGR obtained was of the low order of magnitude seen in other plant systems, and this apparent success has yet to be exploited.

A striking exception to this generally negative experience is provided by the moss, *P.patens*, in which targeted integration of transforming DNA was first conclusively demonstrated by Schaefer and Zryď (8), and which has since been increasingly exploited as a model organism in which reverse genetic manipulation by TGR can be undertaken with relative ease (10,39,40).

Among other complex eukaryotes, most effort has been directed toward the manipulation of mouse embryo-derived stem (ES) cells, with their high potential for regeneration as transgenic mice (41). TGR occurs at low frequencies in these cells and requires longer homologous sequences than are observed in fungi and protists. The maximum rate of homologous recombination in mouse ES cells is reported to be about 2 · 10⁻⁴ achieved with ~14 kb of homologous sequence (42,43).

In *S.cerevisiae*, the frequency of homologous recombination is much higher but varies between loci, the GC content of the homologous sequence being a critical factor (29). There

![Figure 3.](image-url)

**Figure 3.** Relationship between recombination frequency and homology length. (A) Relationship between recombination frequency and homology length: terminal homology. The rate of recombination occurring between either the 5'-arm (black diamonds) or 3'-arm (black triangles) of each terminally-homologous PCR-amplified targeting construct and the target locus is plotted, as a percentage, against the length of the targeting sequence. The regression line is for frequency on sequence length. (B) Relationship between targeting frequency and homology length: terminal non-homology. The rate of recombination occurring between either the 5'-arm (black diamonds) or 3'-arm (black triangles) of each restriction enzyme-digested targeting construct (with 40 bp of terminal non-homology) and the target locus is plotted, as a percentage, against the length of the targeting sequence. The regression line is for frequency on the logarithm of sequence length. (C) Relationship between allele replacement and shorter length of homology. The rate of recombination events occurring in both arms of a targeting construct and hence to allele replacement at the targeted locus, is plotted, for all constructs at all loci, as a percentage, against the length of the shorter homologous sequence. The regression line is for frequency on sequence length, for constructs homologous throughout their length (solid line and black triangles), and on the logarithm of sequence length for constructs with terminal non-homology (dashed line and black squares).
are no data from *Saccharomyces* studies that are strictly comparable to the data for *Physcomitrella* presented here, but for most yeast loci, over 200 bp of homology are required to generate a targeting frequency of 50%. The frequency of gene targeting in *Physcomitrella* and the lengths of homologous sequence required are therefore similar to those in *Saccharomyces*.

In *S.cerevisiae* lengths of terminal non-homology as short as 9 bp can reduce TGR at some target loci (44). Some of the constructs used in this study had non-homologous terminal regions of 40 bp while others were terminally homologous (see Table 1). The frequencies of HR by constructs having homologous termini (Figure 3A) are higher than those having non-homologous ends, and for these, the dependency of the frequency of recombination on length of homology fits best a linear relationship. For constructs with non-homologous termini, the relationship between length of homology and frequency of HR is logarithmic. We suggest that terminal non-homology reduces the efficiency of homology searching by the invading DNA strand. In wild-type *S.cerevisiae* strains, non-homologous terminal sequences must be removed from transforming DNA before targeted integration. This requires both the RAD1/RAD10 nucleotide excision nuclease and the MSH2/MSH3 mismatch repair proteins (45). It may be that a comparable system exists in *P.paten*s and that the need for this to intervene before HR can occur accounts for the different relationship. This is an area where more research, specifically designed to investigate the effects of terminal non-homology, will be fruitful.

The frequency of TGR in *Saccharomyces* was observed to be about 2.5 times higher in a construct having flanking sequences of 200 and 50 bp of homologous DNA, compared with a construct having only 50 bp on either side of the selection cassette (29). TGR requires HR to occur in both flanking sequences, and so the presence of a longer sequence on one side of the cassette increases the rate of HR in the sequence on the other side, presumably by stabilising the recombination complex. We observe a similar but less marked interaction between asymmetric flanking sequences (e.g. compare Table 1, KH-S1, S3, S5 and S7 or KH-S2, S4, S6 and S8).

