GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

A transient assay for recombination demonstrates that *Arabidopsis SNM1* and *XRCC3* enhance non-homologous recombination

R.A. Johnson[1,2], R.P. Hellens[1] and D.R. Love[2,3]

[1]The New Zealand Institute for Plant & Food Research Limited, Sandringham, Auckland, New Zealand
[2]School of Biological Sciences, University of Auckland, Auckland, New Zealand
[3]LabPlus, Diagnostic Genetics, Auckland Hospital, Auckland, New Zealand

Part of a PhD thesis presented by R.A. Johnson to the University of Auckland
Corresponding author: R.A. Johnson
E-mail: rjoh064@aucklanduni.ac.nz

Genet. Mol. Res. 10 (3): 2104-2132 (2011)
Received February 28, 2011
Accepted August 29, 2011
Published September 16, 2011
DOI http://dx.doi.org/10.4238/vol10-3gmr1347

ABSTRACT. Replacement of endogenous genes by homologous recombination is rare in plants; the majority of genetic modifications are the result of transforming DNA molecules undergoing random genomic insertion by way of non-homologous recombination. Factors that affect chromatin remodeling and DNA repair are thought to have the potential to enhance the frequency of homologous recombination in plants. Conventional tools to study the frequencies of genetic recombination often rely on stable transformation-based approaches, with these systems being rarely capable of high-throughput or combinatorial analysis. We developed a series of vectors that use chemiluminescent (*LUC* and *REN*) reporter genes to assay the relative frequency of homologous and non-homologous recombination in plants. These transient assay vectors were used to screen 14 candidate

genes for their effects on recombination frequencies in *Nicotiana benthamiana* plants. Over-expression of *Arabidopsis* genes with sequence similarity to *SNM1* from yeast and *XRCC3* from humans enhanced the frequency of non-homologous recombination when assayed using two different donor vectors. Transient *N. benthamiana* leaf systems were also used in an alternative assay for preliminary measurements of homologous recombination frequencies, which were found to be enhanced by over-expression of *RAD52*, *MIM* and *RAD51* from yeast, as well as *CHR24* from *Arabidopsis*. The findings for the assays described here are in line with previous studies that analyzed recombination frequencies using stable transformation. The assays we report have revealed functions in non-homologous recombination for the *Arabidopsis SNM1* and *XRCC3* genes, so the suppression of these genes' expression offers a potential means to enhance the gene targeting frequency in plants. Furthermore, our findings also indicate that plant gene targeting frequencies could be enhanced by over-expression of *RAD52*, *MIM*, *CHR24*, and *RAD51* genes.

Key words: Extra-chromosomal recombination; *Agro*-infiltration; Non-homologous recombination; Homologous recombination; Plant genetic modification; Transient Dual-Luciferase® assays

INTRODUCTION

The stable genetic modification of plants is based on DNA inserting essentially at random into the genome (Kim et al., 2007). A long-sought aim of plant genetic engineering is to regenerate plants by gene targeting (GT); the directed alteration of a chromosomal target by exchanging it with a homologous construct, termed homologous recombination (HR) (Britt and May, 2003). GT can potentially be used as a tool to substitute or disrupt endogenous genes and/or their *cis*-acting elements with candidate genetic material for functional analysis or directed breeding purposes. GT also avoids the disruption of non-target genomic elements that can occur during non-homologous recombination (NHR) (Iida and Terada, 2005). The endogenous frequencies of GT were low in plants when selectable marker genes were used for positive selection of GT, or when a 'positive-negative selection' strategy that involves counter-selection against cells integrating whole T-DNAs by NHR was used. These GT strategies are similar to those that are routinely used in vertebrates but can sometimes fail to detect GT events in plants. Alternative strategies that have targeted either chromosomally inserted genes for replacement, or endogenous genes that, when disrupted, lead to acquired herbicide-resistance, have also reported low frequencies of GT (Iida and Terada, 2005). A 27-fold enhancement in the GT frequency was observed when *RAD54*, a gene encoding a factor that enhances DNA interactions by remodeling chromatin-bound histone proteins as part of the endogenous DNA repair mechanism in yeast, was over-expressed in *Arabidopsis thaliana* (Shaked et al., 2005). This previous report has critically highlighted the potential for enhancing the GT frequency by manipulating the recombination machinery in plants. In addition to *RAD54*, there are a number of

candidate 'effector' genes that have described or predicted roles in HR, and as such, these genes are appropriate targets for manipulation in order to enhance the GT frequency (Hanin et al., 2000; Molinier et al., 2004; Durrant et al., 2007). Furthermore, a large number of additional candidate genes have not yet been characterized *in planta* and rarely has the simultaneous action of two genes on recombination frequencies been considered in the same study (Schuermann et al., 2005).

Reporter gene systems, such as defective β-*glucuronidase* (*GUS*) genes that can be restored by HR to form a functional gene, have been used to quantify HR events through detecting the instances of GUS activity. These defective genes have been used as extra-chromosomal molecules (Shalev et al., 1999), or within a chromosomally integrated T-DNA molecule, and the recombination substrates in either inverted or direct repeats (Li et al., 2004). Rather than using stable transgenic plant-based methods for screening candidate effector genes, here we describe rapid extra-chromosomal T-DNA systems to assay for recombination, which are based on previously reported recombination assays (Lyznik et al., 1991; Hrouda and Paszkowski, 1994), with this study testing the influence of candidate genes. The assays for recombination use infiltration of *Agrobacterium tumefaciens* into the leaves of *Nicotiana benthamiana* (Hellens et al., 2005). The bacteria harbor vectors, either to measure the levels of recombination using the *Luciferase* (*LUC*) reporter gene, to normalize against the levels of *Agro*-infiltration using the *Renilla Luciferase* (*REN*) reporter gene, or to express genes that may affect recombination frequencies. The assay for HR involves a donor T-DNA containing an *LUC* gene that was fused to an upstream sequence derived from an intron of the e*longation factor 1-alpha* (*EF-1α*) gene in tobacco (Ursin et al., 1991); a second complementary portion of this intron was also included such that HR occurring between the two complementary intron sequence repeats would reconstitute a functional fusion of the *LUC* gene to a 35S promoter. These intron sequences were orientated in inverted repeats as incidences of HR have been more reliably detected using these DNA substrates when compared with constructs that had direct repeats (Li et al., 2004); such non-coding sequences have been reported to be more suitable targets for HR than coding sequences (Hrouda and Paszkowski, 1994; Ilnytskyy et al., 2004; Li et al., 2004). Similarly to these assays for HR, measurements of NHR were made using a previously reported assay (Gorbunova et al., 2000) that has been adapted to rapidly quantify chemiluminescent reporter genes with high sensitivity (Ilnytskyy et al., 2004; Hellens et al., 2005).

Our screen identified two genes from *Arabidopsis*: *sensitive to nitrogen mustard 1* (*SNM1*) and *X-ray repair cross-complementing 3* (*XRCC3*), which, upon over-expression in our transient expression assays, appear to suppress the relative measurements of HR:NHR. The *SNM1* gene from yeast encodes an enzyme that appears to have a primary function as a 5' exonuclease (Li et al., 2005). This gene appears to have three gene orthologues in plants and mammals; the *Arabidopsis SNM1* gene was the only of these genes that was tested for ability to enhance HR in this study due to a reported reduction of intra-chromosomal HR in *snm1* mutant plants (Molinier et al., 2004). The *XRCC3* gene was first identified as a cDNA in humans that could complement the defective DNA repair (Tebbs et al., 1995) in mutant 'irs1SF' Chinese hamster cells (Fuller and Painter, 1988). *XRCC3* is an orthologue of *RAD57* in yeast (Schuermann et al., 2005), and has been described as having roles in vertebrates ranging from early localization to double-stranded DNA break sites (Forget et al., 2004) through stabilizing heteroduplex DNA and resolving

HR intermediates (Brenneman et al., 2002).

Here we describe: i) The development of an independent vector for normalizing the level of *Agro*-infiltration, ii) The testing of Luciferase gene's position relative to the T-DNA borders, iii) The testing of an upstream intron sequence's impact on reporter gene expression, iv) The assessment of 14 candidate genes for their individual impact on homologous frequencies, v) The further testing of eight candidate genes for their impact on non-homologous recombination frequencies, and vi) The validation of the impact from *AtSNM1* and *AtXRCC3* gene over-expression using an alternative assay for non-homologous recombination.

MATERIAL AND METHODS

Isolation of an intron sequence for use in assays of HR

Genomic DNA was obtained from young leaves of *N. tabacum* 'Samsun' using the DNeasy Plant Mini Kit (Qiagen-Biolab Ltd., New Zealand). The *NtEF-1α* (TA2602) (Childs et al., 2007) 5' sequence was obtained using a degenerate polymerase chain reaction (PCR) approach according to manufacturer instructions. This approach used EST sequences from Genbank Nos. D63396 and U04632. PCR amplification used Platinum Taq DNA Polymerase (Invitrogen™) and oligonucleotides RAJ-161 and RAJ-048 (oligonucleotides are shown in Additional File 1). The sequence of a 952-bp intron was obtained after cloning amplified products into pGem-T easy (Promega) according to manufacturer instructions. Molecular biology techniques were performed according to standard established protocols. Specifically, ligations used Rapid DNA Ligation kits (Roche, Global Science and Technology Ltd., Auckland, New Zealand), and were introduced into *Escherichia coli* DH5α cells (Invitrogen New Zealand Ltd., Auckland, New Zealand) by transformation using a heat shock treatment, and a subsequent recovery procedure that were both done according to manufacturer instructions. The transformation mixture transferred onto plates containing Luria Bertani media (Invitrogen) was supplemented with an appropriate antibiotic; selected colonies were cultured in the liquid equivalent of this media (Invitrogen) at 37°C with agitation at 225 revolutions per minute. Plasmid DNA was purified using a QIAprep Spin Miniprep Kit (Qiagen-Biolab Ltd.). Putative clones were sequenced using BigDye Terminator v3.1 sequencing chemistry (Allan Wilson Centre, Massey University, Palmerston North, New Zealand).

Development of effector constructs

The coding sequences of genes that are involved in recombination were amplified from the templates shown in Table 1, using *Pwo* DNA polymerase (Roche), unless otherwise specified, and the oligonucleotides shown in Additional File 1. Candidate genes for over-expression were cloned into the pHEX2 vector using Gateway® recombination technology (Invitrogen), which uses a 35S promoter to drive expression of this effector gene (Hellens et al., 2005). The constructs were sequence-verified, and then introduced by electroporation into *A. tumefaciens* strain *GV3101* (MP90) and/or strain *LBA4404*, for use in transient plant transformation. The plasmid pENTR-*GUS* (Invitrogen) was used to develop an over-expression construct in pHEX2 to serve as a negative control.

Table 1. Genes amplified for testing as effectors of recombination.

Gene [Reference(s)]	Template
ScSPO11 (Schuermann et al., 2005) (*sporulation-specific protein 11*)	*S. cerevisiae* S288C Genomic DNA, Locus: YSCSPO11 [Genbank: J02987]
*AtSNM1*ψ (Molinier et al., 2004) (*A. thaliana sensitive to nitrogen mustard 1*)	ABRC Clone U50597, TAIR Locus: At3g26680 [Genbank: AK117422]
ScRAD52 (Di et al., 2005) (*radiation sensitive 52*)	*S. cerevisiae* S288C Genomic DNA, Locus: YSCRAD52 [Genbank: M10249.1]
ScRAD54 (Shaked et al., 2005) (*radiation sensitive 54*)	*S. cerevisiae* S288C Genomic DNA, Locus: YSCRAD54A [Genbank: M63232]
*AtCHR24*ψ (Shaked et al., 2006) (*A. thaliana chromatin remodeling 24*)	ABRC Clone U21465, TAIR Locus: At5g63950 [Genbank: AK117422]
ScMMS4 (Hanin et al., 2000) (*hypersensitive to MMS, irradiation, and MMC*)	*S. cerevisiae* S288C Genomic DNA [Genbank: X80930]
ScINO80 (Shaked et al., 2006) (*inositol-requiring protein 80*)	*S. cerevisiae* S288C Genomic DNA, Locus: YGL150c [Genbank: Z72672]
*ScRAD51*ψ (Schuermann et al., 2005) (*radiation sensitive 51*)	*S. cerevisiae* S288C Genomic DNA [Genbank: X64270]
*AtRAD51C*ψ (Osakabe et al., 2002) (*A. thaliana radiation sensitive 51 paralogous gene C*)	RIKEN Clone pda11304, TAIR Locus: At2g45280 [Genbank: AK118275]
*AtRAD51D*ψ (Durrant et al., 2007) (*A. thaliana radiation sensitive 51 paralogous gene D*)	RIKEN Clone pda04728, TAIR Locus At1g07745 [Genbank: AY056153]
*AtXRCC3*ψ (Osakabe et al., 2002) (*A. thaliana X-ray repair cross-complementing 3*)	ABRC Clone U85183, TAIR Locus: At5g57450 [Genbank: NM180871]
AtBRCA2-IVψ♠ (Dray et al., 2006) (*A. thaliana breast cancer type 2 susceptibility protein* -IV)	*A. thaliana* ssp *columbia* Genomic DNA, TAIR Locus: AT4G00020 [Genbank: AJ488304]
AtDSS1-1ψ (Dray et al., 2006) (*A. thaliana deleted in split hand/split foot protein 1* -1)	ABRC Clone U11771, TAIR Locus: At1g64750 [Genbank: AY065373]
ScCCE1 (Shaked et al., 2006) (*cruciform cutting exonuclease 1*)	*S. cerevisiae* S288C Genomic DNA, Locus: YSCCCE1A [Genbank: M65275]

ψThe amplification also involved an additional secondary PCR with adapter primers. ♠*AtBRCA2-IV* was amplified using LA Taq Polymerase (TaKaRa™).

