Transgenic Research **14:** 1–14, 2005.

Review

Enhancing gene targeting efficiency in higher plants: rice is on the move

Olivier Cotsaftis* & Emmanuel Guiderdoni
UMR1096 PIA, Biotrop Program, Cirad-Amis, Avenue Agropolis, F-34398 Montpellier Cedex 5, France

Received 27 March 2004; revised 25 June 2004; accepted 28 June 2004

Key words: allele replacement, homologous recombination, illegitimate recombination, knockout, positive and negative selection system, recombination protein

Abstract

Meeting the challenge of routine gene targeting (GT) in higher plants is of crucial interest to researchers and plant breeders who are currently in need of a powerful tool to specifically modify a given locus in a genome. Higher plants have long been considered the last lineage resistant to targeting technology. However, a recent report described an efficient method of T-DNA-mediated targeted disruption of a non-selectable locus in rice [Terada et al., *Nat Biotechnol* **20**: 1030–1034 (2002)]. Though this study was an obvious breakthrough, further improvement of GT frequencies may derive from a better understanding of the natural mechanisms that control homologous recombination (HR) processes. In this review, we will focus on what is known about HR and the factors which may hamper the development of routine GT by HR in higher plants. We will also present the current strategies envisaged to overcome these limitations, such as expression of recombination proteins and refinements in the design of the transformation vector.

Introduction

The plant science community currently suffers from a strange paradox; while tremendous efforts are being invested in the genome sequencing of higher plants, knowledge of gene function remains limited. In rice (*Oryza sativa*), the staple food crop for a vast majority of the world's population and model species for monocotyledonous and cereal plants, draft genome versions of the two agriculturally important sub-species, *indica* and *japonica*, were published in April 2002 (Goff et al., 2002, Yu et al., 2002). Thanks to the International Rice Genome Sequencing Project (IRGSP at http://rgp.dna.affrc.go.jp/IRGSP/), a highly accurate sequence of the *japonica's* cv. Nipponbare 12 chromosomes will also be completed by the end of 2004 (Delseny, 2003). However, of the 60,000 computer-predicted genes in the rice genome, roughly half of them have been assigned an uncertain role on the basis of their sequence while only about 100 have been ascribed a precise and verified function (Cyranoski, 2003).

In an attempt to assign a precise function to a given gene, large insertional mutant libraries have been developed in higher plants over the last decade. Insertional mutagenesis, involving the random insertion of transposable elements or T-DNA into the host genome to act as a molecular tag of the interrupted gene, has proven to be a valuable tool to address gene function through detailed analysis of mutant phenotypes. Systematic isolation and sequencing of genomic DNA flanking the insertion sites (known as FSTs or Flanking Sequence Tag) offers the opportunity to rapidly characterise plants altered in a candidate

*Author for correspondence:
E-mail: cotsaftis@cirad.fr

gene sequence. This approach is notably most useful in fully sequenced genomes such as in *Arabidopsis thaliana* (Krysan et al., 2002; Marsch-Martinez et al., 2002; Raina et al., 2002; Sessions et al., 2002; Alonso et al., 2003), in which approximately 296,000 FSTs are now publicly available, thus nearly saturating the genome with insertion sites (March 15th 2004 update, http://www.arabidopsis.org/links/insertion.jsp).

With 125 Mbp and 26,422 genes, the *Arabidopsis* genome shows very limited synteny with the 420–466 Mbp and 60,000 predicted genes of the rice genome (Schoof & Karlowski, 2003). In contrast, almost all cereal genes are present in rice with a highly conserved macro-colinearity between genomes (Ware & Stein, 2003). For these reasons, there has been a need to specifically develop mutant libraries in rice. Chemically or physically mutagenized populations harbouring alterations from point mutations up to several kilobase deletions have been created (Leung et al., 2001) and their use for PCR-based identification in DNA pools of a mutant line altered in a target candidate sequence has been undertaken using the recently developed TILLING (Till et al., 2003) or Delete-a-Gene (Li et al., 2001) methodologies. In parallel, several T-DNA (Jeon et al., 2000; Jeong et al., 2002; An et al., 2003; Wu et al., 2003; Sallaud et al., 2004), transposon (Izawa et al., 1997; Chin et al., 1999; Greco et al., 2001, 2003; Kohli et al., 2001; Uppadhyaya et al., 2002; Kolesnik et al., 2004) and retrotransposon (Hirochika, 2001; Miyao et al., 2003) tagging populations have been generated. FST databases are now emerging in several countries and will be accessible through a common portal at http://www.iris.irri.org/IRFGC/.

Despite their widespread use, libraries generated by chemical, physical or insertion mutagenesis also have drawbacks that can be common to all mutagens or specific to a given type. Since mutagenesis relies on the random creation of lesions throughout the genome followed by screening for a specific mutation, large collections of lines have to be generated and propagated in order to increase the chance of disrupting a particular gene. This fact, coupled with the growth characteristics of the rice plant, means seed increase has to be accomplished under field conditions, which may be difficult for transgenic materials. Considering that the number of lesions created in a single plant is variable, an observed mutant phenotype may not be correlated to a labelled tag or could be the result of several lesions. Backcrosses with a wild type parent are then often needed to clean up the genetic background from extra lesions and may prove to be tedious, especially in chemical and physical mutagenesis.

To bypass these limitations, the ability to target DNA integration at a given locus is of major interest to functional genomic projects. It would allow specific modification or disruption of endogenous genes, providing a tool for more detailed analysis of gene function. Such a technique would also permit the locus-specific integration of a transgene into a predetermined site of the host genome, avoiding the accidental inactivation of an endogenous gene localised at the insertion site or the unexpected expression profiles of the transgene itself, the so-called position effect (reviewed by Mengiste & Paszkowski, 1999; Vergunst & Hooykaas, 1999; Britt & May, 2003; Hanin & Paszkowski, 2003). Furthermore, one can expect an enormous effort in the next decade in sequencing alleles at loci (coding sequences but also regulatory regions) of interest among genetic resources of crop plants, as well as in establishing association between allele version and the agronomic value of a given trait. From this perspective, developing a method allowing precise and targeted allele replacement in cultivars may become a molecular breeding method which would considerably enhance plant improvement efficiency.

Unfortunately, the natural propensity of DNA to recombine within homologous areas of the genome (i.e., homologous recombination or HR) and therefore direct DNA integration (i.e., gene targeting or GT) happens rarely in higher eukaryotes. Currently, this technique is applied routinely to organisms whose genomes preferentially use HR to mediate DNA integration, such as prokaryotes (Weller et al., 2002), yeast (reviewed by Pâques & Haber, 1999) and the moss *Physcomitrella patens* (reviewed by Schaefer, 2001, 2002). The cytoplasmic genomes of higher eukaryotes are also amenable to targeted modifications due to their prokaryotic origins (Ruf et al., 2001; Inoue et al., 2002). On the contrary, preferred pathways have to be bypassed in higher eukaryote nuclear genomes where DNA mostly integrates in a random manner, even with long stretches of homology shared within genomic template. So far, the use of

stem cells and positive/negative selection systems as well as the development of cloning technology have permitted GT in mice and other mammalian species whereas the recent success of nuclear transfer in zebrafish has been a first step towards gene knockouts in fish (Bradey et al., 1984; Lee et al., 2002; reviewed by Sedivy & Dutriaux, 1999; Clark et al., 2000; Dennin & Priddle, 2003). Moreover, an alternative to the traditional cell culture based approach has been developed in *Drosophila melanogaster* involving a GT system which operates within the cells of intact animals (Rong & Golic, 2000; reviewed by Gong & Rong, 2003).