In an early comparison of the earliest gene targeting experiments undertaken in *Physcomitrella*, Schaefer (7) observed that gene targeting frequencies over a range of 4–100% of stable transformants had been obtained, with an indication that targeting constructs based on cDNA were less efficiently targeted than those in which the targeting sequences comprised contiguous lengths of genomic DNA, and that frequencies obtained with shorter constructs were generally less efficiently targeted. We have now shown that allele replacement frequencies of 50% are routinely obtained if the selection cassette is flanked by homologous sequences no more 400 bp in length on either side: a length substantially shorter than is required for gene targeting in vertebrate cells. However, we also observe that although homology length is a major determinant of targeting efficiency, the symmetry of targeting constructs is equally important in determining the outcome of transformation. Thus, highly asymmetric constructs are integrated preferentially with a single homologous recombination event within the longer arm of the construct, accompanied by a non-homologous end joining event by the shorter arm to integrate the construct into the genome. Such NHEJ events are frequently accompanied by rearrangements at the site of integration (Y. Kamisugi, S. Rensing, A. C. Cuming, R. Reski and D. J. Cove, manuscript in preparation) and in these respects the integration of highly asymmetric constructs appears similar to that observed in mouse ES cells (46), where a highly asymmetric construct was both less efficiently targeted and underwent rearrangement in the shorter arm.

The mechanisms that regulate homologous recombination in plants are not well understood. This is in part due to the general incompetence to undertake homologous recombination exhibited by the flowering plant models that have predominantly been studied (36). The determination of the parameters required for efficient gene targeting in *Physcomitrella* offers the opportunity to illuminate this

### Table 2. Analyses of co-variance of the dependency of rate of recombination on length of homologous sequence

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum-of-squares</th>
<th>Degrees of freedom</th>
<th>Mean-square</th>
<th>F-ratio</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homologous termini&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus (Rac or Pum)</td>
<td>0.084</td>
<td>1</td>
<td>0.084</td>
<td>2.29</td>
<td>0.154</td>
</tr>
<tr>
<td>Location (5’ or 3’)</td>
<td>0.002</td>
<td>1</td>
<td>0.002</td>
<td>0.061</td>
<td>0.808</td>
</tr>
<tr>
<td>Locus x location</td>
<td>0.112</td>
<td>1</td>
<td>0.112</td>
<td>3.04</td>
<td>0.105</td>
</tr>
<tr>
<td>Length of homology</td>
<td>0.408</td>
<td>1</td>
<td>0.408</td>
<td>11.05</td>
<td>0.005</td>
</tr>
<tr>
<td>Error</td>
<td>0.480</td>
<td>13</td>
<td>0.037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-homologous termini&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus (Rac or Pum)</td>
<td>0.041</td>
<td>1</td>
<td>0.041</td>
<td>1.56</td>
<td>0.229</td>
</tr>
<tr>
<td>Location (5’ or 3’)</td>
<td>0.000</td>
<td>1</td>
<td>0.000</td>
<td>0.01</td>
<td>0.977</td>
</tr>
<tr>
<td>Locus x location</td>
<td>0.003</td>
<td>1</td>
<td>0.003</td>
<td>0.10</td>
<td>0.755</td>
</tr>
<tr>
<td>Length of homology</td>
<td>0.913</td>
<td>1</td>
<td>0.913</td>
<td>34.86</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Error</td>
<td>0.445</td>
<td>17</td>
<td>0.026</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Analysis of the sequence length and location of homologous sequence in relation to selection cassette (5’ versus 3’) and of locus (Rac versus Pum) on arsclin square root of recombination rate.

<sup>b</sup>Analysis of the logarithm of sequence length and location of homologous sequence in relation to selection cassette (5’ versus 3’) and of locus (Rac versus Pum) on arsclin square root of recombination rate.

### Table 3. Frequency of TGR

<table>
<thead>
<tr>
<th>Locus</th>
<th>Transformants [n]</th>
<th>HR/HR [%]</th>
<th>Single-copy HR/HR [%]</th>
<th>Single-copy HR/HR no ectopic insertion [%]</th>
<th>TGR [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>PpRac-1</em> 1.1 + 0.5 kb</td>
<td>130</td>
<td>56 (43.1)</td>
<td>14/56 (25)</td>
<td>7/10 (70)</td>
<td>7.5</td>
</tr>
<tr>
<td><em>PpWD-1</em> 1.2 + 1.6 kb</td>
<td>127</td>
<td>53 (41.7)</td>
<td>14/53 (26.4)</td>
<td>11/13 (84.6)</td>
<td>9.3</td>
</tr>
<tr>
<td><em>PpLEA-1</em> 1 + 0.5 kb</td>
<td>45</td>
<td>22 (48.8)</td>
<td>12/22 (54.5)</td>
<td>9/12 (75)</td>
<td>20</td>
</tr>
</tbody>
</table>