Transient Dual-Luciferase® Assays

Transient Dual-Luciferase® Assays were conducted in a modified format to a previous study by Hellens et al. (2005). *N. tabacum* 'Samsun' and *N. benthamiana* plants were grown in glasshouse conditions at 22°C with light supplemented to 16-h days. Plants were grown to the 6-10 leaf stage. Plants were infiltrated and incubated in the conditions described above throughout the course of the experiment. *A. tumefaciens* cells were cultured at 30°C on Luria Bertani media (Invitrogen) with kanamycin-selection (50 µg/mL) for strains carrying pGreenII and pSoup. For *A. tumefaciens* strain *GV3101* (MP90) carrying pHEX2, selection used gentamycin (10 µg/mL), rifampicin (25 µg/mL), and spectinomycin (50 µg/mL). *A. tumefaciens* strain *LBA4404* carrying pHEX2 vectors used selection with spectinomycin (100 µg/mL). Confluent cells were resuspended in infiltration media (10 mM $MgCl_2$, 1 µM acetosyringone) to an optical density (in light with a 600 nm wavelength) of 1.2 ± 0.05, and incubated at room temperature for 2 h without shaking to prepare *Agrobacteria* for T-DNA transfer. A 7-day time-course was used for the assay of effector genes (Hellens et al., 2005) in order for LUC activity to accumulate after effector enzymes could have mediated recombination events. The *N. benthamiana* plants used in these assays appeared to be appropriate for the 7-day period of assay used due to the reported longevity for transforming extra-chromosomal donor T-DNA molecules in a close relative of these plants, tobacco (*N. tabacum*), when compared with an alternative plant species, *A. thaliana* (Orel and Puchta, 2003). LUC and REN activity was assayed using Dual-Luciferase Reporter Assay reagents (Promega). Infiltrated leaf tissue was harvested after the stated incubation period, aggregated between sample runs, and ground in 100 µL passive lysis buffer. Leaf debris were pelleted by centrifugation for 5 s at 9300 rpm, and 1/20 dilution of the clarified extract was then made using 4 µL supernatant used in each sample run. Chemiluminescence measurements of LUC were made after the addition of 40 µL Luciferase Assay Buffer II. Measurements of REN were made following the addition of 40 µL Stop and Glow to the same sample run. Absolute RLU were measured in a Turner 20/20 luminometer, with a 5-s delay and a 10-s measurement. Data of LUC and REN values were collected individually, used to perform regression analysis verification, and then converted to a ratio. Typically, three leaves were used per plant. In line with previous use of these transient expression assays, the trends that have been presented here were checked, and found to be reproducible using a repeat experiment; this step was carried out to control for any variation that could have arisen in the data from day-to-day, plant-to-plant, or leaf-position effects (Hellens et al., 2005). Differences in the LUC:REN ratios that were observed between samples were only considered to be statistically significant if the P value that was obtained from the two-tailed Student *t*-test (with an assumption unequal variance) was less than 0.05. Typically, LUC:REN ratios that were measured for an effector gene as higher than the measurement for the *GUS* gene control also had higher LUC values than this control infiltration, despite comparable REN values. Plasmid DNA was extracted from each of the *A. tumefaciens* strains that were used in these assays, re-transformed into *E. coli* cells and then analyzed by restriction enzyme digestion to verify the integrity of all plasmids.

Luciferase reporter vectors

The Luciferase promoter traps, pGreenII RB-*LUC*-Term and pGreenII LB-*LUC*-

Term, were constructed to assay for NHR in plant genomes. The *LUC* gene in fusion with a cauliflower mosaic virus (CaMV)-derived Terminator was amplified from pGreenII 1598-6 using *Pwo* DNA polymerase (Roche) and primers RPH-063 and RPH-059. The subsequent *LUC*-Term amplification product was introduced into *Hpa*I and *Stu*I double-digested pGreenII 0000. All clones that were developed underwent screening using with predictive restriction endonuclease digestion of extracted DNA; sequencing was also used to verify the clone in this instance. The two different orientations of the insert were identified by sequencing for their subsequent use in transient assays. After confirmation, pGreenII plasmids and pSoup, *REN* were introduced into *A. tumefaciens* strains *GV3101* (MP90) and *LBA4404* by electroporation.

The pGreenII *LUC*-Term-*NOS* (nopaline synthase gene-derived promoter) vector was developed as a substrate for the assay of NHR between T-DNA ends in plants. The vector pGreenII 0049 was digested with *Not*I and *Psp*OMI, religated, and then amplified with *Pwo* Superyield DNA polymerase (Roche) and the primers, RAJ-360 and RAJ-361, that carried *Not*I recognition sites. The subsequent *LUC*-Term-*NOS* amplification product was then cloned into the same sites of pGreenII 0000, after its sequential digestion with *Sac*II-*Stu*I and *Hpa*I-*Spe*I endonucleases and re-ligation; the clones were sequence-verified.

The plasmid pGreenII:*LUC* (pGreenII 0579-1) was constructed so that a firefly-derived Luciferase gene (*LUC*) could be over-expressed in plants. An *Asp*718I-*Xba*I fragment was removed from pGreenII 1598-6, and T4 DNA polymerase was used to blunt the plasmid's ends before proceeding to self-ligation.

The *NtEF-1α* (TA2602) locus was amplified using oligonucleotides RAJ-200 and RAJ-190 to introduce *Not*I sites in the 5' untranslated region (5' UTR) intron and downstream intergenic sequence. The resulting clone was sequenced, and then used as a template in an inverse PCR amplification with oligonucleotides RAJ-183 and RAJ-184, which excluded the *EF-1α* codon sequence; this amplification product was then used for development of the pGreenII *EF-LUC*-Term vector.

The pGreenII *EF-LUC*-Term vector was constructed to quantify the *NtEF-1α* (TA2602) 5' UTR sequence for promoter activity. A *LUC*-Term fragment (described above in the development of pGreenII RB-*LUC*-Term) was ligated into the gene-excluded *NtEF-1α* (TA2602) locus clone. An *EF-LUC*-Term fragment was amplified using RPH-063 and RAJ-178 with *Pwo* DNA polymerase (Roche). The subsequent amplification product was ligated into the *Not*I and *Hpa*I sites of pGreenII 0000, after the prior removal of an *Stu*I-*Sac*II fragment from this vector.

The pGreenII *EF-LUC*-Term-*EF*-35S vector was constructed to detect intra-molecular homologous recombination after delivery into plants. The pGreenII *EF-LUC*-Term vector was used as a template for an inverse amplification with *Pwo* Superyield DNA polymerase (Roche), and primers RAJ-370 and RAJ-371 carrying *Xma*I recognition sites. An *EF*-35S fragment was amplified from pHEX2 + *NtEF-1α* (TA2602) using Platinum Taq DNA Polymerase High Fidelity (Invitrogen™), with the primers RAJ-373 and RAJ-431 that carried *Xma*I recognition sites. This amplicon was cloned into *Xma*I-digested pGreenII *EF-LUC*-Term; the subsequent clone was sequence-verified.

The vector pHEX2 + *NtEF-1α* (TA2602) was developed as an intermediate step in constructing the pGreenII *EF-LUC*-Term-*EF*-35S vector. An *att*B site-introducing PCR amplification of pCR4-TOPO + *NtEF-1α* (TA2602) was conducted using LA Taq™

(TaKaRa™) with oligonucleotides RAJ-362 and RAJ-363. Using Gateway® recombination technology (Invitrogen), a standard BP recombination cloning was conducted, with the clone subsequently sequenced; the amplified locus was then sub-cloned into pHEX2 using an LR reaction.

The normalization vector, pSoup: *REN* (pSoup 0800), was constructed to over-express a *Renilla reniformis*-derived Luciferase gene (*REN*) cassette in plants. The T-DNA region from pGreenII 0800 was isolated as a *Bgl*II fragment and cloned into the *Bam*HI sites of pSoup 0000; the clones were then sequenced across the cloning site.

Further information regarding the plasmids has been reported elsewhere and can be found at www.pGreen.ac.uk.

RESULTS

Development of an independent vector for normalizing the level of *Agro*-infiltration

Previously, in experiments that have not been surveying recombination, the *LUC* and *REN* reporter genes have been used when they were contained within the same T-DNA (Hellens et al., 2005). This study used delivery of the *LUC* and *REN* reporter genes on separate T-DNA molecules by co-transfection using a single *Agrobacterium* strain; a pGreenII construct contained a recombination donor cassette, and a separate pSoup: *REN* (pSoup 0800) vector for normalization when used in co-transfection assays with pGreenII donor constructs. Positioning the *REN* gene expression cassette outside the recombination-assay donor was used to avoid instability in this pGreenII vector T-DNA. The separate delivery of *LUC* and *REN* gene-housing T-DNA molecules first required relocating the *REN* reporter gene to the pSoup plasmid and determining that this vector was capable of normalizing the levels of *Agro*-infiltration in subsequent assays for recombination-based LUC activity (Additional File 2).

We found that the LUC:REN was 0.11 ± 0.0026 (SE) when the *REN* gene was located on the pSoup normalization plasmid and co-transfected separately to the *LUC* gene-housing pGreenII plasmid. On the other hand, the LUC:REN was 0.033 ± 0.0019 (SE) when both the *LUC* and *REN* genes were *Agro*-infiltrated as components of one pGreenII plasmid. Despite this 3.2-fold difference in ratio, the absolute values of LUC and REN correlated well, suggesting the pSoup:*REN* (pSoup 0800) vector could be used to normalize the levels of LUC from constructs delivered using the pGreenII vector in assays of HR and NHR.

The testing of a Luciferase gene's position relative to the T-DNA borders

We examined the location of the *LUC* reporter gene within the T-DNA due to the differences in the nature of the border elements; the right T-DNA border (RB) contains the overdrive sequence and is at the leading end of the T-DNA's transfer into the plant (Gelvin, 2003). We observed a 1.7-fold (SE range 1.2-2.3) higher level of normalized LUC signal when the *LUC* gene was located next to the RB compared with at the left T-DNA border (Figure 1); these data showed that the configuration of *LUC* gene constructs in downstream positions of the RB gave higher reporter gene activity measurements and may be the best location to detect rare recombination events.