In higher plants, since the first report of targeted reconstruction of a transgene in the tobacco genome (Paszkowski et al., 1988), GT events have been described at several loci but only in dicotyledonous plants and always with a very low frequency ranging from 10^{-3} to 10^{-6} (Lee et al., 1990; Offringa et al., 1990; Miao & Lam, 1995; Risseeuw et al., 1995, 1997; Kempin et al., 1997; Reiss et al., 2000; Hanin et al., 2001). However, a recent report described an efficient method of T-DNA-mediated targeted disruption of a non-selectable locus (the *waxy* gene) in rice based on a large homologous region of 13 kb and highly expressed, duplicated diphtheria toxin gene-based counter-selectable markers (Terada et al., 2002; reviewed by lida & Terada, 2004). The GT events represented roughly 1 GT event in 1500 potential transformants and 1 GT event for 99 escapes, resulting in an observable frequency of 1%.

Although this method may be applicable to other genomic loci and transferable to other plant species for which well-established transformation systems already exist, a better understanding of the natural mechanisms that trigger HR processes may improve GT frequencies and may reveal why there is such a difference in GT success rate among organisms. In this review, we will first focus on what is known about HR and the inefficiency of GT and then on strategies for enhancing GT frequencies in higher plants.

Natural features of somatic homologous recombination

HR is most commonly known as the process resulting in DNA recombination during meiosis. In mitotic cells, however, HR is a basic mechanism to repair DNA damage and more precisely to repair DNA double-strand breaks (DSBs) (reviewed by Britt, 1999; Tuteja et al., 2001). If left unrepaired, DNA DSBs can affect cellular metabolism and lead the cell through anarchic development or even to cell death. Nevertheless, DSBs sometimes play a crucial role in biological processes such as DNA replication (Barnes & Rio, 1997) or meiosis, through the activity of the Spo11 protein (Smith & Nicolas, 1998; Grelon et al., 2001). The vital rearrangement of the V(D)J region for antibody production in vertebrate immune systems is also processed *via* DSB formations (Roth, 2003).

While HR mediates a precise repair by the copying of homologous templates (e.g., homologous chromosomes or sister-chromatids), potentially lethal DNA DSBs can also be repaired by illegitimate recombination (IR) *via* non-homologous end joining (NHEJ), where essentially any two ends can fuse (reviewed by Puchta & Hohn, 1996; Gorbunova & Levy, 1999; van den Bosch et al., 2002; Pastwa & Blasiak, 2003; Dudas & Chovanec, 2004) (Figure 1). Thus, this random

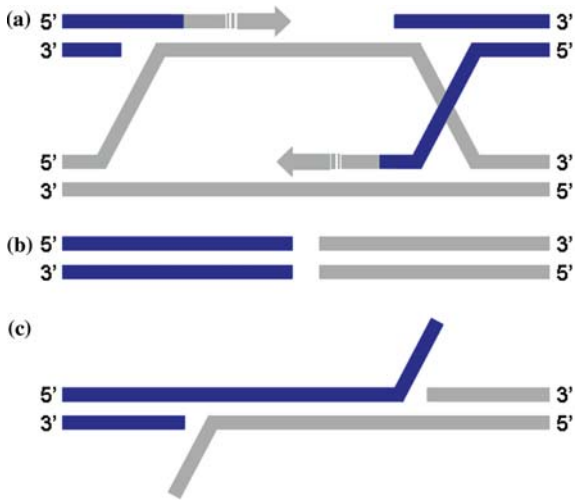

Figure 1. The three different DSB repair pathways. Repair by HR (a): the break is repaired by copying the information on a homologous template (in gray) resulting in the retrieval of the original sequence. Repair by IR via non-homologous end joining (b): the break is repaired by simple ligation of two double-stranded DNA ends. Basically, any two ends can fuse, a process leading to frequent genome rearrangements, Repair by SSA (c): the break is repaired by the annealing of complementary 5′-protruding sequences located on the DNA ends. As in IR, the repair is said to be non conservative as there is a loss of genetic information. Each color represents a distinct DNA molecule.

rejoining of broken ends is often associated with chromosomal instability and a loss of genetic information. HR-mediated repair between non-allelic loci can also lead to genome rearrangements but spontaneous recombination between ectopic sites is considered to be quite rare in plants and animals (Puchta, 1999). Single-strand annealing (SSA), a third path of repair, requires the presence of repeated sequences on both sides of a break (Figure 1). After exonuclease degradation of the 5′-ends, repair occurs by annealing of the two complementary sequences, a process leading to the loss of the genetic information contained between these repeats.

With respect to the species preferential DSB repair pathway, HR but also IR mediates transgene integration. This second aspect explains the inefficiency of GT in higher plants (see above) which use HR as a minor pathway of repair. Thus, despite the fact that transgene integration processes are still unclear in plants (reviewed by Makarevitch et al., 2003; Kohli et al., 2003; Tzfira & Citovsky, 2003; Somers & Makarevitch, 2004), transgenic DNA would be preferentially integrated by end joining whether or not sharing homology within the host genome.

Homologous recombination molecular mechanisms

Several distinct pathways have been described for the repair of DSBs and/or transgene integration. However, the genomic rearrangements often observed in higher plants after these processes have occurred can only be explained by a combination of different mechanisms common to HR and IR. Thus, a description of recombination mechanisms as a whole is necessary. Moreover, as the published material is mainly focused on DNA repair, recombination mechanisms will be described in this section as DSB repair events.

After a DSB occurs within a genome, two options exist for the newly-formed double-stranded ends: if not simply ligated to another end, free 3′-ends are formed via exonuclease-catalyzed digestion of the released 5′-ends. The simple ligation of double-stranded ends without single-strand formation is however considered to be rare in plants (Gorbunova & Levy, 1999). For example, a recent study reported 7 simple ligations of T-DNA ends out of 67 plant genomic/T-DNA junctions analysed (Windels et al., 2003). Other reports proved that NHEJ-mediated repair is commonly achieved through the annealing of 3–5 bp micro-homology localised on both extremities involved in the repair (Takano et al., 1997; Brunaud et al., 2002; Windels et al., 2003). One difference between NHEJ and SSA in plants would then be the length of homology: whereas NEHJ is carried out at micro-homological sites within a few bp, SSA requires larger size of homology; the common feature being the loss of the genetic information contained between the repeats. Nevertheless, SSA-mediated repair was proposed in higher plants to explain extrachromosomal (Puchta & Hohn, 1991) as well as intrachromosomal HR events at closely linked chromosomal direct repeats (Lowe et al., 1992; Siebert & Puchta, 2002), but not GT.

Often associated with deletions, rejoined breaks in plants are also prone to DNA insertion at the repaired site (Gorbunova & Levy, 1997; Takano et al., 1997; Salomon & Puchta, 1998; Kirik et al., 2000; Chilton & Que, 2003; Tzifira et al., 2003; Windels et al., 2003). This frequent outcome was first demonstrated in the tobacco genome where up to 1.2 kb of filler DNA was found after sequence analysis of end junctions (Gorbunova & Levy, 1997). Introduction of filler DNA can be best explained by the Synthesis-Dependent Strand Annealing (SDSA) mechanism where DSB-induced free 3'-termini invade a double-stranded template and prime DNA synthesis, generating long stretches of single-stranded DNA (Lin et al., 1984). After displacement of the molecule from the template, end joining or SSA then process the repair, resulting in the insertion of newly synthesized DNA at the junction. Template switching during DNA synthesis may be responsible for the complex DNA pattern often obtained at the repaired site. However, filler DNA can also be explained by IR-mediated insertion of extrachromosomal DNA such as transposable elements and transgenes (Salomon & Puchta, 1998; Chilton & Que, 2003; Tzifira et al., 2003) or, for insertions of a few base pairs, by a misannealing in end joining at micro-homological sites (Windels et al., 2003).

Similar to SDSA is the One-Sided Invasion (OSI) mechanism, which describe events where only one single-stranded 3′-end invades a template (Belmaaza & Chartrand, 1994). Thus, if a

3′-termini invades a homologous template and synthesis proceeds without a template switch, OSI or SDSA-mediated repair can lead to a HR event (Puchta, 1998) (Figure 2). Moreover, since the two DNA ends of a targeting vector can react independently, it was also proposed to explain some observed outcomes of GT experiments that one end would serve in a OSI-mediated HR event while the other end would integrate by NHEJ (Hohn & Puchta, 2003).