This summarizes the analysis of three gene targeting experiments using a single targeting construct at each of three loci. The targeting constructs used were *PpRac-1 ’RacKXH’, PpWD-1 ’YWD’* and *PpLEA-1 ’Lea’, (Figure 1). The values for single-copy and single-insertion events as a percentage of the total number of stable transformants for each locus is corrected, for the *PpRac-1* and *PpWD-1* loci, to allow for the fact that not all the single-copy replacement lines were tested by Southern blot analysis. The assumption made is that 70 and 84.6, respectively, of all the lines targeted by HR/HR would be single-insertion transformants.
enigmatic area of plant DNA metabolism. In animal cells, as in yeast, genes comprising the Rad52 epistasis group, which regulate double-strand breakage (DSB), strand invasion and heteroduplex formation at recombination sites have been clearly implicated by observing the consequences of mutations in these genes for gene targeting (47), as have genes involved in mismatch repair (48–50). In both yeast and mammalian cells, HR mediated DSB repair is catalysed by proteins of the Rad52 epistasis group (Rad51, 52, 54, 55, 57 and the MRN complex) which use an intact copy of the damaged region as a template for repair. HR is initiated by the formation of long 3' single stranded tails at the DSB which are used to locate and invade an homologous sequence in the genome. DNA synthesis is primed by the invading strand associated with proteins of the Rad52 epistasis group (51,52). In plant genomes sequenced to date, no homologue of the key Rad52 gene has been identified, leading to speculation that this might be a reason for the low efficiencies of homologous recombination obtained. However, other higher-plant recombination-associated genes have been identified (53,54) and very recently it was reported that overexpression of a yeast Rad54 transgene in Arabidopsis results in a substantial enhancement of gene targeting efficiency in this previously intractable model (6). Physcomitrella homologues of some members of the Rad52 epistasis group have been identified, including Rad51, by cloning and functional analysis (55,56), and Rad54 and RPA complex homologues can be found by database searches. It is therefore possible that other genes functionally equivalent to those identified in yeast and mammalian cells exist, but that they are insufficiently similar to be detected using standard sequence homology searches. With the sequencing of the Physcomitrella genome ongoing, the opportunity to identify recombination-associated genes will be markedly enhanced, and should provide the capability to utilize the technology of homologous recombination to disrupt or otherwise alter the Physcomitrella homologues of these genes, in situ, and examine the consequences for the efficiency of homologous recombination-mediated integration of transforming DNA.

**Figure 4.** Targeted versus ectopic insertions. (A) Identification of single-copy replacement events. This illustrates single-copy replacement events in the PpWD-1 gene, identified by PCR using gene-specific primers external to the targeting construct. Amplification of the wild-type locus (w.t.) generates a fragment of ~4.2 kb, whereas replacement by a single-copy of the targeting construct containing the nptII cassette cloned between XhoI (Xh) and ScaI (Sc) sites results in the amplification of an ~5 kb fragment. When additional copies are inserted at the PpWD-1 locus, the length of the sequence between the external primers is too long to be amplified under standard experimental conditions. Tracks 1–10 correspond to ten independent transformants. 'M' corresponds to a HindIII digestion of bacteriophage lambda DNA. (B) Southern blot detection of ectopic insertions. TGR at the PpWD-1 gene, among single-copy replacement transformants, was confirmed by Southern blot analysis using a hybridization probe containing sequences found in both arms of the targeting construct (hatched bars corresponding to 1.2 kb and 1.6 kb fragments). Digestion with HindIII generates two fragments of 5 kb and 3.5 kb from the wild-type PpWD-1 locus (w.t.). A single-copy replacement event results in the appearance of a 4.2 kb band, derived from the 3' arm of the transforming construct, generated by the insertion of the selection cassette between the XbaI (X) and SmaI (S) sites, 0.1 kb from the HindIII site (H) internal to the gene. In all but two of the transformants analysed, no other hybridizing fragments are observed. In transformants [6] and [7], a complex ladder of additional hybridizing fragments is observed, indicating multiple ectopic insertions and rearrangements (these tracks are also shown following a shorter exposure of the X-ray film).
SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by the European Union Framework V as a part of the ‘Pregene’ network (D.J.C. and Y.K.: Contract QRLT 2000-00365), and through the BBSRC supported ‘PEP’ programme (A.C.C. Grant No. 24/P11357). D.J.C. gratefully acknowledges the award of a Leverhulme Trust Emeritus Fellowship. We are grateful to Dr Mike Deeks and Dr Brendan Davies for providing the Ppmago gene used in this study and to Dr Peter Van Zandt and Dr Stuart McDaniel for advice on the statistical analyses. Funding to pay the Open Access publication charges for this article was provided by JISC.

Conflict of interest statement. None declared.

REFERENCES