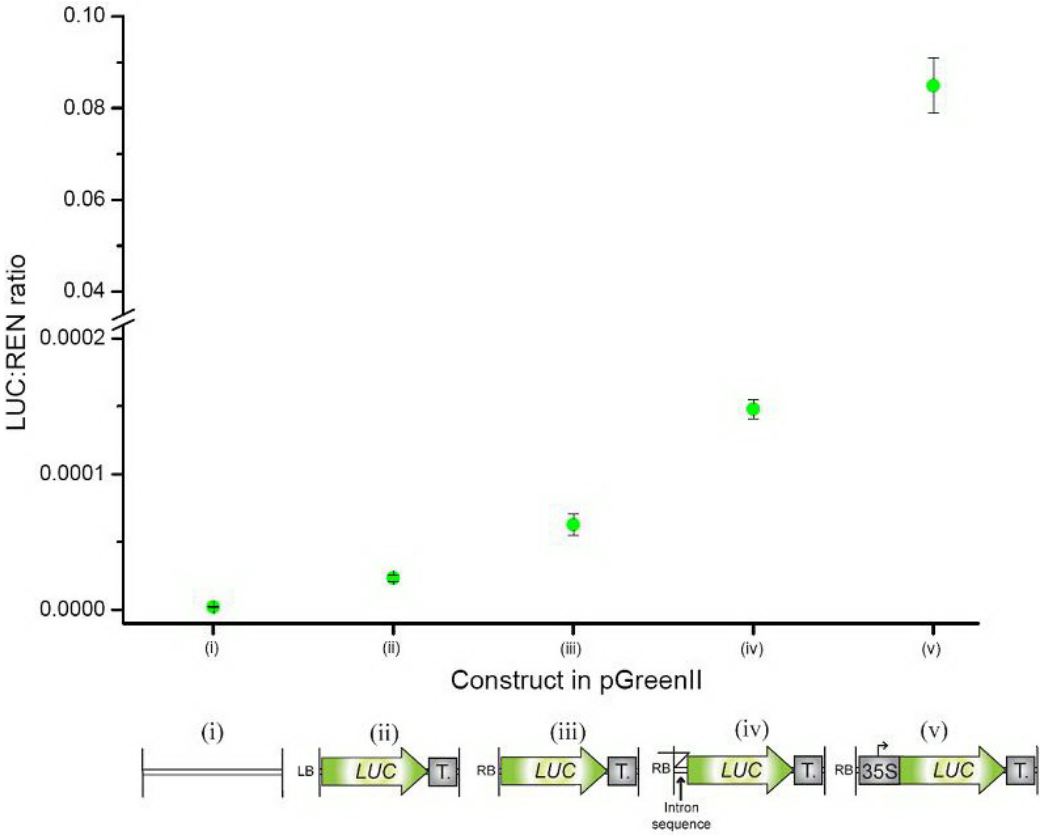


Figure 1. Effect of T-DNA border and *NtEF-1a* sequences fused to *LUC* upon transient expression in *Nicotiana benthamiana* leaves. The pGreenII vectors used for these plots are i) pGreenII 0000; ii) pGreenII LB-*LUC*-Term; iii) pGreenII RB-*LUC*-Term; iv) pGreenII RB-*EF*-*LUC*-Term, and v) pGreenII:*LUC*, respectively. The levels of *Agro*-infiltration were standardized using the pSoup:*REN* (or pSoup 0800) normalization vector T-DNA. *Agrobacterium tumefaciens* strain *GV3101* (MP90), and an incubation period of 3 days were used. Data points use ≥6 sample runs from ≥3 infiltrated *N. benthamiana* plants, and are representative of at least two separate experiments. Error bars are standard error. The label '35S' refers to a 35S promoter, 'T.' refers to a cauliflower mosaic virus (CaMV) terminator, and 'Intron sequence' refers to a partial *NtEF-1α* (TA2602) 5' UTR intron. LB = left T-DNA border; RB = right T-DNA border.

The testing of an upstream intron sequence's impact on reporter gene expression

In many plant species the *EF-1α* gene exists as a family of genes that contain an intron sequence in the 5' UTR (Chung et al., 2006). Intron fragments were amplified from a number of *EF-1α* 5' UTRs in tobacco for subsequent use in directing HR; *N. tabacum* was used to derive introns, rather than the *N. benthamiana* relative that was used in the transient expression assays. This species was used due to a greater number of expressed sequence tags being available for tobacco in the Plant Transcript Assemblies Database (Childs et al., 2007). The largest 5' UTR-intron of 951 bp was found for the *NtEF-1α* gene that had the identification code of TA2602 (Childs et al., 2007). The intron sizes determined for other gene members are shown

in Additional File 3. We predicted that the inclusion of this 5' UTR intron sequence in a vector without a splice-donor site should exhibit lower-level intrinsic promoter activity than would be possible if an upstream promoter sequence were to be used. To determine the transcriptional promoter activity of this sequence, a number of constructs were developed, and the experimental data from the testing of these constructs are also shown in Figure 1. We found that the relative level of *LUC* expression was 1.4-fold (SE range 1.0-1.8) higher when the 5' UTR intron fragment was located upstream of the *LUC* gene compared with constructs where the *LUC* gene was located directly adjacent to the RB. This difference was considerably low in relation to the 570-fold (SE range 510-650) higher level of reporter gene activity that was detected when the *LUC* gene was fused to a highly active constitutive cauliflower mosaic virus 35S promoter (35S); the finding suggested that this partial intron sequence could provide upstream homologous sequence in a promoter-less donor construct that was used in assays for HR.

The assessment of 14 candidate genes for their individual impact on homologous recombination frequencies

The components of the assays for recombination that were used to screen candidate genes are summarized in Figure 2. The vector to assay extra-chromosomal, intra-molecular HR (Figure 2A) contains a second complementary intron sequence that does not contain a splice acceptor site. Analysis of plasmid DNA by restriction enzyme digestion did not indicate that the HR assay vector had undergone modification or rearrangement while in either an *E. coli* or *A. tumefaciens* host. Full-length genes of enzymes involved in recombination were isolated from yeast (*Saccharomyces cerevisiae*) genomic DNA or *Arabidopsis* cDNA clones (Table 1). The *AtBRCA2*-IV gene (AT4G00020), however, was amplified from *Arabidopsis* genomic DNA and therefore contains introns (Dray et al., 2006). These genes were then fused to the 35S promoter in the plant transformation vector, pHEX2 (Hellens et al., 2005) and used in transient assays by mixing this effector gene vector-containing strain, with the strain that contains both *LUC* and *REN* genes to survey recombination frequencies, for co-infiltration into *N. benthamiana* leaves; the constructs are shown in Figure 2C. Effector genes were then tested for their influence on recombination frequencies and normalized relative to a control measurement that used the *GUS* gene, as this reporter gene would not be expected to alter recombination frequencies. The relative LUC levels in both HR and NHR assays is summarized in Figure 3 as plots of fold change in the average LUC:REN signal ratio. The raw data, shown in Additional File 4, has been processed to show this relative enhancement of LUC:REN signal by candidate genes with respect to the control infiltration, as a means to allow for cross-comparisons between data points (Hellens et al., 2005).

The data in Figure 3 show that eight genes, when over-expressed, had elevated signals for HR; these are *ScRAD52* (Di Primo et al., 2005), *AtSNM1* (Molinier et al., 2004), *ScRAD54* (Shaked et al., 2005), *AtCHR24* (Shaked et al. 2006), *ScMIM* (Hanin et al., 2000), *ScRAD51* (Schuermann et al., 2005), *AtRAD51D* (Durrant et al., 2007), and the *AtXRCC3* gene (Osakabe et al., 2002). The fold change in LUC:REN signal that was measured in the presence of these genes, with respect to the control, was determined to have P values that were less than 0.05. The remainder of the over-expressed gene constructs did not show increases in HR that could be considered as statistically significant; these alterations were either not beyond the margins of error (the upper and lower limits of fold change) or were determined to have P values that were greater than 0.05.

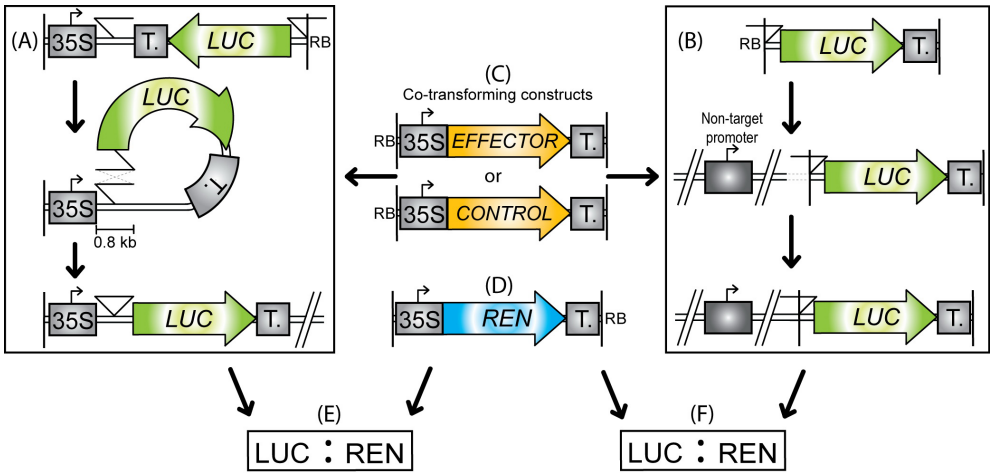


Figure 2. The use of Luciferase vectors to assay for effectors of HR and NHR. **A.** The T-DNA vector used for assaying intra-molecular HR from LUC activity that arises from reconstitution of a promoter-driven 5' UTR-*LUC* gene fusion (pGreenII RB-*EF*-*LUC*-Term-*EF*-35S) (Hellens et al. 2005). **B.** The T-DNA vector used to detect the LUC activity that arises from NHR-based non-target promoter trapping in assays for HR (pGreenII RB-*EF*-*LUC*-Term). **C.** Recombination-effector constructs that can thus be surveyed for any elicited impact on normalized Luciferase signal. **D.** A normalization vector T-DNA that can be used for standardizing the level of *Agro*-infiltration (pSoup:*REN* or pSoup 0800). **E.** The LUC:REN ratios for HR. **F.** The LUC:REN ratios for NHR. These ratios are compared to infer the roles of surveyed recombination-effector molecules. The vertical black lines show T-DNA borders and partial inverted triangles represent incomplete introns. '35S' refers to the CaMV 35S promoter, and 'T.' is its corresponding terminator. RB = right T-DNA border.

Further testing of eight candidate genes for their impact on non-homologous recombination frequencies

Eight genes, *AtSNM1*, *ScRAD54*, *AtCHR24*, *ScMIM*, *ScRAD51*, *AtRAD51D*, *ScRAD52*, and *AtXRCC3*, were also assessed for their impact on NHR using an assay for a donor molecule trapping promoters of an undetermined nature; these promoters could potentially include the 35S promoter of the effector gene's T-DNA molecule, or promoters in the *N. benthamiana* genome. Vectors with an *LUC* gene fused to a partial *NtEF-1α* 5' UTR intron, which was described for testing the intron's transcriptional promoter activity (Figure 1), was also used to measure NHR-activated *LUC* gene expression (Figure 2B). This analysis was carried out to account for any *LUC* gene expression that may have arisen from the HR vector T-DNA, in addition to intra-molecular HR. We observed that over-expression of the *AtSNM1* and *AtXRCC3* genes resulted in an apparent increase in NHR that was greater than the increases detected in assays for HR, with data that were found to be statistically significant. In contrast, over-expression of the *ScRAD52*, *AtCHR24*, *ScMIM*, and *ScRAD51* genes had enhanced measurements of NHR, but to a lesser extent compared with the measurements of HR. The greatest influence on the HR:NHR ratio was detected for a homologue of the *ScRAD54* gene, *ScMIM* (Hanin et al., 2000). The enhancement in the levels of NHR for the *ScMIM* gene may not have been statistically significant (P = 0.060). The apparent enhancement in NHR, however, may

have been significant in the cases of the *AtCHR24* (P = 0.048) and the *ScRAD54* (P = 3.2 x 10^{-6}) gene over-expression constructs. It would appear that LUC:REN in assays for NHR is not significantly altered in the presence of the *ScRAD52* (P = 0.63) and *ScRAD51* (P = 0.11) gene over-expression constructs. For the *AtRAD51D* and *ScRAD54* gene over-expression constructs, the increased average LUC:REN ratios observed in assays for NHR appear to be equal in magnitude to the LUC:REN increases observed in assays for HR. In order to determine if these candidate gene effectors of HR may function synergistically, combinatorial assessments of these effector genes were conducted by co-infiltration of various *Agrobacterium* strains (Hellens et al., 2005); however, the various combinations of these genes did not exhibit greater increases on HR compared with individual genes.