The HR events can be also explained through the classical DSB repair (DSBR) model of Szostak (Szostak et al., 1983) (Figure 2). In this model, both single-stranded 3′-ends invade a homologous template and prime DNA synthesis prior to the formation of two Holliday junctions. Depending on their cleavage orientation, Holliday junctions can then be resolved by either gene-conversion, where the two DNA molecules retain their flanking sequences, or by crossing-over, where they exchange them. Alternatively, displacement of the newly synthesized strands from the homologous template before Holliday junction formation can lead to a SDSA-mediated

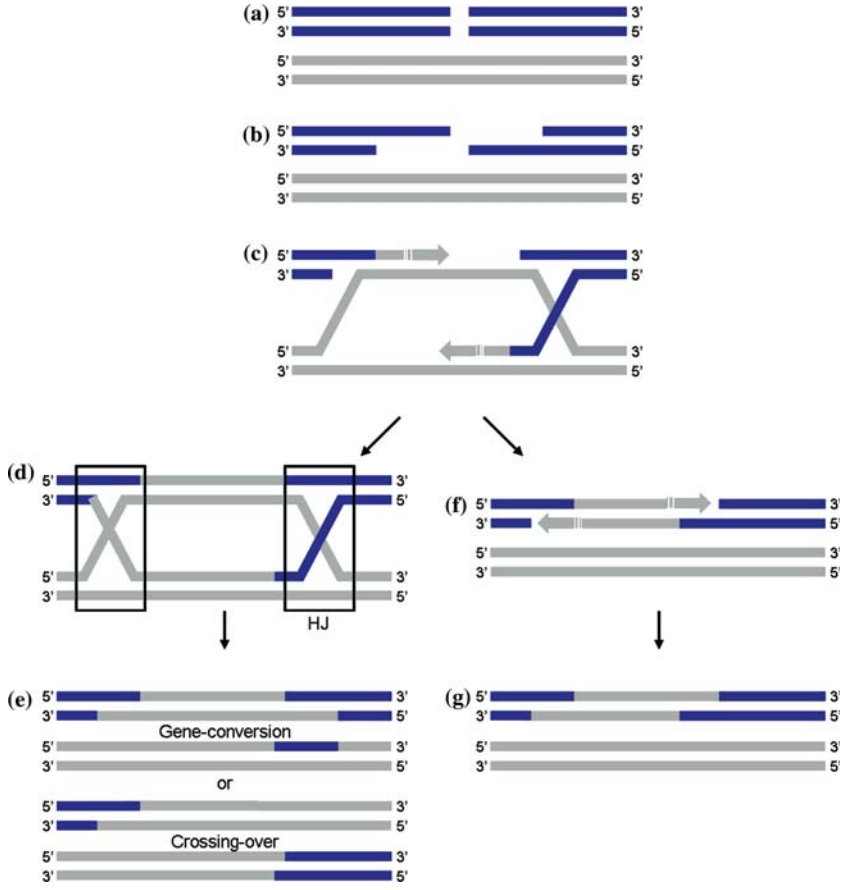

Figure 2. HR-mediated DSB repair mechanisms *via* the DSBR model of Szostak or SDSA/OSI. When a DNA DSB occurs within a genome (a). If not simply ligated to another end, free 3′-ends are formed via exonuclease-catalyzed digestion of the released 5′-ends (b). These 3′-protruding ends are then used as primers for DNA synthesis on a homologous template such as a homologous chromosome, a sister chromatid or any DNA sequence available sharing homology (c). In the DSBR model of Szostak, Holliday junctions (HJ) are formed (d). Resolved by either gene-conversion, where the two DNA molecules retain their flanking sequences, or by crossing-over, where they exchange them (e). Alternatively, displacement of the newly synthesized strands from the homologous template prior to HJ formation and without re-annealing on a non-homologous template leads to a SDSA or OSI mechanism depending on whether both or only one single stranded 3′-end(s), respectively, is(are) involved in the repair process (f). The break is then repaired by IR, resulting in the retrieval of the original sequence in the broken molecule and no changes of the homologous template (g). No genetic information is lost during these processes. The broken DNA molecule is represented in blue whereas the homologous template is in gray.

HR event (Prado & Aguilera, 2003). The DSBR model associated with crossing-over was proposed as a possibility to explain gene knockouts.

In summary, the literature shows that DSB repair is an error-prone mechanism in higher plants. The prevalence of IR-mediated repair almost inevitably produces deletions which are often associated with DNA insertions (such as transgenes) and chromosomal rearrangements. Mechanistically, end joining at micro-homological sites, SSA and SDSA/OSI appear to be the prominent paths of repair. HR-mediated DSB repair is therefore a minor process (Puchta, 1999; Gisler et al., 2002) that can be explained either by SDSA/OSI or DSBR (Puchta et al., 1996). It is possible that SDSA/OSI functions as a HR repair in somatic cells, whereas DSBR occurs mostly during meiosis, when the fidelity of the exchange is assured by the pairing of homologs (Zickler & Kleckner, 1999). Moreover, SDSA/OSI-mediated repair may be a safer option than DSBR in somatic cells as it reduces the risks associated with crossover events such as translocations.

Why is GT so inefficient in higher plants?

The consequence of the preferential DSB repair pathway in higher plants is that natural frequencies of GT by HR remain very low, ranging from 10^{-3} to 10^{-6} (Paszkowski et al., 1988; Lee et al., 1990; Offringa et al., 1990; Miao & Lam, 1995; Risseeuw et al., 1995, 1997; Kempin et al., 1997; Reiss et al., 2000; Hanin et al., 2001; Terada et al., 2002). Thus, the main question to address is why the recombination machinery is so efficient in some organisms, such as yeast or the moss *Physcomitrella patens*, but not in higher plants.

Based on the observation that GT frequencies above 1 % in eukaryotic cells seem to be restricted to primitive and/or unicellular organisms (Schaefer, 2001), Schaefer and Zryd (1997) correlated the efficient targeted integration of transgenes in these organisms with their constant or part-time haploid state throughout their life-cycle. Haploid organisms, with only one copy of their genome, would have to repair DSBs by HR in order to maintain genome integrity. However, both haploid and diploid strains of yeast integrate exogenous DNA by HR with as little as 50 bp of homology shared within their genome (reviewed by Pâques & Haber, 1999). Moreover, in the model dicot plant *Arabidopsis*, the common transformation procedure consisting of *Agrobacterium*-infiltration of female gametophytic cells (Bechtold et al., 2000) does not enhance GT frequency (Hanin et al., 2001). Thus, a high ratio of GT vs. random integration might not be a question of ploidy but, perhaps, of cell-type.

Mouse embryonic stem cells or chicken DT 40 pre-B cells are known for their ability to naturally carry out GT with higher frequencies than other animal cell-types (reviewed by Capecchi, 1989; Winding & Berchtold, 2001). As both of these cells can be considered undifferentiated (i.e., they are not the final outcome of the cell lineage), it may be possible that a high frequency of GT is correlated with this so-called undifferentiated state. For instance, in the "Rong and Golic" targeting system developed for *Drosophila* (Rong & Golic, 2000), the difference in the germline targeting efficiency in favour of females could relate to its undifferentiated state; whereas male germ-cells differentiate to start spermatogenesis, female germ-cells remain undifferentiated until oogenesis occurs (reviewed by Lin, 1997). In addition, the *P. patens* protoplasts used for transformation are obtained from a culture of chloronema cells, one of the two cell-types generated once a haploid spore has germinated; chloronema cells differentiate into caulonema from which leafy buds will emerge (Hohe et al., 2004). In multi-cellular organisms, undifferentiated cells would have to repair DSBs by HR in order to preserve genome integrity in the cell-lines, organs or organisms that they will be transformed into whereas differentiated cells would accumulate repair-errors until death without threatening, to a certain limit, the whole organism. On the contrary, DSB repair by HR would be a survival necessity in unicellular organisms. Considering that all of the eukaryote cell-types known to undertake GT are the precursors of cell lineage development (e.g., chicken pre-B cells and *Drosophila* female germline) or of an entire organism (e.g., mice embryonic stem cells and *P. patens* chloronema cells), one can speculate that this particular state is correlated to a high GT frequency. Targeting experiments in microspores or

meristematic tissues may answer this question in higher plants (for a review on plant stem cells, see Byrne et al., 2003).