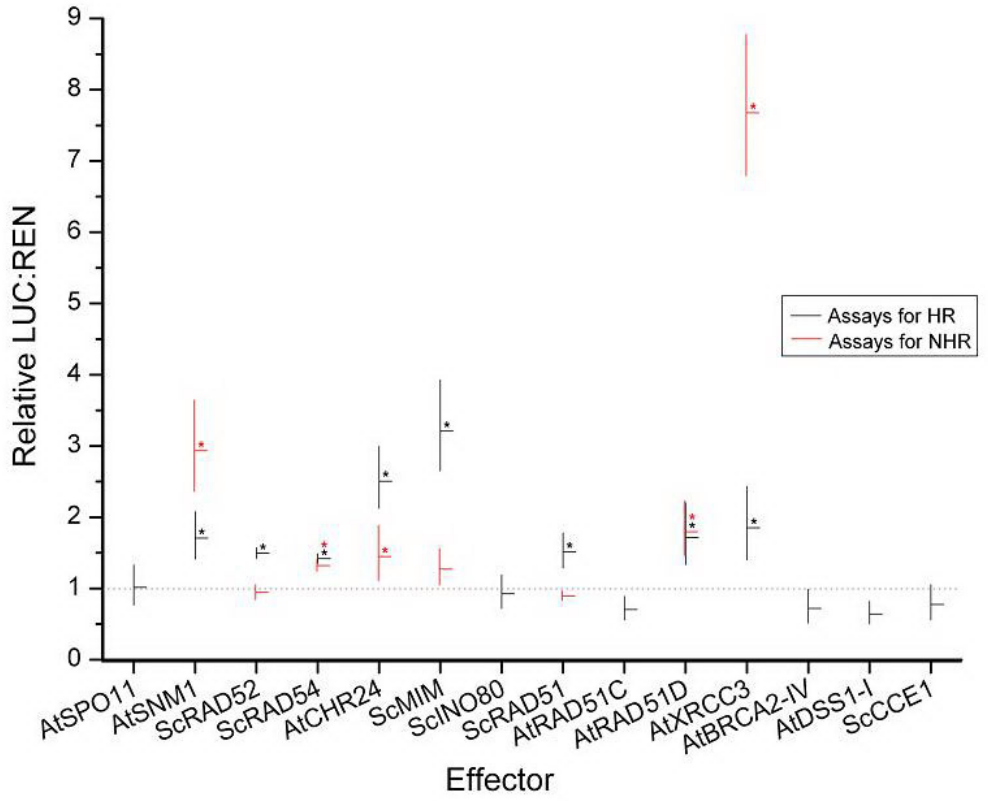


Figure 3. Relative effect of candidate over-expressed genes upon homologous recombination (HR) and non-homologous recombination (NHR) using assay for transient *LUC* expression in transfected *Nicotiana benthamiana* plant leaves. Plots shown are standardized to a control measurement that uses the *GUS* gene. Co-infiltrated over-expressed gene effectors are in the pHEX2 vector. The error bars show fold change between the effector gene and the *GUS* gene control at the upper (effector (average + SE)) / (control (average - SE)) and lower (effector (average - SE)) / (control (average + SE)) limits, respectively. Asterisks show a statistically significant (P < 0.05) difference between the measurements for the effector and control genes. Full-gene names are shown in Table 1. *Agrobacterium tumefaciens* strain *LBA4404* was used, with a 7-day incubation period. Data points are an average of ≥6 sample runs from ≥4 infiltrated *N. benthamiana* plants, and are representative of at least two separate experiments. Raw data is shown in Additional File 4.

Validating the impact of *AtSNM1* and *AtXRCC3* gene over-expression using an alternative assay for non-homologous recombination

A transient expression-based assay for the activation of *LUC* gene expression was configured to show extra-chromosomal joining of T-DNA ends after *Agrobacterium*-mediated leaf transformation. This vector has a *LUC* gene and promoter at distal ends; the action of an NHR event can fuse the 'promoter-less' *LUC* gene to the *NOS* promoter. The assay of such a vector is similar to a previously reported assay that was based on a linearized plasmid being introduced into plant leaves using biolistics (Gorbunova et al., 2000). The chemiluminescence measurements that were made in this study lend themselves to an increase in sample throughput, as the recombination events do not rely on manual counting of reporter gene expression events. A description of the constructs that were used in this assay is presented in Figure 4.

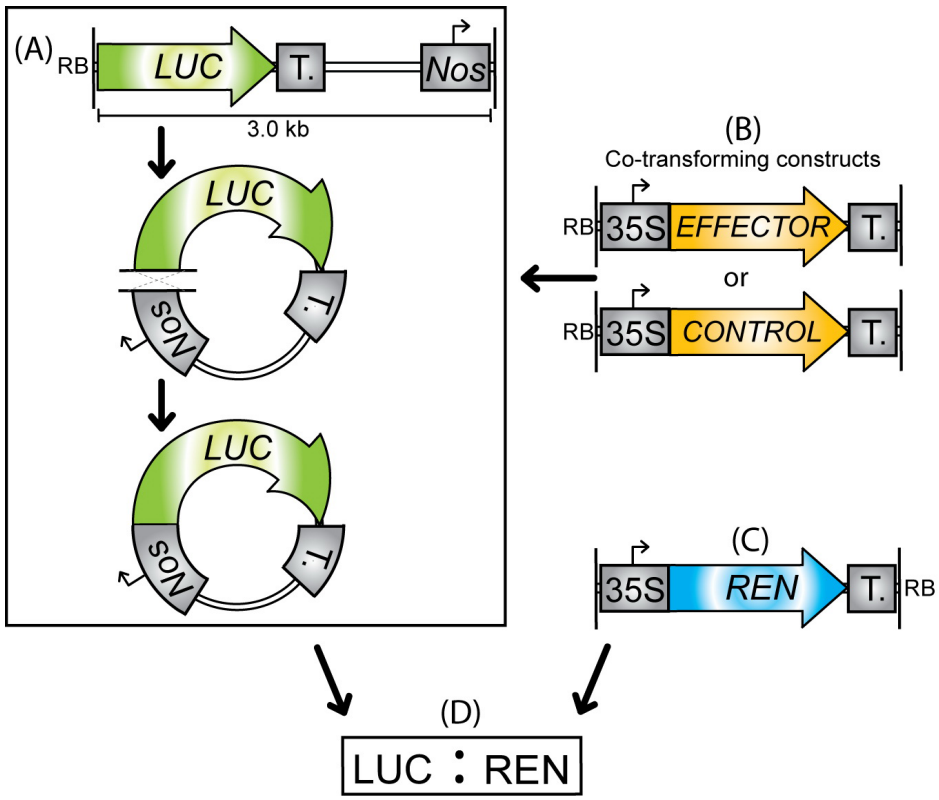


Figure 4. The use of Luciferase vectors to further assay effectors of NHR. **A.** The T-DNA vector can be used for assaying extra-chromosomal T-DNA end-joining from LUC activity that arises from reconstitution of a nopaline synthase (*NOS*) promoter-driven *LUC* gene fusion (pGreenII RB-*LUC*-Term-*NOS*) (Gorbunova et al., 2000). **B.** Recombination-effector constructs that can thus be surveyed for any elicited impact on normalized Luciferase signal. **C.** A normalization vector T-DNA that can be used for standardizing the level of *Agro*-infiltration. **D.** The LUC:REN ratios for NHR, which are used to infer the roles of surveyed recombination-effector molecules. The vertical black lines show T-DNA borders. '35S' is a CaMV 35S promoter, and 'T.' is its corresponding terminator. RB = right T-DNA border.

Figure 5 shows data regarding evaluation of the *AtSNM1* and *AtXRCC3* genes, both of which show significantly enhanced NHR upon over-expression (the P values were lower than 0.01). Consistent with earlier observations (Figure 3), over-expression of the *AtXRCC3* gene appears to confer a greater increase in NHR signal than over-expression of the *AtSNM1* gene. It must be noted that the fold changes determined for the *AtSNM1* and *AtXRCC3* gene-enhancement of NHR shown in Figure 5 are much less than the data shown in Figure 3; this difference may arise from the nature of the assays involved, namely the following three factors. Firstly, i) the simple end-to-end self-ligation of a T-DNA (Figure 5) may have occurred more rapidly than NHR assayed by non-target promoter-trapping (Figure 3), and so may have understated the impact of an effector gene whose protein product required expression to subsequently interact with an NHR vector molecule (Gelvin, 2003). In addition, ii) *LUC* gene expression in Figure 5 was driven by the *NOS* promoter, where it was driven by the 35S promoter in the assay whose data are shown in Figure 3. Alternatively, iii), the *A. tumefaciens* strain *GV3101* (MP90) was used to deliver vector T-DNAs into plant cells for the data shown in Figure 5, whereas *A. tumefaciens* strain *LBA4404* was used to derive the data shown in Figure 3.

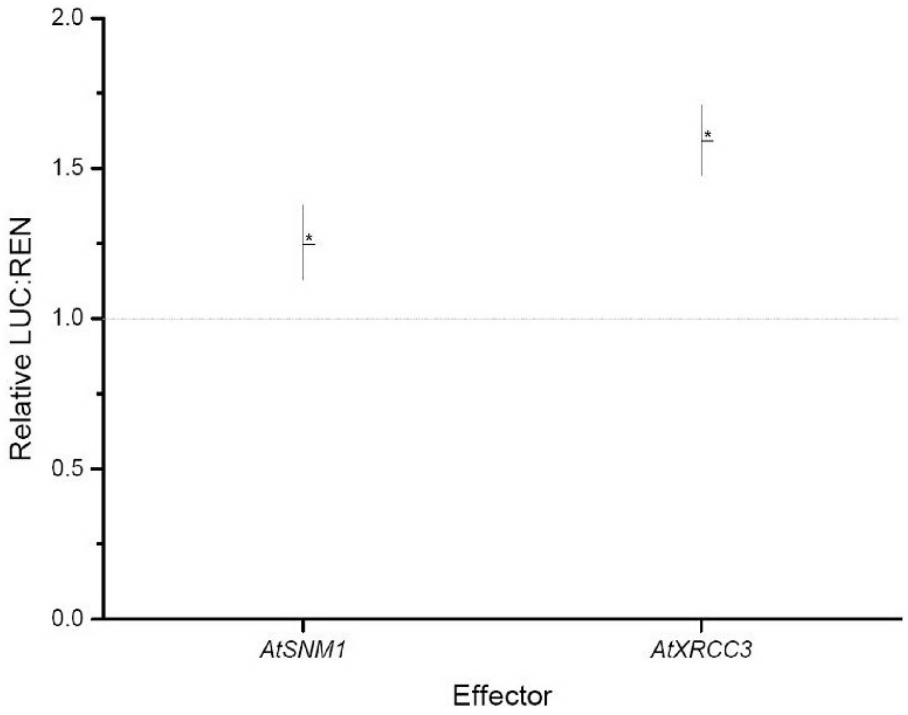


Figure 5. Relative effect of over-expressing the *AtSNM1* and *AtXRCC3* genes on NHR using a transient *LUC* expression assay in transfected *Nicotiana benthamiana* plant leaves. Plots shown are standardized to a control measurement that uses the *GUS* gene. Asterisks show a statistically significant ($P < 0.05$) difference between the measurements for the effector and control genes. *Agrobacterium tumefaciens* strain *GV3101* (MP90) was used, with a 7-day incubation period. Data points are an average of ≥6 sample runs from ≥3 infiltrated *N. benthamiana* plants, and are representative of at least two separate experiments. The error bars note the maximum upper and lower fold changes between the effector (± SE) and the control (± SE) measurements (Figure 3). Raw data is shown in Additional File 4.

DISCUSSION

The data presented in this study describe a series of vectors designed for transient transformation that were configured to rapidly assess the potential of candidate genes in influencing extra-chromosomal HR and/or NHR. These extra-chromosomal assays were developed and used to rapidly screen effectors for determining those yeast- or *Arabidopsis*-derived factors with the potential to influence HR in the heterologous *N. benthamiana* leaf system. The assays are also intended to overcome the constraints that are associated with assessing effector genes in stable transgenic systems, which involve longer time frames and potential chromosome positioning effects (Puchta and Hohn, 1991). Reproducible differences in normalized LUC activity were observed for a number of effector genes that have previously been reported to play a role in recombination. The findings were that the *AtSNM1* and *AtXRCC3* effector genes enhanced signal in assays for NHR more than that for HR, and vice versa for the *ScRAD52*, *AtCHR24*, *ScMIM*, and *ScRAD51* genes; the difference in the recombination pathways affected by these genes would suggest that the assay for HR was not primarily measuring NHR, and vice versa. In regard to transforming donor molecules having potentially undergone insertion into the *N. benthamiana* genome during assays for recombination, such integrations are typically rare compared with the majority of transforming molecules that become transcriptionally competent extra-chromosomal entities (Lyznik, 1991; Orel and Puchta, 2003). The majority of donor molecules undergoing genomic-insertion would do so by way of NHR (Iida and Terada, 2005); the findings of our two types of assay for NHR would therefore not be distorted if this same kind of recombination event caused LUC activity from the genomic integration of donor molecules. Furthermore, as previously introduced, genomic insertion by HR occurs very infrequently in plants, so was considered to cause negligible false signal in these assays, particularly with the use of recombination substrates that employed little or no homology to the *N. benthamiana* genome. The enhancements in NHR that were determined in this study are discussed immediately below, with the enhancements in HR being presented subsequently.

Two effector genes, *AtSNM1* (Molinier et al., 2004) and *AtXRCC3* (Osakabe et al., 2002), were found to positively regulate the level of NHR. These genes were also found to positively regulate HR, an observation that has previously been reported (Bleuyard and White, 2004; Molinier et al., 2004); however, greater relative enhancements in NHR were detected for the genes in question. Data derived from a promoter-trap assay or NHR were corroborated using a T-DNA end-joining assay, which uses a vector that lacks both homology to the *N. benthamiana* genome and repeats of homologous sequence. Both assays for NHR showed that the greatest enhancement in this form of recombination was achieved using the *AtXRCC3* gene. These data strongly suggest that the SNM1 and XRCC3 enzymes can stimulate homologous and non-homologous recombination and therefore may be key regulators of DNA repair and the GT frequency in plants (Bleuyard et al., 2006). These effector genes have been studied with respect to their individual alteration of homologous recombination (Bleuyard and White, 2004; Hemphill et al., 2008), but not in regard to their effects on non-homologous recombination. While homologous and non-homologous recombination may compete for substrates (Britt and May, 2003), these enzymes may be involved in up-regulating both recombination processes, as opposed to one type specifically; these functions are discussed below.

The observation that SNM1 may contribute to NHR in plants is consistent with the report of a previous study in mammalian cells that the SNM1 enzyme is involved in pathways

that are separate from HR (Hemphill et al., 2008). The substrates of SNM1 are thought to include the intermediary molecules during restoration of DNA interstrand crosslinks (ICLs) to duplex chromosomal DNA (Li and Moses, 2003). Despite the involvement of *SNM1* in ICL-repair in mammals (Hemphill et al., 2008), the SNM1 enzyme was reported not to be involved in plant ICL-repair (Molinier et al., 2004). SNM1 has been reported to function in the repair of oxidative stress-induced DNA damage in plants (Molinier et al., 2004) and may have alternative functions in plants (Schuermann et al., 2005). The *SNM1* gene, however, is up-regulated in response to various types of DNA damage in rice (Kimura et al., 2005), consistent with a role in DNA repair. SNM1 has exonuclease activity (Li et al., 2005) and this may affect NHR through such mechanisms as DNA end-blunting by exonuclease activity.

The increase in NHR elicited by over-expressing *AtXRCC3* is unexpected given this gene is a paralogue of the *ScRAD51* gene, the core recombinase in yeast HR. A previous study found that *Arabidopsis xrcc3* mutant plants showed chromosomal fragmentation during meiosis and were hypersensitive to DNA-damage in meiosis, which suggested defective HR (Bleuyard and White, 2004). Our data strongly suggest that XRCC3 operates not only in the processes of HR, but also NHR, which is a novel finding to our knowledge. This finding is in line with studies reporting that the *XRCC3* gene plays a role in DNA damage repair (Kurumizaka et al., 2001; Brenneman et al., 2002), the dominant means of which appears to be NHR in *Arabidopsis* (Iida and Terada, 2005; Kim et al., 2007). The enzyme is known to catalyze homologous DNA-pairing in a protein complex with RAD51C termed the 'CX3' enzyme complex (Kurumizaka et al., 2001). It appears that there is a difference in plants and vertebrates with respect to the role of XRCC3 in HR; *xrcc3* mutant *Arabidopsis* plants show normal inter-chromosomal HR (Bleuyard and White, 2004), yet *xrcc3* mutant 'irs1SF' Chinese hamster cells exhibit reduced intra-chromosomal HR, which could be complemented by the *XRCC3* gene (Brenneman et al., 2000). We suggest that XRCC may also facilitate NHR, in addition to HR, through its early localization to double-stranded DNA break sites (Forget et al., 2004) and its role in DNA-binding (Kurumizaka et al., 2001).

Due to the low frequency of HR occurring in plants (Britt and May, 2003), the following control measures were taken during assays for HR to avoid interference from alternative processes. Firstly, the inclusion of an upstream intron (without a splice donor site) was used, not only to guide HR events, but also to limit *LUC* gene expression from any donor molecules undergoing NHR within *N. benthamiana* genomes (Hrouda and Paszkowski, 1994). This partial intron sequence was also included in the donor molecule used to assay for the level of LUC activity that could have arisen from the donor molecule trapping any non-target promoter elements by NHR in assays for HR. This pair-wise testing of effector genes for their impact on both HR and NHR offered means to determine the major recombination mechanism that each effector gene impacted. Secondly, a design feature of the vector that was used to assay for HR was the locating of the RB in an upstream position of the *LUC* gene in order to exclude the reporter gene from any T-DNAs that underwent recombination between intron sequences in an *E. coli* or *A. tumefaciens* host; however, restriction enzyme analysis of plasmid DNA isolated from these hosts did not indicate that this event had occurred. The different findings for effector genes in assays for HR strongly suggest that the HR vector does not always undergo recombination within the *A. tumefaciens* host by self-associating either before or during delivery into plant cells, when the T-DNA molecule is at the stage of the single-stranded T-strand or T-complex (Gelvin, 2003). Thirdly, the design feature of using the pSoup-derived *REN*

gene expression vector to normalize the levels of *Agro*-infiltration in assays for LUC activity, compared with the alternative of having a more complex donor vector with a T-DNA region containing both of these reporter gene expression cassettes, may have reduced the chances of non-target recombination occurring within a donor T-DNA molecules in assays for recombination. The 3.2-fold difference in LUC:REN ratio that was measured for these respective scenarios may have arisen due to differences in the transfection efficiency of these pGreenII and pSoup T-DNAs (Gelvin, 2003).

It must also be considered that this extra-chromosomal, intra-molecular system used to assay for recombination may differ from a chromosomal assay system for HR in terms of i) The level of histone proteins bound to the DNA (Coates et al., 1987); the positive effect observed for the over-expression of chromatin remodeling factor genes would indicate that some histone protein complexes may be present on the extra-chromosomal HR vector T-DNA construct. ii) A difference in the methylation modification of the bacterial-derived HR vector T-DNA when compared with the eukaryotic genomic DNA that effector enzymes may normally interact with (Coates et al., 1987). iii) The relative importance of homology search preceding HR in experiments, that assay for inter-chromosomal HR or GT (Puchta and Hohn, 1991). iv) In some experiments, the background 'noise' may contribute to transient expression measurements in the form of inaccurate termination of the T-strand as it is being excised from the plant transformation vector, as part of the T-DNA's delivery into the plant cell (Kim et al., 2007).

The over-expression of the *ScRAD52*, *AtCHR24*, *ScMIM*, and *ScRAD51* genes was found to enhance the level of HR, relative to NHR, in the transient expression assays of recombination. The observation for over-expressing the *RAD51* gene enhancing HR to a greater extent than NHR has also been found in studies expressing the *E. coli* orthologue of the *RAD51* gene: the *REC-A* factor (Reiss et al., 1996). It has also been reported that the expression of the *RAD51* gene in *Arabidopsis* is enhanced after DNA damage (Bleuyard et al., 2006), indicating a potential importance of this factor in DNA repair. The enhancement of HR by the *ScRAD52* gene relative to NHR is in line with previous research (Di Primo et al., 2005). As there is no RAD52 orthologue present in plants, when introduced into plant cells, RAD52 may facilitate a similar role to that conducted in the highly HR-proficient yeast cells in binding DNA ends and recruiting factors to conduct HR-based repair (Cotsaftis and Guiderdoni, 2005).

The *ScMIM*, *AtCHR24*, *ScRAD54*, and also the *ScINO80* genes are close homologues (Shaked et al., 2006). Their enzymes are believed to participate in chromatin remodeling, which involves the repositioning of nucleosomes along DNA strands during the stage of homology searching as part of DNA strand invasion in HR (Shaked et al., 2005). The data we report are in line with a previous study that reported the role of the *MIM* gene in HR (Hanin et al., 2000). The *AtCHR24* gene was reported as being up-regulated in response to DNA damage (Shaked et al., 2006); however, it does not seem that the *AtCHR24* gene has been previously reported to enhance HR. The effect of over-expressing these genes was found to be similar to that reported for the closely homologous *ScRAD54* gene, which has been previously reported to elicit an increase in the GT frequency in stable transgenic *Arabidopsis* plants (Shaked et al., 2005). The *ScRAD54* gene was found to enhance the measurements of both HR and NHR in the *N. benthamiana*-based transient expression assays of recombination presented here; accordingly, the chromatin configuration of the *Arabidopsis* chromosomal system, which was used in a previous study of *ScRAD54* (Shaked et al., 2005), may differ from that in this extra-chromosomal T-DNA system. Additionally, the RAD54 enzyme is known to partner with

RAD52, RAD51, RAD55, and RAD57 (XRCC3) enzymes to perform HR in yeast (Bleuyard et al., 2006), and may perform additional roles to chromatin remodeling. It must be considered that the RAD54 enzyme may interact differently with enzyme partners in the transient *N. benthamiana* system. Such chromatin or enzyme-partnering differences may also be behind why our transient expression assays did not reveal an enhancement in HR in response to *ScINO80* gene (Shaked et al., 2006). It does not yet appear reported whether the *Sc*MIM, *At*CHR24 and *Sc*INO80 enzymes undergo the same biological interactions as RAD54 *in planta* (Schuermann et al., 2005). Our tests strongly suggested that the *AtRAD51D* gene also leads to an increase in HR, as well as in NHR, which is consistent with earlier study that measured its influence on HR and the level of DNA damage repair (Durrant et al., 2007).

The enzymes produced from effector genes that were not found to alter recombination frequencies in our assay system, *AtSPO11*, *ScRAD54*, *AtRAD51C*, *AtRAD51D*, *ScINO80*, *AtBRCA2-IV, AtDSS1-I*, and *ScCCE1*, may not have compensated for endogenous rate-limiting reactions in the HR process. It is also possible that these genes may have been incompatible with *N. benthamiana* cells' endogenous recombination machinery, possibly due to a divergence in the evolutionary history of these genes, or these enzymes were unable to interact with an extra-chromosomal T-DNA vector to enhance the level of HR during the time period of the assay. These genes may also perform roles in HR that could not be measured in our transient expression assay for HR, which uses extra-chromosomal T-DNA molecules.

The candidate recombination enhancing genes that have been identified in this transient expression assay screening, particularly for HR, will require validation in other experimental systems including stable transformation. On the other hand, the transient assay system that we describe here provides a means of rapid and potentially high throughput screening of effector genes, and may allow for various combinations of these genes to be assessed.

CONCLUSIONS

We have developed vectors that are specifically designed to assess the extent of extra-chromosomal HR and NHR in a transient expression assay. Using these constructs we have found that *AtSNM1* and *AtXRCC3* are candidates for suppression as a means of enhancing the GT frequency in plants. Moreover, our data showed that the *ScRAD52*, *AtCHR24*, *ScMIM*, and *ScRAD51* genes are candidates for over-expression as an additional means of enhancing the GT frequency in plants. To our knowledge, there are no studies at present that individually report these many candidate genes with the potential to be used in enhancing plant GT frequencies. While the *AtSPO11*, *ScRAD54*, *AtRAD51C*, *AtRAD51D*, *ScINO80*, *AtBRCA2-IV*, *AtDSS1-I*, and *ScCCE1* genes had less clearly defined impacts on recombination frequencies in our *N. benthamiana*-based transient assays, we cannot rule out the possibility that the influence of these genes was not measured in the 7-day time course of this HR assay that uses extra-chromosomal vector T-DNA molecules.

ACKNOWLEDGMENTS

We thank the researchers at The New Zealand Institute for Plant & Food Research Limited including Niels Nieuwenhuizen, Andrew P. Dare, Andrew C. Allan, Richard V. Espley, Kui Lin-Wang, Supinya (Nicky) Dejnoprat, and Sean Bulley for advice over the course of the re-

search. Further thanks to Andrew P. Gleave and Philippa J. Barrell for editing of the manuscript, and to Julie Nicholls and G. (Wade) Wadasinghe for the upkeep of the glasshouse facilities. Funding was provided by Doctoral Scholarships from the Agriculture Marketing Research and Development Trust (AGMARDT) and the University of Auckland. We thank the ABRC, INRA-CNRGV, and RIKEN BIORESOURCE CENTRE for cDNA clones, and the Wine Science Research Group (the University of Auckland) for yeast genomic DNA. We declare no competing interests.

REFERENCES

Bleuyard JY and White CI (2004). The *Arabidopsis* homologue of Xrcc3 plays an essential role in meiosis. *EMBO J.* 23: 439-449.

Bleuyard JY, Gallego ME and White CI (2006). Recent advances in understanding of the DNA double-strand break repair machinery of plants. *DNA Repair* 5: 1-12.

Brenneman MA, Weiss AE, Nickoloff JA and Chen DJ (2000). XRCC3 is required for efficient repair of chromosome breaks by homologous recombination. *Mutat. Res.* 459: 89-97.

Brenneman MA, Wagener BM, Miller CA, Allen C, et al. (2002). XRCC3 controls the fidelity of homologous recombination: roles for XRCC3 in late stages of recombination. *Mol. Cell* 10: 387-395.

Britt AB and May GD (2003). Re-engineering plant gene targeting. *Trends Plant Sci.* 8: 90-95.