The connection between cell-cycle phases and HR frequency has also been identified as a condition for efficient GT in certain organisms. In vertebrate cells, it was demonstrated that end joining DSB repair plays a dominant role during G1/early S phase while HR repair is preferentially used in late S/G2 phase (Takata et al., 1998). In yeast, the kinetics of recombination first appeared to be cell-cycle independent (reviewed by Pâques and Haber, 1999) but a recent study showed that IR activity is predominant in G1 (Karathanasis & Wilson, 2002). Furthermore, Reski (1998) proposed that the high ratio of targeted vs. random integration of transgenes in *P. patens* is correlated to the fact that chloronema cells (i.e., the transformed cells) are specifically arrested most often at the G2/M boundary of the cell-cycle (Schween et al., 2003). Cell-cycle arrest in G2 (i.e., after DNA replication) would increase the chance of meeting a homolog, especially in haploid organisms that possess only one copy of their genome. At the same time, this theory would explain the inefficiency of GT in higher plants where cells are known to be predominantly arrested in G1/S.

Can the higher plant protein machinery be responsible for low efficiency?

Over the past two decades, the study of yeast mutants has lead to the characterisation of proteins involved in DNA repair and recombination. These mutants mostly fall into two general classes: those involved primarily in HR, such as *rad51-52-54-55-57-59*, or IR, such as *ku*, *1ig4* and *xrcc4* (reviewed by Pâques & Haber, 1999). The IR pathway in vertebrate cells also requires a DNA protein kinase catalytic subunit for which no homologue has been found in yeast (Pastwa & Blasiak, 2003). However, one protein complex encoded by the *rad50*, *mre11* and *xrs2* genes has been implicated in both repair pathways. All of these proteins seem to be evolutionarily conserved and homologs have been found in several species. For example, the Rad50, Rad51 and Mre11 proteins are highly conserved among eukaryotes and have also been identified in prokaryotes and archaea (Symington, 2002). Thus, although several theories have been speculated to explain the inefficiency of GT in higher plants, only functional analysis of homologs involved in recombinational processes may unravel such an issue (progress on DNA repair and recombination protein in plants has been reviewed by Ray & Langer, 2002).

Surprisingly, while there is strong evidence that Rad52 is essential for any repair by HR in yeast (Sugawara et al., 2003; Wolner et al., 2003; Lisby et al., 2003; reviewed by Symington, 2002), this protein has never been found in plants. Ray and Langer (2002) postulated a correlation between this absence and the low levels of HR commonly obtained in plant somatic cells. However, Rad52 homologs have been discovered in animals, which are known to perform GT at low levels with the exception of some characterised cell-lines (see above). Moreover, homologs of most other proteins involved in HR have been cloned in plants and somatic GT events do occur despite the absence of Rad52 (Paszkowski et al., 1988; Lee et al., 1990; Offringa et al., 1990; Miao & Lam, 1995; Risseeuw et al., 1995, 1997; Kempin et al., 1997; Reiss et al., 2000; Hanin et al., 2001; Terada et al., 2002). It is also possible that Rad52 function is accomplished by an unknown component in plants and animals.

More generally, recent studies tend to contradict the idea that emerged in the late 1990s that HR and IR compete for available DNA ends by the fixation of either the Rad52 protein or the Ku70/80 heterodimer, respectively, to the extremities of a DSB (van Dyck et al., 1999; Haber, 1999). An *in vitro* assay has demonstrated that whereas Ku proteins preferentially bind to free DNA ends, Rad52 preferentially attaches itself to single-stranded DNA but not necessarily at a strand extremity (Ristic et al., 2003). As the authors concluded, it is therefore unlikely that these proteins compete as gatekeepers between HR and IR.

In addition, it was shown in yeast that end joining at micro-homological sites (i.e., the common form of IR in plants, see above) is Ku70 independent (Ma et al., 2003). Similarly, Ku80 seems not to be implicated in T-DNA integration in *Arabidopsis* (Gallego et al., 2003). Ku proteins might then only be required for the simple ligation of double-stranded ends but not for end

joining at micro-homological sites. Moreover, considering that DNA ligase IV, a crucial enzyme for NHEJ T-DNA insertion in yeast, also seems not to be required in *Arabidopsis* (van Attikum et al., 2001, 2003), IR-mediated T-DNA integration mechanisms may be species-specific. T-DNA insertion mechanisms would depend on the prevalence of simple ligation vs. end joining at micro-homological sites to repair a DSB. By contrast, Friesner and Britt (2003) reported that Ku80 and DNA ligase IV mutant plants are defective in T-DNA integration. Thus, this issue is still controversial in plants.

A recent study also demonstrated that even if the Rad50-Mre11-Xrs2 tri-proteic complex is required for random T-DNA integration by IR in yeast (van Attikum et al., 2001), these proteins do not participate in HR-targeted integration events (van Attikum & Hooykaas, 2003). In other words, this means that Rad50, Mre11 and Xrs2, which were thought to be implicated in both recombination processes, are not involved in GT. Other reports have proved the importance of this complex in DNA repair (Daoudal-Cotterell et al., 2002; Trujillo et al., 2003) and the study of the *Arabidopsis rad50* null mutant showed an 8- to 10-fold increase of the mitotic intrachromosomal recombination level (Gherbi et al., 2001). Considering the hypothesis that inhibition of IR would result in a stimulation of HR, mutants of these proteins may increase GT frequency in higher plants. We recently cloned and characterised the homolog of the *rad50* gene in rice cv. Nipponbare (see Genbank accession number AY277897; Cotsaftis and Guiderdoni, unpublished data) and identified a mutant for this gene in the T-DNA insertion line library created in our laboratory (Sallaud et al., 2004). This mutant line is very valuable for the investigation of the role of Rad50 in GT efficiency.

Finally, other proteins have been studied in an attempt to determine the key factors that regulate HR. Markmann-Mulisch et al. (2002) proposed that the unique organisation of the *P. patens rad51* gene (i.e., duplicated and intronless) among multicellular organisms could be related to the high GT frequency observed in this species as introns could serve regulatory functions (see also Ayora et al., 2002). It was also demonstrated that SMC-like proteins (Structural Maintenance of Chromosome) play a role in the HR pathway (Mengiste et al., 1999) and that the over-expression of the MIM gene (i.e., coding for a SMC protein) is associated with higher level of intrachromosomal recombination in *Arabidopsis* (Hanin et al., 2000). Thus, the chromatin structure at the target loci may have a role in GT efficiency.

To summarize, recently published results have provided some new insights into the mechanisms of DNA repair and recombination, but only the isolation and characterisation of plant gene mutants involved in those mechanisms will help to further understand GT processes. Nevertheless, in *P. patens*, DNA molecules integrate 10 times more efficiently by HR (10 in 10^5 cells) than by IR (1 in 10^5 cells) (Schaefer & Zryd, 1997). Thus, the reported absolute HR frequency in the cells used for transformation does not differ much between this moss and higher plants but the absolute frequency of IR appears to be much lower in *Physcomitrella*. The answer to the question "Why is GT so inefficient in higher plants?" might not then be a question of the inefficiency of HR but more of the important efficiency of IR.

Towards efficient GT in higher plants

'Increasing the frequency of GT involves not only maximising the frequency of HR but also improving selection techniques for the identification of rare recombinants' (Mengiste & Paszkowski, 1999). Indeed, the only report displaying a high frequency of GT (~1%) in higher plants used the hygromycin resistance gene as a disruptive gene (i.e., a positive selection gene) and duplicated highly expressed diphtheria toxin gene-based counter-selectable markers (i.e., a negative selection gene) at both ends of an ends-out configuration transformation vector (Terada et al., 2002) (Figure 3(a)).

Placed in the target construct outside the region of homology, negative selection genes allow the elimination of randomly inserted clones, leading to an increase of the GT frequency (Figure 3(b)). Coding sequence deletion and silencing of the counter-selectable marker, however, limit their potential by generating escapes (e.g., the 99% escapes in the experiment cited

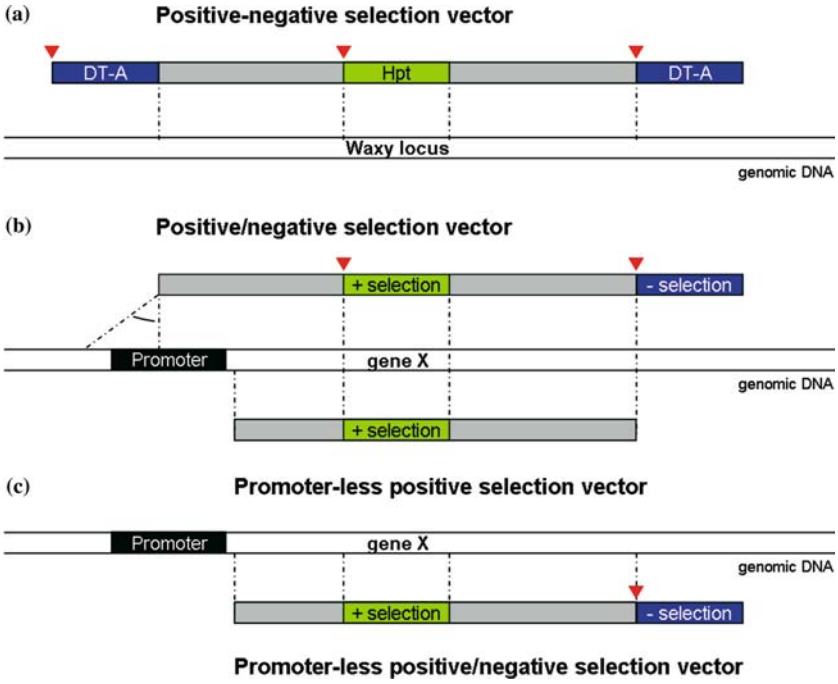

Figure 3. (a) Representation of the targeting vector used by Terada et al. (2002) for the disruption of the *waxy* locus. The hygromycin resistance gene (Hpt), in green, is used as the disruptive gene and two negative selection diphtheria toxin genes (DT-A), in blue, are placed at both ends of the transformation vector. Promoters are represented by red triangles. In gray are the regions of homology shared with the chromosomal target on the transformation vector. (b) Two different strategies for the enrichment of rare knockout events by GT using an ends-out configuration targeting vector (such as a T-DNA). In the positive-negative selection system, both markers are independent expression units while the positive selectable marker in the promoter-less strategy is under control of the native promoter of the targeted gene X after a GT event occurs. Thus, the targeting vector in the promoter-less strategy does not have to share homology with the promoter region of the gene to be targeted while this genomic region may or may not be included in the positive-negative selection system. The disruptive positive selectable marker is represented in green and the negative selection gene is in blue. (c) The promoter-less positive/negative selection vector is designed to further ameliorate these strategies. The vector is based on the promoter-less positive selection system with the addition of a negative selection gene driven by a strong promoter at the vector extremity that is homologous with the 3′ region of the gene to be targeted.

above). Moreover, targeting events accompanied by random insertions of the targeting vectors or transient expression of the negative selection gene are lost. Despite these drawbacks, positive/negative selection systems have proved their usefulness to enrich GT frequencies in rice (Terada et al., 2002, 2004; reviewed by lida & Terada, 2004) as well as in mammalian somatic cells (reviewed by Sedivy & Dutriaux, 1999).

In the specific case of HR-mediated knockouts (as opposed to HR-mediated locus-specific modifications), another enrichment method relies on the use of a promoter-less positive selectable marker whose expression is made conditional on recombination at the homologous target site (Figure 3(b)). In other words, expression of the selectable marker is placed under control of the promoter of the gene to be targeted itself while no expression is supposed to be found during an IR event, leading to the non-selection of these events. In this method, particular care has to be given to the location of the start codon of the target gene: if placed before the positive selection gene in the transformation vector, the two gene sequences have to be aligned in the same reading frame. Sufficient distance also has to be left between the promoter of the target gene and the selectable marker to allow efficient HR. Despite generation of escapes that are recovered when an ectopic promoter allows sufficient expression of the selectable marker during an IR event, it has been demonstrated that promoter-less transformation vectors can achieve enrichments of 20- to 250-fold over a positive/negative selection system in mammalian somatic cells (reviewed by Sedivy & Dutriaux, 1999). This strategy, however, has never been attempted in higher plants.

To further enhance knockout efficiencies, a valuable approach would be to design an ends-out targeting vector based on the promoter-less positive selection system by adding a negative selection gene driven by a strong promoter at the vector extremity which is homologous with the 3′ region of the gene to be targeted (Figure 3(c)). Such a vector, which could be called a promoter-less positive/negative selection vector, would combine the advantages of both previous systems and may allow the observation of higher GT frequencies as both selective agents would select for GT events.

The length of sequence homology between the targeting vector and the target locus has also been suggested to affect GT efficiency. For instance, targeting efficiency was positively correlated in an exponential manner to the extent of homology in mouse embryonic stem cells (Deng & Capecchi, 1992). In yeast, a homology increase of only 200 bp on one end of the disruption vector increased GT efficiency (Gray & Honigberg, 2001). On the contrary, targeting experiments in higher plants using sequence homology ranging from 0.4 to 22 kb did not show any significant difference in GT frequency (Mengiste & Paszkowski, 1999). In rice cv. Niponbarre, the natural frequency found for the disruption of the *waxy* gene with a 13 kb homology ($\sim 10^{-4}$; Terada et al., 2002) is in the same magnitude of a gene reconstruction experiment using the *gfp* marker with only 500 bp homology (Cotsaftis & Guiderdoni, unpublished data). It therefore appears that the extent of homology between the transformation vector and the target locus does not influence GT frequency in higher plants. However, since the absolute number of GT events are dwarfed relative to the absolute number of IR events, this observation could simply be due to biased statistics and only experiments using systems described above to increase the observed frequency of GT would address that issue.

Induced DSBs (Puchta et al., 1993; Chiurazzi et al., 1996; Orel et al., 2003), physiological stresses (such as high salinity concentration (Puchta et al., 1995) or pathogen infection (Kovalchuck et al., 2003)), expression of exogenous recombination proteins (Shalev et al., 1999) or exposure to DNA damaging agents (Kovalchuck et al., 2000; Molinier et al., 2004) have also been reported to enhance intrachromosomal HR level in higher plant somatic cells. These factors may then increase GT frequency as well. However, specifically introducing a DSB at a chosen target locus remains a challenge and the genomic modifications caused by genetic transformation and the effects of the external agents listed above would likely modify other loci. Therefore, the use of these methods for increasing GT frequency may not be practical for the characterisation of knockouts or locus-specific modifications. By contrast, transient expression of a recombination protein gene could be an option if it does not integrate a *posteriori*. For example, transient expression of the yeast rad 52 gene could increase GT efficiency but it may be necessary to reengineer an exogenous version so that it interacts properly with the corresponding plant-specific complexes (as proposed by Ray & Langer, 2002).