Childs KL, Hamilton JP, Zhu W, Ly E, et al. (2007). The TIGR plant transcript assemblies database. *Nucleic Acids Res.* 35: D846-D851.

Chung BY, Simons C, Firth AE, Brown CM, et al. (2006). Effect of 5' UTR introns on gene expression in *Arabidopsis thaliana*. *BMC Genomics* 7: 120.

Coates D, Taliercio EW and Gelvin SB (1987). Chromatin structure of integrated T-DNA in crown gall tumors. *Plant Mol. Biol.* 8: 159-168.

Cotsaftis O and Guiderdoni E (2005). Enhancing gene targeting efficiency in higher plants: rice is on the move. *Transgenic Res.* 14: 1-14.

Di Primo C, Galli A, Cervelli T, Zoppe M, et al. (2005). Potentiation of gene targeting in human cells by expression of *Saccharomyces cerevisiae* Rad52. *Nucleic Acids Res.* 33: 4639-4648.

Dray E, Siaud N, Dubois E and Doutriaux MP (2006). Interaction between *Arabidopsis* Brca2 and its partners Rad51, Dmc1, and Dss1. *Plant Physiol.* 140: 1059-1069.

Durrant WE, Wang S and Dong X (2007). *Arabidopsis* SNI1 and RAD51D regulate both gene transcription and DNA recombination during the defense response. *Proc. Natl. Acad. Sci. U. S. A.* 104: 4223-4227.

Forget AL, Bennett BT and Knight KL (2004). Xrcc3 is recruited to DNA double strand breaks early and independent of Rad51. *J. Cell. Biochem.* 93: 429-436.

Fuller LF and Painter RB (1988). A Chinese hamster ovary cell line hypersensitive to ionizing radiation and deficient in repair replication. *Mutat. Res.* 193: 109-121.

Gelvin SB (2003). *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol. Mol. Biol. Rev.* 67: 16-37, table.

Gorbunova V, Avivi-Ragolski N, Shalev G, Kovalchuk I, et al. (2000). A new hyperrecombinogenic mutant of *Nicotiana tabacum*. *Plant J.* 24: 601-611.

Hanin M, Mengiste T, Bogucki A and Paszkowski J (2000). Elevated levels of intrachromosomal homologous recombination in *Arabidopsis* overexpressing the *MIM* gene. *Plant J.* 24: 183-189.

Hellens RP, Allan AC, Friel EN, Bolitho K, et al. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* 1: 13.

Hemphill AW, Bruun D, Thrun L, Akkari Y, et al. (2008). Mammalian SNM1 is required for genome stability. *Mol. Genet. Metab.* 94: 38-45.

Hrouda M and Paszkowski J (1994). High fidelity extrachromosomal recombination and gene targeting in plants. *Mol. Gen. Genet.* 243: 106-111.

Iida S and Terada R (2005). Modification of endogenous natural genes by gene targeting in rice and other higher plants. *Plant Mol. Biol.* 59: 205-219.

Ilnytskyy Y, Boyko A and Kovalchuk I (2004). Luciferase-based transgenic recombination assay is more sensitive than beta-glucoronidase-based. *Mutat. Res.* 559: 189-197.

Kim SI and Gelvin SB (2007). Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. *Plant J.* 51: 779-791.

Kimura S, Saotome A, Uchiyama Y, Mori Y, et al. (2005). The expression of the rice (*Oryza sativa* L.) homologue of Snm1 is induced by DNA damages. *Biochem. Biophys. Res. Commun.* 329: 668-672.

Kurumizaka H, Ikawa S, Nakada M, Eda K, et al. (2001). Homologous-pairing activity of the human DNA-repair proteins Xrcc3.Rad51C. *Proc. Natl. Acad. Sci. U. S. A.* 98: 5538-5543.

Li L, Santerre-Ayotte S, Boivin EB, Jean M, et al. (2004). A novel reporter for intrachromosomal homoeologous recombination in *Arabidopsis thaliana*. *Plant J.* 40: 1007-1015.

Li X and Moses RE (2003). The β-lactamase motif in Snm1 is required for repair of DNA double-strand breaks caused by interstrand crosslinks in *S. cerevisiae*. *DNA Repair* 2: 121-129.

Li X, Hejna J and Moses RE (2005). The yeast Snm1 protein is a DNA 5'-exonuclease. *DNA Repair* 4: 163-170.

Lyznik LA, McGee JD, Tung PY, Bennetzen JL, et al. (1991). Homologous recombination between plasmid DNA molecules in maize protoplasts. *Mol. Gen. Genet.* 230: 209-218.

Molinier J, Stamm ME and Hohn B (2004). SNM-dependent recombinational repair of oxidatively induced DNA damage in *Arabidopsis thaliana*. *EMBO Rep.* 5: 994-999.

Orel N and Puchta H (2003). Differences in the processing of DNA ends in *Arabidopsis thaliana* and tobacco: possible implications for genome evolution. *Plant Mol. Biol.* 51: 523-531.

Osakabe K, Yoshioka T, Ichikawa H and Toki S (2002). Molecular cloning and characterization of *RAD51*-like genes from *Arabidopsis thaliana*. *Plant Mol. Biol.* 50: 71-81.

Puchta H and Hohn B (1991). 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.

Reiss B, Klemm M, Kosak H and Schell J (1996). RecA protein stimulates homologous recombination in plants. *Proc. Natl. Acad. Sci. U. S. A.* 93: 3094-3098.

Schuermann D, Molinier J, Fritsch O and Hohn B (2005). The dual nature of homologous recombination in plants. *Trends Genet.* 21: 172-181.

Shaked H, Melamed-Bessudo C and Levy AA (2005). High-frequency gene targeting in *Arabidopsis* plants expressing the yeast RAD54 gene. *Proc. Natl. Acad. Sci. U. S. A.* 102: 12265-12269.

Shaked H, Avivi-Ragolsky N and Levy AA (2006). Involvement of the *Arabidopsis* SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. *Genetics* 173: 985-994.

Shalev G, Sitrit Y, Avivi-Ragolski N, Lichtenstein C, et al. (1999). Stimulation of homologous recombination in plants by expression of the bacterial resolvase RuvC. *Proc. Natl. Acad Sci U. S. A.* 96: 7398-7402.

Tebbs RS, Zhao Y, Tucker JD, Scheerer JB, et al. (1995). Correction of chromosomal instability and sensitivity to diverse mutagens by a cloned cDNA of the XRCC3 DNA repair gene. *Proc. Natl. Acad. Sci. U. S. A.* 92: 6354-6358.

Ursin VM, Irvine JM, Hiatt WR and Shewmaker CK (1991). Developmental analysis of elongation factor-1 alpha expression in transgenic tobacco. *Plant Cell* 3: 583-591.

ADDITIONAL FILES

Additional File 1. Oligonucleotide primers used.

Table showing sequences and applications of the oligonucleotide primer sequences that were used in the study.

Name	Sequence (5' - 3')	Application
Forward adapter	GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT	GenomeWalker™ Library Preparation [93]
Reverse adapter	(P)-ACCAGCCC-NH_2	GenomeWalker™ Library Preparation [93]
RAJ-044	GTGCCTTGGTTTTCCAAGGTGGCTCAGAAG	Gene-specific primer 2, 1st *NtEF-1α* (TA2602) 3' GenomeWalker™ amplification (GW)
RAJ-048	CCACACGACCAACAGGGACAGTACC	Reverse, amplification of *NtEF-1α* (TA2602) locus
RAJ-051	TCCCCTTTGTCCCCATCTCTGGT	Forward, Positive Control PCR amplification on *NtEF-1α* (TA2602) Locus with RAJ-048
RAJ-059	GCACCCAGGCATACTTGAATGACCTCTTGTTCATCTCAG	Reverse, amplification of *NtEF-1α* (TA2602) Locus upstream sequence
RAJ-069	GTCAGGCAATTAAATCTGTGGATATGCATCTGCGGTT	Forward, amplification of *NtEF-1α* (TA2602) Locus upstream sequence
RAJ-105	CTGTATGCGACGGCCATAAGAAACAATACTAATAGATTTTTCCAC	Forward, amplification of *NtEF-1α* (TA2602) Locus upstream sequence
RAJ-123	AAGCTGAACTCATCCCGCAGCCCAACA	Forward, amplification of *NtEF-1α* (TA2602) Locus upstream sequence
RAJ-155	GGAACCAGCAGGCAGCAGGCAGAAGG	Forward, amplification of *NtEF-1α* (TA2602) Locus downstream sequence
RAJ-161	CTTCTCCTTCTGCTGTGCGGCA	Forward, amplification from the *NtEF-1α* (TA2602) 5' UTR, pre-intron
RAJ-164	CTTGATGAAGTCCCTGTGTCCAGGAGCATC	Reverse, amplification of *NtEF-1α* (TA2602) Locus upstream sequence
RAJ-167	GGAGATACTTCACCTGGGCAGAGAATCCTAATTC	Gene-specific primer 1, 7th *NtEF-1α* (TA2602) 3' GW
RAJ-168	CACTGGAGCTCATAACTTCACAACTTTAAGCGATC	Gene-specific primer 2, 7th *NtEF-1α* (TA2602) 3' GW
RAJ-169	CCCAATACCAACCCTGAAACCCTAAT	Forward, amplification from the *NtEF-1α* (TA2602) 5' UTR, pre-intron
RAJ-178	CTTTTAAGGTTAGATCGCGGCCGCTTTCGTTTCTCTTCAGTTTC	Forward, amplification from *NtEF-1α* (TA2602) 5' UTR splice donor, *Not*I site-introducing
RAJ-184	ACGTTGCAGTTCAATTCTGGTGGTACT	Forward, amplification of *NtEF-1α* (TA2602) Locus downstream sequence
RAJ-187	CGTACTTTTATAGCCACACAAATGACATGG	Reverse, amplification of *NtEF-1α* (TA2602) downstream sequence
RAJ-189	TTACATTTTATACACGGCGATCTTTCCG	Reverse, amplification from end of Luciferase, with stop codon
RAJ-198	CTTAGTTTACCCGCCAATATATCCTGTCA	Forward, Right Border of pGreenII, PCR colony screening
RAJ-199	AGATCTTGGCAGGATATATTGTGGTGTAAC	Reverse, Left Border of pGreenII, PCR colony screening
RAJ-206	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTCGACAGCACAGA AAGCTAAGATATTGC	Forward, amplification of the *ScCCE1* (codon sequence, cds) gene
RAJ-207	GGGGACCACTTTGTACAAGAAAGCTGGGTACACAATTTCTTAGTCATT GTTGTAAGTGTTCTGC	Reverse, amplification of the *ScCCE1* (cds) gene
RAJ-208	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGGCGTTTTTAAGCTA TTTTGCCACTG	Forward, amplification of the *ScRAD52* (cds) gene
RAJ-209	GGGGACCACTTTGTACAAGAAAGCTGGGTTGTTTCAAGTAGGCTTGC GTGCATG	Reverse, amplification of the *ScRAD52* (cds) gene
RAJ-210	GGGGACAAGTTTGTACAAAAAAGCAGGCTAAATGGCAAGACGCAGA TTACCAGACA	Forward, amplification of the *ScRAD54* (cds) gene
RAJ-211	GGGGACCACTTTGTACAAGAAAGCTGGGTTAGTATGTAAGAGATCAA TGTGAAATATATTGAAATGC	Reverse, amplification of the *ScRAD54* (cds) gene
RAJ-212	GGGGACAAGTTTGTACAAAAAAGCAGGCTAAATGGCTTTGGAGGG ATTGCG	Forward, amplification of the *ScSPO11* (cds) gene
RAJ-213	GGGGACCACTTTGTACAAGAAAGCTGGGTACTTATTCATTTGTATTCA AAAATTCTGGCATC	Reverse, amplification of the *ScSPO11* (cds) gene

Continued on next page

Additional File 1. Continued.

Name	Sequence (5' - 3')	Application
RAJ-229	AAAAAGCAGGCTGCCTACTAATTTGTTATCGTCATATGTCTCAAGTTC	Forward, amplification of the *ScRAD51* (cds) gene
RAJ-230	AGAAAGCTGGGTACCTACTCGTCTTCTTCTCTGGGGTCAC	Reverse, amplification of the *ScRAD51* (cds) gene
RAJ-245	GGGGACAAGTTTGTACAAAAAAGCAGGCT	Forward, amplification from the *att*B1 adapter
RAJ-246	GGGGACCACTTTGTACAAGAAAGCTGGGT	Reverse, amplification from the *att*B2 adapter
RAJ-265	AAAAAGCAGGCTATGGCGCCTCTCAAACATCTGGAG	Forward, amplification of the *AtRAD51D* (cds) gene
RAJ-266	AGAAAGCTGGGTTTATGGACATTGTTGATTCTCCTTGCGA	Reverse, amplification of the *AtRAD51D* (cds) gene
RAJ-267	AAAAAGCAGGCTATGGATTTTTCTGATGAAGACGACGACG	Forward, amplification of the *AtSNM1* (cds) gene
RAJ-268	AGAAAGCTGGGTTCAGCGTCTGAGCCATTCTCTGAAAC	Reverse, amplification of the *AtSNM1* (cds) gene
RAJ-279	AAAAAGCAGGCTATGATTTCATTTGGGCGGCGTAAATC	Forward, amplification of the *AtRAD51C* (cds) gene
RAJ-280	AGAAAGCTGGGTTTACATCATCTTGACTCGCTTGCTACTCG	Reverse, amplification of the *AtRAD51C* (cds) gene
RAJ-281	AAAAAGCAGGCTATGGCGGCAGAACCGAAGGCA	Forward, amplification of the *AtDSS1* –1 (cds) gene
RAJ-282	AGAAAGCTGGGTTTATTTCTTGTCAGTACCATTCTCAAGCTCCTTC	Reverse, amplification of the *AtDSS1* –1 (cds) gene
RAJ-289	AAAAAGCAGGCTATGGCGGAAAATACGGCCAGC	Forward, amplification of the *AtCHR24* (cds) gene
RAJ-290	AGAAAGCTGGGTCTACAGATTCAATCCTTTTTGCATCTTCCGAC	Reverse, amplification of the *AtCHR24* (cds) gene
RAJ-295	GGGGACAAGTTTGTACAAAAAAGCAGGCTGACATGTCACTGGCAGT TCTACTCAATAAGGAGGACAAGG	Forward, amplification of the *ScINO80* (cds) gene, secondary amplification
RAJ-296	GGGGACCACTTTGTACAAGAAAGCTGGGTTCATGCCAATGCACTTG CCGCTTCC	Reverse, amplification of the *ScINO80* (cds) gene, secondary amplification
RAJ-297	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGTATGATTTCGACTACG ATTAGTGGTAAAAGACCTATAGAGCAGG	Forward, amplification of the *ScMIM* (cds) gene
RAJ-298	GGGGACCACTTTGTACAAGAAAGCTGGGTGCTGCCGTGAAAACGAA TTCAAATAGACGATTACAC	Reverse, amplification of the *ScMIM* (cds) gene
RAJ-317	GAGCATTTATAGACATGTCACTGGCAGTTC	Forward, amplification of the *ScINO80* (cds) gene, primary amplification
RAJ-318	CCTTTTTCATGCCAATGCACTTGC	Reverse, amplification of the *ScINO80* (cds) gene, primary amplification
RAJ-338	ACCTGCAGGCTGATCATCGACGAATTAATTCCAATCC	Forward, start of CaMV 35S promoter on pHEX2, *BclI* site-introducing
RAJ-339	TCACTAGTGCTGATCAAGATTTAGGTGACACTATAG	Reverse, end of CaMV terminator on pHEX2, *BclI* site-introducing
RAJ-352	GTCAAGCAGATCGTTCAAACATTTGG	Reverse, from the end of the *NOS* terminator
RAJ-356	AAAAAGCAGGCTATGCAAAATGGGAAAATTAAGCCGGAGAATC	Forward, amplification of the *AtXRCC3* (cds) gene
RAJ-357	AGAAAGCTGGGTCTACGCTTGAACCGCACAAATCC	Reverse, amplification of the *AtXRCC3* (cds) gene
RAJ-360	GTTCAATCATGCGGCCGCCGAGTTGAGAGTGAATATGAG	Forward, from the *NOS* Promoter on pGreenII 0049, *Not*I-introducing
RAJ-361	TGGAGAGGGCGGCCGCAGCTCATGGAAGACGCCAAAAAC	Reverse, from start of *LUC* on pGreenII 0049, *Not*I-introducing
RAJ-362	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCTTCTCCTTCTGCTGTG CGGCAACAGATC	Forward, from start of *NtEF-1α* (TA2602), *att*B1
RAJ-363	GGGGACCACTTTGTACAAGAAAGCTGGGTGCCACACAAATGACATGG CAGGTTTAGC	Reverse, from end of *NtEF-1α* (TA2602), *att*B2
RAJ-370	CCAGACCCGGGTTACACCACAATATATCCTG	Forward, pGreenII *EF-LUC*-Term, *Xma*I-introducing
RAJ-371	TGGTGTAACCCGGGTCTGGATTTTAGTACTGG	Reverse, pGreenII *EF-LUC*-Term, *Xma*I-introducing
RAJ-380	AAAAAGCAGGCTAATGTCGACGTGGCATTTATTTTCC	Forward, amplification of the *AtBRCA2* -IV (cds) gene
RAJ-381	AGAAAGCTGGGTTCAGCATGAAGGTGATTTACAAGCACC	Reverse, amplification of the *AtBRCA2* -IV (cds) gene

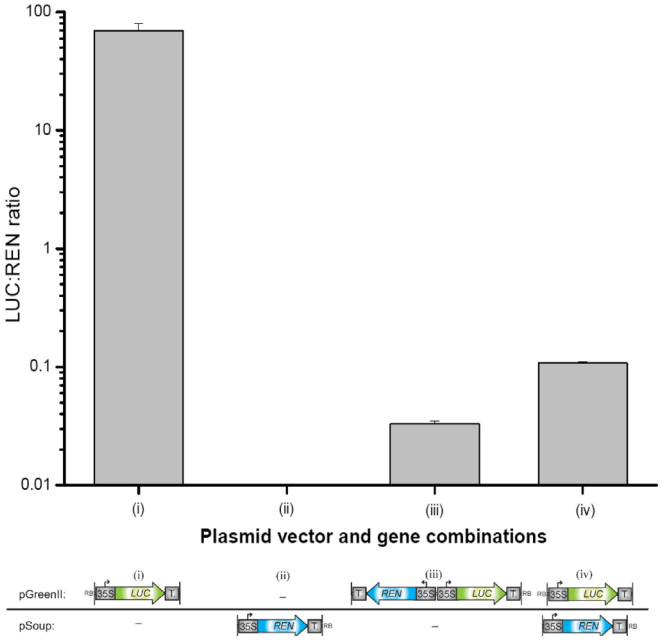
Continued on next page

Additional File 1. Continued.

Name	Sequence (5' - 3')	Application
RAJ-385	CGAATTCCCATGGATTACACGGCGATC	Reverse, from the end of *LUC* with *Nco*I
RAJ-387	CTCAACACCATGGGCATTTCGCAG	Forward, from start of *LUC* with *Nco*I
RAJ-389	CTCACTATAGGGCGGCCGCGGTACCCCTACTCC	Forward, from start of CaMV 35S enhancers on pGreenII 1598-6 with *Not*I
RAJ-390	CGAAGTCATGGGCCCTAGCTAGTGCATGCATATGTG	Reverse, from end of CaMV 35S enhancers on pGreenII 1598-6 with *PspOMI*
RAJ-393	CTAAAATCCAGACCATGGCAGGTATAACATGTC	Forward, from pGreenII *EF-LUC*-Term-*EF*-35S with *Nco*I
RAJ-394	GTTATACCTGCCATGGTCTGGATTTTAGTACTGG	Reverse, from pGreenII *EF-LUC*-Term-*EF*-35S with *Nco*I
RAJ-429	ACCTGCAGGCCCCGGGTCGACGAATTAATTCCAATCC	Forward, from the start of the CaMV 35S promoter on pHEX2
RPH-059	ATGGAAGACGCCAAAAACATAAAGAAAG	Forward, amplification from the start of *LUC*, with start codon
RPH-060	GATACATGAGAATTAAGGGAGTCACGT	Forward, amplification from the *NOS* promoter
RPH-061	AGCTTGCATGCCGGTCGATCTAGTAAC	Reverse, amplification from the *NOS* terminator
RPH-062	CGTACCCCTACTCCAAAAATGTCAAAG	Forward, amplification from start of the CaMV 35S promoter
RPH-063	CGATCTGGATTTTAGTACTGGATTTTG	Reverse, amplification from the CaMV terminator
RPH-149	GTAATACGACTCACTATAGGGC	GW Adapter Primer 1
RPH-150	ACTATAGGGCACGCGTGGT	GW Adapter Primer 2
RPH-249	GTCTCAATAGCCCTTTGGTCTTCTGAG	Reverse, amplification from the CaMV 35S promoter
RPH-428	CTCTTCTTCAAATCTATACATTAAGACG	Reverse, amplification from middle of *LUC*

Amplifications of effector genes generally used the Adapter-PCR amplification method as described by Invitrogen with the adapter-specific primers RAJ-245 and RAJ-246, although when this method proved unsuccessful, amplifications used *att*B site-containing gene-specific primers.

Additional File 2. Luciferase gene activity in transfected *Nicotiana benthamiana* plant leaves, for different combinations of transient *LUC* and *REN* gene expression vectors. Figure showing the Luciferase activity that was measured as part of testing the normalization vector pSoup:*REN* (pSoup 0800). Measurements were made after an incubation of 72 h. Error bars are standard error. i) pGreenII 0579-1 and pSoup; ii) pGreenII 0000 and pSoup:*REN*; iii) pGreenII 0800-*LUC*+ and pSoup 0000; iv) pGreenII 0579-1 and pSoup. RB = right T-DNA border. The vertical black lines show T-DNA borders. '35S' is a CaMV 35S promoter, and 'T.' is its corresponding terminator.



Luciferase data complementing the above figure.

Label	pGreenII vector	pSoup vector	LUC	REN
(i)	pGreenII 0579-1	pSoup	473.4	4.457
			515.6	10.98
			423.4	5.461
			315.4	6.357
			330.2	3.788
			357.3	7.324
(ii)	pGreenII 0000	pSoup:REN	0.038	10105
			0.043	9799
			0.05	9908
			0.054	9982
			0.048	9123
			0.044	8368
(iii)	pGreenII 0800-LUC+	pSoup	282.3	8892
			239.3	7715
			195.4	8351
			353.6	9179
			368.5	9764
			368.7	10308
(iv)	pGreenII 0579-1	pSoup:REN	1186	11019
			1222	11376
			861.8	8854
			1416	12268
			1073	9813
			1259	11927

Additional File 3. Intron sequences obtained from tobacco *EF-1a* gene members.

Table showing the molecular size of intron sequences in both the 5' UTR and the codon sequence of tobacco *EF-1a* genes.

	Transcript assembly number [52]	5' UTR intron size (bp)	Codon sequence intron size (bp)
A	TA2598	900	79
B	TA2600	791	86
C	TA2599	765	78
D	TA2602	951	76
E	TA2597	932	92

Additional File 4. Raw data.

A. Table showing data that were obtained for 14 candidate genes regarding their impact on frequencies of homologous recombination, as shown in Figure 3.

Effector gene	#	LUC	REN	LUC:REN Ratio	Average LUC:REN	SE	Average LUC	Average REN
GUS[a]	1	1.826	6686	0.0002731				
	2	11.98	11781	0.001017				
	3	4.018	8318	0.0004831				
	4	11.17	10435	0.001070				
	5	10.81	11013	0.0009816				
	6	9.752	11086	0.0008797				
	7	8.293	9070	0.0009143				
	8	10.25	12655	0.0008100	0.0008036	9.916×10^{-5}	8.512	10130
ScSPO11[a]	1	3.945	8399	0.0004679				
	2	7.288	8772	0.0008308				
	3	2.851	4801	0.0005938				
	4	10.31	10201	0.001011				
	5	6.116	7922	0.0007720				
	6	7.943	6539	0.001215	0.0008153	0.0001214	6.409	7772
AtSNM1[a]	1	1.163	1079	0.001078				
	2	2.946	1741	0.001692				
	3	2.637	1828	0.001443				
	4	2.543	2089	0.001217				
	5	6.706	4757	0.001410				
ScRAD52[b]	1	7.241	5321	0.001361				
	2	4.069	2917	0.001395				
	3	5.367	3735	0.001437				
	4	9.995	7068	0.001414				
	5	12.96	9224	0.001405				
	6	4.579	3786	0.001209	0.001370	3.695×10^{-5}	7.369	5342
ScRAD54[b]	1	6.662	5490	0.001213				
	2	9.417	7316	0.001287				
	3	8.722	6187	0.00141				
	4	7.539	6042	0.001248				
	5	10.86	8141	0.001334				
	6	11.49	8790	0.001307	0.001230	3.074×10^{-5}	9.115	6994
AtCHR24[a]	1	1.461	6037	0.0002420				
	2	26	12800	0.002031				
	3	11.54	4777	0.002416				
	4	16.66	8193	0.002033				
	5	12.59	6725	0.001872				
	6	16.42	8706	0.001886				
	7	11.39	6169	0.001846	0.002014	9.522×10^{-5}	13.72	7630
ScMIM[a]	1	8.547	4375	0.001954				
	2	18.88	7390	0.002555				
	3	19.9	6269	0.003174				
	4	20.14	6433	0.003131				
	5	11.24	5206	0.002159				
	6	17.61	6735	0.002615				
	7	18.57	7535	0.002464	0.002579	0.0001857	16.41	6278

Continued on next page

Additional File 4. Continued.