Finally, the high efficiency of *Agrobacterium*-mediated transformation in rice (Sallaud et al., 2003; Terada et al., 2004) allowed us to set up a reliable, low-cost and relatively rapid GT test based on the reconstruction of a 3′-deleted *gfp* transgene by HR through two successive rounds of transformation (Cotsaftis & Guiderdoni, unpublished data). In this system, the co-culture of embryogenic calli induced from ~100 seeds with *Agrobacterium* routinely yields between 5000 and 10,000 independently transformed cell lines that can be transformed with a second T-DNA. The entire process may be completed by one person in about 4 months. Screening for GFP activity resulting from HR-mediated reconstruction of an expressed *gfp* gene can be achieved by examining the new cell lines under a UV light in a single day. Basal GT frequency within a wild-type Nipponbare was estimated in the first experiments at 2.23×10^{-4}. This test can be used to observe increased or decreased GT frequency in mutants altered in recombination processes. It may also generate data that will enhance our knowledge of the homologous and illegitimate recombination processes in rice and other higher plant species.

Conclusion

Meeting the challenge of routine GT in higher plants is of crucial interest to researchers and plant breeders who are currently in need of a

powerful tool to specifically modify a given locus in a genome. Higher plants have long been considered to be the last lineage resistant to efficient targeting technology. Although gene disruption with an efficiency of 1% has so far been achieved in a single gene in rice and still has to be tested on other loci and species, this report was a true breakthrough which placed GT efficiency in rice at the same level as in mice. However, increased knowledge of the HR and IR plant protein machinery through mutant analyses and experimentation on different plant tissues, as they may have variable targeting efficiencies (see 'Why is GT so inefficient in higher plants?'), may increase GT frequency further and reduce both the time and labour needed for experiments where thousands of transformation events are initially required. Indeed, even a few percent increase would be valuable.

Acknowledgements

We wish to warmly thank Shigeru lida and Charles I. White for their helpful comments on the manuscript. We also thank Alexander A.T. Johnson and Darren Plett for English language review.

References

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P et al. (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* **301**: 653–657.

An S, Park S, Jeong DH, Lee DY, Kang HG, Yu JH et al. (2003) Generation and analysis of end sequence database for T-DNA tagging lines in rice. *Plant Physiol* **133**: 2040–2047.

Ayora S, Piruat JI, Luna R, Reiss B, Russo VE, Aguilera A et al. (2002) Characterization of two highly similar Rad51 homologs of Physcomitrella patens. *J Mol Biol* **316**: 35–49.

Barnes G and Rio D (1997) DNA double-strand-break sensitivity, DNA replication, and cell cycle arrest phenotypes of Ku-deficient Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* **94**: 867–872.

Bechtold N, Jaudeau B, Jolivet S, Maba B, Vezon D, Voisin R et al. (2000) The maternal chromosome set is the target of the T-DNA in the in planta transformation of Arabidopsis thaliana. *Geneties* **155**: 1875–1887.

Belmaaza A and Chartrand P (1994) One-sided invasion events in homologous recombination at double-strand breaks. *Mutat Res* **314**: 199–208.

Bradley A, Evans M, Kaufman MH and Robertson E (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**: 255–256.

Britt AB (1999) Molecular genetics of DNA repair in higher plants. *Trends Plant Sci* **4**: 20–25.

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

Brunaud V, Balzergue S, Dubreucq B, Aubourg S, Samson F, Chauvin S et al. (2002) T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion site. *EMBO Rep* **3**: 1152–1157.

Byrne ME, Kidner CA and Martienssen RA (2003) Plant stem cells: divergent pathways and common themes in shoots and roots. *Curr Opin Genet Dev* **13**: 551–557.

Capecchi MR (1989) The new mouse genetics: altering the genome by gene targeting. *Trends Genet* **5**: 70–76.

Chilton MD and Que Q (2003) Targeted integration of T-DNA into the tobacco genome at double-stranded breaks: new insights on the mechanism of T-DNA, integration. *Plant Physiol* **133**: 956–965.

Chin HG, Choe MS, Lee SH, Park SH, Koo JC, Kim NY et al. (1999) Molecular analysis of rice plants harboring an Ac/Ds transposable element-mediated gene trapping system. *Plant J* **19**(5): 615–623.

Chiurazzi M, Ray A, Viret JF, Perera R, Wang XH, Lloyd AM et al. (1996) Enhancement of somatic intrachromosomal homologous recombination in Arabidopsis by the HO endonuclease. *Plant Ce1l* **8**: 2057–2066.

Clark AJ, Burl S, Denning C and Dickinson P (2000) Gene targeting in livestock: a preview. *Transgenic Res* **9**: 263–275.

Cyranoski D (2003) Rice genome: a recipe for revolution? *Nature* **422**: 796–798.

Daoudal-Cotterell S, Gallego ME and White CI (2002) The plant Rad50-Mre11 protein complex. *FEBS Lett* **516**: 164–166.

Delseny M (2003) Towards an accurate sequence of the rice genome. *Curr Opin Plant Biol* **6**: 101–105.

Deng C and Capecchi MR (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol Cell Biol* **12**: 3365–3371.

Denning C and Priddle H (2003) New frontiers in gene targeting and cloning: success, application and challenges in domestic animals and human embryonic stem cells. *Reproduction* **126**: 1–11.

Dudas A and Chovanec M (2004) DNA double-strand break repair by homologous recombination. *Mutat Res* **566**: 131–167.

Friesner J and Britt AB (2003) Ku80 – and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integratin. *Plant J* **34**: 427–440.

Gallego ME, Bleuyard JY, Daoudal-Cotterell S, Jallut N and White CI (2003) Ku80 plays a role in non-homologous recombination but is not required for T-DNA integration in Arabidopsis. *Plant J* **35**: 557–565.

Gherbi H, Gallego ME, Jalut N, Lucht JM, Hohn B and White CI (2001) Homologous recombination in planta is stimulated in the absence of Rad50. *EMBO Rep* **2**: 287–291.

Gisler B, Salomon S and Puchta H (2002) The role of double-strand break-induced allelic homologous recombination in somatic plant cells. *Plant J* **32**: 277–284.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. Japonica). *Science* **296**: 92–100.

Gong M and Rong YS (2003) Targeting multi-cellular organisms. *Curr Opin Genet Dev* **13**: 215–220.

Gorbunova VV and Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res* **25**: 4650–4657.

Gorbunova VV and Levy AA (1999) How plants make ends meet: DNA double-strand break repair. *Trends Plant Sci* **4**: 263–269.

Gray M and Honigberg SM (2001) Effect of chromosomal locus, GC content and length of homology on PCR-mediated targeted gene replacement in Saccharomyces. *Nucleic Acids Res* **29**: 5156–5162.

Greco R, Ouwerkerk PB, De Kam RJ, Sallaud C, Favalli C, Colombo L et al. (2003) Transpositional behaviour of an Ac/Ds system for reverse genetics in rice. *Theor Appl Genet* **108**: 10–24.

Greco R, Ouwerkerk PB, Taal AJ, Favalli C, Beguiristain T, Puigdomenech P et al. (2001) Early and multiple Ac transpositions in rice suitable for efficient insertional mutagenesis. *Plant Mol Biol* **46**: 215–227.

Grelon M, Vezon D, Gendrot G and Pelletier G (2001) AtSPO11-1 is necessary for efficient meiotic recombination in plants. *EMBO J* **20**: 589–600.

Haber JE (1999) DNA repair. Gatekeepers of recombination. *Nature* **398**: 665–667.

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.

Hanin M and Paszkowski J (2003) Plant genome modification by homologous recombination. *Curr Opin Plant Biol* **6**: 157–162.

Hanin M, Volrath S, Bogucki A, Briker M, Ward E and Paszkowski J (2001) Gene targeting in Arabidopsis. *Plant J* **28**: 671–677.

Hirochika H (2001) Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr Opin Plant Biol* **4**: 118–122.

Hohe A, Egener T, Lucht JM, Holtorf H, Reinhard C, Schween G et al. (2004) An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, Physcomitrella patens. *Curr Gene* **44**: 339–347.

Hohn B and Puchta H (2003) Some like it sticky: targeting of the rice gene Waxy. *Trends Plant Sci* **8**: 51–53.

Ida S and Terada R (2004) A tale of two integrations, transgene and T-DNA: gene targeting by homologous recombination in rice. *Curr Opin Biotechnol* **15**: 132–138.

Inoue K, Ogura A and Hayashi J (2002) Production of mitochondrial DNA transgenic mice using zygotes. *Methods* **26**: 358–363.

Izawa T, Ohnishi T, Nakano T, Ishida N, Enoki H, Hashimoto H et al. (1997) Transposon tagging in rice. *Plant Mol Biol* **35**: 219–229.

Jeon JS, Lee S, Jung KH, Jun SH, Jeong DH, Lee J et al. (2000) T-DNA insertional mutagenesis for functional genomics in rive. *Plant J* **22**: 561–570.

Jeong DH, An S, Kang HG, Moon S, Han JJ, Park S et al. (2002) T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol* **130**: 1636–1644.

Karathanasis E and Wilson TE (2002) Enhancement of Saccharomyces cerevisiae end-joining efficiency by cell growth stage but not by impairment of recombination. *Genetics* **161**: 1015–1027.

Kempin SA, Liljegren SJ, Block LM, Rounsley SD, Yanofsky MF and Lam E (1997) Targeted disruption in Arabidopsis. *Nature* **389**: 802–803.

Kohli A, Twyman RM, Abranches R, Wegel E, Stoger E and Christou P (2003) Transgene integration, organization and interaction in plants. *Plant Mol Biol* **52**: 247–258.

Kohli A, Xiong J, Greco R, Christou P and Pereira A (2001) Tagged Transcriptome Display (TTD) in indica rice using Ac transposition. *Mol Genet Genomics* **266**: 1–11.

Kirik A, Salomon S and Puchta H (2000) Species-specific double-strand break repair and genome evolution in plants. *EMBO J* **19**: 5562–5566.

Kolesnik T, Szeverenyi I, Bachmann D, Kumar CS, Jiang S, Ramamoorthy R et al. (2004) Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. *Plant J* **37**: 301–314.

Kovalchuk I, Kovalchuk O, Kalck V, Boyko V, Filkowski J, Heinlein M et al. (2003) Pathogen-induced systemic plant signal triggers DNA rearrangements. *Nature* **423**: 760–762.

Kovalchuk O, Arkhipov A, Barylyak I, Karachov I, Titov V, Hohn B et al. (2000) Plants experiencing chronic internal exposure to ionizing radiation exhibit higher frequency of homologous recombination than acutely irradiated plants. *Mutat Res* **449**: 47–56.

Krysan PJ, Young JC, Jester PJ, Monson S, Copenhaver G, Preuss D et al. (2002) Characterization of T-DNA insertion sites in Arabidopsis thaliana and the implications for saturation mutagenesis. *OMICS* **6**: 163–174.

Lee KY, Huang H, Ju B, Yang Z and Lin S (2002) Cloned zebrafish by nuclear transfer from long-term-cultured cells. *Nat Biotechnol* **20**: 795–799.

Lee KY, Lund P, Lowe K and Dunsmuir P (1990) Homologous recombination in plant cells after Agrobacterium-mediated transformation. *Plant Cell* **2**: 415–425.

Leung H, Wu C, Baraoidan M, Bordeos A, Ramos M, Madamba S et al. (2001) Deletion mutants for functional genomics: progress in phenotyping, sequence assignment, and database development. In: Khush GS, Brar DS and Hardy B (eds), Rice Genetics IV (pp. 239–51) Science Publishers, Inc. & International Rice Research Institute, New Delhi.

Li X, Song Y, Century K, Straight S, Ronald P, Dong X et al. (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant J* **27**: 235–242.

Lin FL, Sperle K and Sternberg N (1984) Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. *Mol Cell Biol* **4**: 1020–1034.

Lin H (1997) The tao of stem cells in the germline. *Annu Rev Genet* **31**: 455–491.

Lisby M, Antunez de Mayolo A, Mortensen UH and Rothstein R (2003) Cell cycle-regulated centers of DNA double-strand break repair. *Cell Cycle* **2**: 479–483.

Lowe B, Mathern J and Hake S (1992) Active Mutator elements suppress the knotted phenotype and increase recombination at the Kn1-O tandem duplication. *Genetics* **132**: 813–822.

Ma JL, Kim EM, Haber JE and Lee SE (2003) Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol* **23**: 8820–8828.

Makarevitch I, Svitashev SK and Somers DA (2003) Complete sequence analysis of transgene loci from plants transformed via microprojectile bombardment. *Plant Mol Biol* **52**: 421–432.

Markmann-Mulisch U, Hadi MZ, Koepchen K, Alonso JC, Russo VE, Schell J et al. (2002) The organization of Physcomitrella patens Rad51 genes is unique among eukaryotic organisms. *Proc Natl Acad Sci USA* **99**: 2959–2964.

Marsch-Martinez N, Greco R, Van Arkel G, Herrera-Estrella L and Pereira A (2002) Activation tagging using the En-I maize transposon system in Arabidopsis. *Plant Physiol* **129**: 1544–1556.

Mengiste T and Paszkowski J (1999) Prospects for the precise engineering of plant genomes by homologous recombination. *Biol Chem* **380**: 749–758.

Mengiste T, Revenkova E, Bechtold N and Paszkowski J (1999) An SMC-like protein is required for efficient homologous recombination in Arabidopsi. EMBO *J* **18**: 4505–4512.

Miao ZH and Lam E (1995) Targeted disruption of the TGA3 locus in Arabidopsis thaliana. *Plant J* **7**: 359–365.

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K et al. (2003) Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* **15**: 1771–1780.

Molinier J, Ries G, Bonhoeffer S and Hohn B (2004) Interchromatid and Interhomolog Recombination in Arabidopsis thaliana. *Plant Cell* **16**: 342–352.

Offringa R, de Groot MJ, Haagsman HJ, Does MP, van den Elzen PJ and Hooykaas PJ (1990) Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation. *EMBO J* **9**: 3077–3084.

Orel N, Kyryk A and Puchta H (2003) Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. Plant *J* **35**: 604–612.

Pâiques F and Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisia. *Microbiol Mol Biol Rev* **63**: 349–404.

Pastwa E and Blasiak J (2003) Non-homologous DNA end joining. *Acta Biochim Pol* **50**: 891–908.

Paszkowski J, Baur M, Bogucki A and Potrykus I (1988) Gene targeting in plants. *EMBO J* **7**: 4021–4026.

Prado F and Aguilera A (2003) Control of cross-over by single-strand DNA resection. *Trends Genet* **19**: 428–431.

Puchta H (1998) Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequence. *Plant J* **13**: 331–339.

Puchta H (1999) Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cell. *Genetics* **152**: 1173–1181.

Puchta H, Dujon B and Hohn B (1993) Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res* **21**: 5034–5040.

Puchta H, Dujon Band Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc Natl Acad Sci USA* **93**: 5055–5060.

Puchta H and Hohn B (1991) The mechanism of extrachromosomal homologous DNA recombination in plant cells. *Mol Gen Genet* **230**: 1–7.

Puchta H and Hohn B (1996) From centiMorgans to base pairs: homologous recombination in plants. *Trends Plant Sci* **1**: 340–348.

Puchta H, Swoboba P and Hohn B (1995) Induction of intrachromosomal homologous recombination in whole plants. *Plant J* **7**: 203–210.

Raina S, Mahalingam R, Chen F and Fedoroff N (2002) A collection of sequenced and mapped Ds transposon insertion sites in Arabidopsis thaliana. *Plant Mol Biol* **50**: 93–110.

Ray A and Langer M (2002) Homologous recombination: ends as the means. *Trends Plant Sci* **7**: 435–440.