Effector gene	#	LUC	REN	LUC:REN Ratio	Average LUC:REN	SE	Average LUC	Average REN
ScINO80 [a]	1	7.322	8123	0.0009014				
	2	5.171	5710	0.0009056				
	3	1.541	3238	0.0004759				
	4	2.179	4570	0.0004768				
	5	4.732	5230	0.0009048				
	6	3.916	4897	0.0007997	0.0007440	9.447×10^{-5}	4.144	5295
ScRAD51 [c]	1	35.23	8093	0.004353				
	2	13.95	8401	0.001661				
	3	14.8	5894	0.002511				
	4	41.09	9677	0.004246				
	5	23.47	8985	0.002612				
	6	33.27	10428	0.003190				
	7	35.25	11596	0.003040				
	8	22.41	7582	0.002956				
	9	19.47	6068	0.003209	0.003086	0.0002954	26.55	8525
AtRAD51C [a]	1	5.316	10779	0.0004932				
	2	3.408	8030	0.0004244				
	3	4.285	6436	0.0006660				
	4	1.253	4322	0.0002900				
	5	3.082	5382	0.0005730				
	6	7.648	9344	0.0008185				
	7	4.277	6650	0.0006432				
	8	3.607	5836	0.0006181	0.0005657	6.128×10^{-5}	4.110	7097
AtRAD51D [a]	1	5.298	6258	0.0008466				
	2	5.124	4987	0.001027				
	3	6.406	5386	0.001189				
	4	7.865	5179	0.001519				
	5	11.65	5404	0.002156				
	6	9.666	6857	0.001410				
	7	13.13	8729	0.001504	0.001379	0.0001734	8.448	6114
AtXRCC3 [a]	1	3.23	8453	0.000382				
	2	14.35	9145	0.001569				
	3	12.01	11132	0.001079				
	4	20.92	6740	0.003104				
	5	9.498	5950	0.001596				
	6	8.221	6391	0.001286				
	7	14.14	9317	0.001518				
	8	12.03	7974	0.001509				
	9	7.68	5488	0.001399				
	10	8.356	5807	0.001439	0.001488	0.0002242	11.04	7640
AtBRCA2-IV [a]	1	5.925	8632	0.0006864				
	2	4.382	10636	0.0004120				
	3	14.11	10393	0.001358				
	4	1.599	7332	0.0002181				
	5	2.327	7243	0.0003213				
	6	6.369	11238	0.0005667				
	7	3.047	5963	0.0005110				
	8	6.003	10755	0.0005582				
	9	6.731	11555	0.0005825	0.0005793	0.0001151	5.610	9305
AtDSS1-I [a]	1	1.126	6517	0.0001728				
	2	4.79	8013	0.0005978				
	3	2.267	5531	0.0004099				
	4	6.489	9027	0.0007188				
	5	2.83	5327	0.0005313				
	6	5.051	8583	0.0005885				
	7	3.492	6342	0.0005506				
	8	2.823	5046	0.0005595	0.0005161	6.152×10^{-5}	3.609	6798
ScCCE1 [a]	1	3.201	7920	0.0004042				
	2	1.816	5335	0.0003404				
	3	1.772	4318	0.0004104				
	4	1.969	3297	0.0005972				
	5	13.54	10257	0.001320				

Continued on next page

Additional File 4. Continued.

Effector gene	#	LUC	REN	LUC:REN Ratio	Average LUC:REN	SE	Average LUC	Average REN
	6	4.369	6173	0.0007077				
	7	4.088	6458	0.0006330				
	8	3.023	5216	0.0005796	0.0006241	0.0001168	4.222	6122
GUS[b]	1	8.9	10049	0.0008856				
	2	5.691	6691	0.0008506				
	3	7.274	8096	0.0008985				
	4	10.82	10987	0.0009848				
	5	8.49	9013	0.0009420				
	6	6.001	6394	0.0009385	0.0009167	2.138×10^{-5}	7.863	8538
GUS[c]	1	15.71	10569	0.001486				
	2	24.83	9594	0.002588				
	3	9.366	7234	0.001295				
	4	8.771	4223	0.002077				
	5	12.91	5912	0.002184				
	6	19.32	7736	0.002497				
	7	12.13	6135	0.001977				
	8	16.81	8213	0.002047				
	9	10.48	4783	0.002191				
	10	15.35	7639	0.002009	0.002035	0.0001325	14.57	7204

[a], [b] and [c] indicate the measurements of the *GUS* control gene that correspond to the measurements of each effector gene.

B. Table showing data that were obtained for eight candidate genes regarding their impact on frequencies of non-homologous recombination, as shown in Figure 3.

Effector gene	#	LUC	REN	LUC:REN ratio	Average LUC:REN	SE	Average LUC	Average REN
GUS[a]	1	0.425	5490	7.741×10^{-5}				
	2	0.414	8251	5.018×10^{-5}				
	3	0.427	7068	6.041×10^{-5}				
	4	0.464	11189	4.147×10^{-5}				
	5	0.271	6823	3.972×10^{-5}				
	6	0.331	9127	3.627×10^{-5}	4.627×10^{-5}	5.019×10^{-6}	0.3783	8547
AtSNM1[a]	1	0.4	2085	0.0001919				
	2	0.16	1052	0.0001521				
	3	0.452	1970	0.0002294				
	4	0.152	1940	7.835×10^{-5}				
	5	0.348	3079	0.0001130				
	6	0.713	6245	0.0001142				
	7	0.392	2312	0.0001696				
	8	0.441	5220	8.448×10^{-5}	0.0001367	1.439×10^{-5}	0.4434	3505
ScRAD52[b]	1	0.551	6984	7.890×10^{-5}				
	2	0.566	6284	9.007×10^{-5}				
	3	0.453	5381	8.419×10^{-5}				
	4	0.677	7948	8.518×10^{-5}				
	5	0.621	7421	8.368×10^{-5}				
	6	0.707	8332	8.485×10^{-5}				
	7	0.729	8112	8.987×10^{-5}				
	8	0.49	10310	4.753×10^{-5}	8.053×10^{-5}	5.217×10^{-6}	0.5993	7597
ScRAD54[b]	1	0.581	5336	0.0001089				
	2	0.694	6834	0.0001016				
	3	1.063	8963	0.0001186				
	4	1.045	9286	0.0001125				
	5	0.98	8450	0.0001160				
	6	1.145	9825	0.0001165				
	7	1.011	9029	0.0001120				
	8	0.831	7564	0.0001099	0.0001120	2.039×10^{-6}	0.9188	8161
AtCHR24[c]	1	1.515	11022	0.0001375				
	2	3.072	11781	0.0002608				
	3	1.528	6495	0.0002353				
	4	0.829	8617	9.621×10^{-5}				

Continued on next page

Additional File 4B. Continued.

Effector gene	#	LUC	REN	LUC:REN ratio	Average LUC:REN	SE	Average LUC	Average REN
	5	1.633	9462	0.0001726				
	6	1.651	9415	0.0001754				
	7	1.599	9533	0.0001677	0.0001779	2.273×10^{-5}	1.690	9475
ScMIM[c]	1	0.709	6743	0.0001052				
	2	1.291	9408	0.0001372				
	3	1.664	8969	0.0001855				
	4	1.406	8066	0.0001743				
	5	1.548	9070	0.0001707				
	6	1.645	9099	0.0001808				
	7	1.373	8694	0.0001579				
	8	0.82	6417	0.0001278				
	9	1.249	7393	0.0001689	0.0001565	9.635×10^{-6}	1.414	8266
ScRAD51[d]	1	1.087	6487	0.0001676				
	2	1.136	7409	0.0001533				
	3	1.219	6353	0.0001919				
	4	2.115	9487	0.0002229				
	5	2.105	9475	0.0002222				
	6	2.166	9749	0.0002222				
	7	1.264	6894	0.0001834				
	8	1.116	6680	0.0001671				
	9	1.463	7629	0.0001918	0.0001914	9.323×10^{-6}	1.519	7796
AtRAD51D[c]	1	2.881	9912	0.0002907				
	2	2.068	8233	0.0002512				
	3	1.439	10521	0.0001368				
	4	2.286	9241	0.0002474				
	5	1.713	8310	0.0002061				
	6	1.644	7757	0.0002119				
	7	2.085	9801	0.0002127				
	8	1.998	9813	0.0002036	0.0002200	1.704×10^{-5}	2.014	9199
AtXRCC3[a]	1	1.599	4369	0.0003660				
	2	3.548	9573	0.0003706				
	3	3.068	8439	0.0003636				
	4	2.131	6328	0.0003368				
	5	2.682	7395	0.0003627				
	6	1.999	5928	0.0003372	0.0003561	6.748×10^{-6}	2.505	7005
GUS[b]	1	0.75	8783	8.539×10^{-5}				
	2	0.626	7259	8.624×10^{-5}				
	3	0.761	10577	7.195×10^{-5}				
	4	0.541	7428	7.283×10^{-5}				
	5	0.731	7063	0.0001035				
	6	0.659	7291	9.039×10^{-5}				
	7	0.521	7613	6.844×10^{-5}				
	8	0.632	7985	7.915×10^{-5}				
	9	0.651	7370	8.833×10^{-5}				
	10	0.798	9061	8.807×10^{-5}	8.343×10^{-5}	3.506×10^{-6}	0.6670	8043
GUS[c]	1	0.629	9384	6.703×10^{-5}				
	2	0.806	10686	7.543×10^{-5}				
	3	0.624	6815	9.156×10^{-5}				
	4	0.99	12184	8.125×10^{-5}				
	5	1.79	8552	0.0002093				
	6	1.967	9140	0.0002152				
	7	0.814	7567	0.0001076				
	8	1.193	9217	0.0001294				
	9	1.315	10195	0.0001290				
	10	1.082	9761	0.0001109				
	11	1.045	9106	0.0001148	0.0001210	1.567×10^{-5}	1.114	9328
GUS[d]	1	2.124	10334	0.0002055				
	2	2.029	10050	0.0002019				
	3	2.17	9331	0.0002326				

Continued on next page

Additional File 4B. Continued.

Effector gene	#	LUC	REN	LUC:REN ratio	Average LUC:REN	SE	Average LUC	Average REN
	4	1.569	7603	0.0002064				
	5	1.647	8003	0.0002058				
	6	1.231	6522	0.0001888				
	7	1.637	8276	0.0001978				
	8	2.13	9100	0.0002341				
	9	1.563	7835	0.0001995				
	10	2.304	10871	0.0002119	0.0002084	4.836×10^{-6}	1.840	8793

[a], [b], [c], and [d] indicate the measurements of the *GUS* control gene that correspond to the measurements of each effector gene.

C. Table showing data that were obtained for two candidate genes regarding their impact on frequencies of non-homologous recombination, as shown in Figure 5.

Effector gene	#	LUC	REN	LUC:REN ratio	Average LUC:REN	SE	Average LUC	Average REN
GUS	1	0.442	5234	8.445×10^{-5}				
	2	0.32	4884	6.552×10^{-5}				
	3	0.315	5090	6.189×10^{-5}				
	4	0.316	4323	7.310×10^{-5}				
	5	0.342	4292	7.968×10^{-5}				
	6	0.46	6268	7.339×10^{-5}				
	7	0.427	5837	7.315×10^{-5}				
	8	0.41	6069	6.756×10^{-5}	7.234×10^{-5}	2.790×10^{-6}	0.3790	5250
SNM1	1	0.655	6225	0.0001052				
	2	0.473	6538	7.235×10^{-5}				
	3	0.733	7867	9.317×10^{-5}				
	4	0.714	7803	9.150×10^{-5}				
	5	0.567	5399	0.0001050				
	6	0.55	7555	7.280×10^{-5}				
	7	0.556	6024	9.230×10^{-5}	9.034×10^{-5}	5.481×10^{-6}	0.6069	6773
XRCC3	1	0.687	6101	0.0001126				
	2	0.596	5617	0.0001061				
	3	0.651	5196	0.0001253				
	4	0.675	5513	0.0001224				
	5	0.748	5848	0.0001279				
	6	0.717	6754	0.0001062				
	7	0.691	6597	0.0001047	0.0001150	4.069×10^{-6}	0.6807	5947

SE = standard error.