Reiss B, Schubert I, Kopchen K, Wendeler E, Schell J and Puchta H (2000) RecA stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not gene targeting, in plants transformed by Agrobacterium. *Proc Natl Acad Sci USA* **97**: 3358–3363.

Reski R (1998) Development, genetics and molecular biology of mosses. *Bot Acta* **111**: 1–15.

Risseeuw E, Franke-van Dijk ME and Hooykaas PJ (1997) Gene targeting and instability of Agrobacterium T-DNA loci in the plant genome. *Plant J* **11**: 717–728.

Risseeuw E, Offringa R, Franke-van Dijk ME and Hooykaas PJ (1995) Targeted recombination in plants using Agrobacterium coincides with additional rearrangements at the target locus. *Plant J* **7**: 109–119.

Ristic D, Modesti M, Kanaar R and Wyman C (2003) Rad52 and Ku bind to different DNA structures produced early in double-strand break repair. *Nucleic Acids Res* **31**: 5229–5237.

Rong YS and Golic KG (2000) Gene targeting by homologous recombination in Drosophila. *Science* **288**: 2013–2018.

Roth DB (2003) Restraining the V(D)J recombines. *Nat Rev Immunol* **3**: 656–666.

Ruf S, Hermann M, Berger IJ, Carrer H and Bock R (2001) Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat Biotechnol* **19**: 870–875.

Sallaud C, Gay C, Larmande P, Bès M, Piffanelli P, Piégu B et al. (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards *in silic* reverse genetics. *Plant J* **39**: 450–464.

Sallaud C, Meynard D, Van Boxtel J, Gay C, Bes M, Brizard JP et al. (2003) Highly efficient production and characterization of T-DNA plants for rice (Oryza sativa L) functional genomics. *Theor Appl Genet* **106**: 1396–1408.

Salomon S and Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J* **17**: 6086–6095.

Schaefer DG (2001) Gene targeting in Physcomitrella patens. *Curr Opin Plant Biol* **4:** 143–150.

Schaefer DG (2002) A new moss genetics: targeted mutagenesis in Physcomitrella patens. *Annu Rev Plant Biol* **53**: 477–501.

Schaefer OG and Zryd JP (1997) Efficient gene targeting in the moss Physcomitrella patens. *Plant J* **11**: 1195–1206.

Schoof H and Karlowski WM (2003) Comparison of rice and Arabidopsis annotation. *Curr Opin Plant Biol* **6**: 106–112.

Schween G, Gorr G, Hohe A and Reski R (2003) Unique tissue-specific cell cycle in Physcomitrella. *Plant Biol* **5**: 50–58.

Sedivy JM and Dutriaux A (1999) Gene targeting and somatic cell genetics – a rebirth or a coming of age? *Trends Genet* **15**: 88–90.

Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D et al. (2002) A high-throughput Arabidopsis reverse genetics system. *Plant Cell* **14**: 2985–2994.

Shalev G, Sitrit V, Avivi-Ragolski N, Lichtenstein C and Levy AA (1999) Stimulation of homologous recombination in plants by expression of the bacterial resolvase ruvC. *Proc Natl Acad Sci USA* **96**: 7398–7402.

Siebert R and Puchta H (2002) Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. *Plant Cell* **14**: 1121–1131.

Smith KN and Nicolas A (1998) Recombination at work for meiosis. *Curr Opin Genet Dev* **8**: 200–211.

Somers DA and Makarevitch I (2004) Transgene integration in plants: poking or patching holes in promiscuous genomes? *Curr Opin Biotechnol* **15**: 126–131.

Sugawara N, Wang X and Haber JE (2003) *In vivo* roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. *Mol Ce1l* **12**: 209–219.

Symington LS (2002) Role of Rad52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiol Mol Biol Rev* **66**: 630–670.

Szostak JW, Orr-Weaver TL, Rothstein RJ and Stahl FW (1983) The double-strand-break repair model for recombination. *Cell* **33**: 25–35.

Takano M, Egawa H, Ikeda JE and Wakasa K (1997) The structures of integration sites in transgenic rice. *Plant J* **11**: 353–361.

Terada R, Assao H and Iida S (2004) A large *Agrobacterium-mediated* transformation procedure with a strong positive-negative selection for gene targeting in rice (*Oryza sativa* L.) *Plant Cell Rep* **22**: 653–659.

Terada R, Urawa H, Inagaki V, Tsugane K and lida S (2002) Efficient gene targeting by homologous recombination in rice. *Nat Biotechnol* **20**: 1030–1034.

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE et al. (2003) Large-: scale discovery of induced point mutations with high-throughput TILLING. *Genome Res* **13**: 524–530.

Trujillo KM, Roh DH, Chen L, Van Komen S, Tomkinson A and Sung P (2003) Yeast *Xrs2* binds DNA and helps target. Rad50 and Mre11 to DNA ends. J *Biol Chem* **278**: 48957–48964.

Tuteja N, Singh MB, Misra MK, Bhalla PL and Tuteja R (2001) Molecular mechanisms of DNA damage and repair: progress in plants. *Crit Rev Biochem Mol Biol* **36**: 337–397.

Tzfira T and Citovsky V (2003) The Agrobacterium-plant cell interaction. Taking biology lessons from a bug. *Plant Physiol* **133**: 943–947.

Tzfira T, Frankman LR, Vaidya M and Citovsky V (2003) Site-specific integration of Agrobacterium tumefaciens T-DNA via double-stranded intermediates. *Plant Physiol* **133**: 1011–1023.

Upadhyaya NM, Zhou XR, Zhu QH, Ramm K, Wu L, Eamens A et al. (2002) An iAc/Ds gene and enhancer trapping system for insertional mutagenesis in rice. *Funcl Plant Biol* **29**: 547–559.

van Attikum H, Bundock P and Hooykaas PJ (2001) Non-homologous end-joining proteins are required for Agrobacterium T-DNA integration. *EMBO J* **20**: 6550–6558.

van Attikum H, Bundock P, Overmeer RM, Lee LV, Gelvin SB and Hooykaas PJ (2003) The Arabidopsis AtLlG4 gene is required for the repair of DNA damage, but not for the integration of Agrobacterium T-DNA *Nucleic Acids Res* **31**: 4247–4255.

van Attikum Hand Hooykaas PJ (2003) Genetic requirements for the targeted integration of Agrobacterium T-DNA in Saccharomyces cerevisiae. *Nucleic Acids Res* **31**: 826–832.

van den Bosch M, Lohman PH and Pastink A (2002) DNA double-strand break repair by homologous recombination. *Biol Chem* **383**: 873–892.

Van Dyck E, Stasiak AZ, Stasiak A and West SC (1999) Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* **398**: 728–731.

Vergunst AC and Hooykaas PJ (1999) Recombination in the plant genome and its application in biotechnology. *Cri Rev Plant Sci* **18**: 1–31.

Ware D and Stein L (2003) Comparison of genes among cereals. *Curr Opin Plant Biol* **6**: 121–127.

Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, Della M et al. (2002) Identification of a DNA non homologous end-joining complex in bacteria. *Science* **297**: 1686–1689.

Windels P, De Buck S, Van Bockstaele E, De Loose M and Depicker A (2003) T-DNA integration in Arabidopsis chromosomes. Presence and origin of filler DNA sequences. *Plant Physiol* **133**: 2061–2068.

Winding P and Berchtold MW (2001) The chicken B cell line DT40: a novel tool for gene disruption experiments. *J Immunol Methods* **249**: 1–16.

Wolner B, van Komen S, Sung P and Peterson CL (2003) Recruitment of the recombinational repair machinery to a DNA double-strand break in yeast. *Mol Cell* **12**: 221–232.

Wu C, Li X, Yuan W, Chen G, Kilian A, Li J et al. (2003) Development of enhancer trap lines for functional analysis of the rice genome. *Plant J* **35**: 418–427.

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science* **296**: 79–92.

Zickler D and Kleckner N (1999) Meiotic chromosomes: integrating structure and function. *Annu Rev Genet* **33**: 603–754.