Gene targeting in plants: 25 years later

HOLGER PUCHTA* and FRIEDRICH FAUSER

Botanical Institute II, Karlsruhe Institute of Technology, Karlsruhe, Germany

ABSTRACT Only five years after the initiation of transgenic research in plants, gene targeting (GT) was achieved for the first time in tobacco. Unfortunately, the frequency of targeted integration via homologous recombination (HR) was so low in comparison to random integration that GT could not be established as a feasible technique in higher plants. It took another 25 years and great effort to develop the knowledge and tools necessary to overcome this challenge, at least for some plant species. In some cases, the overexpression of proteins involved in HR or the use of negative selectable markers improved GT to a certain extent. An effective solution to this problem was developed in 1996, when a sequence-specific endonuclease was used to induce a double-strand break (DSB) at the target locus. Thus, GT frequencies were enhanced dramatically. Thereafter, the main limitation was the absence of tools needed to induce DSBs at specific sites in the genome. Such tools became available with the development of zinc finger nucleases (ZFNs), and a breakthrough was achieved in 2005 when ZFNs were used to target a marker gene in tobacco. Subsequently, endogenous loci were targeted in maize, tobacco and Arabidopsis. Recently, our toolbox for genetic engineering has expanded with the addition of more types of site-specific endonucleases, meganucleases, transcription activator-like effector nucleases (TALENs) and the CRISPR/Cas system. We assume that targeted genome modifications will become routine in the near future in crop plants using these nucleases along with the newly developed in planta GT technique.

KEY WORDS: plant biotechnology, gene technology, synthetic nucleases, transformation, double-strand break repair

Introduction

In this review, we intend to give a concise overview of the history and current developments of GT in plants. Due to space limitations, we will not be able to discuss all of the approaches that have been attempted, especially ones that did not help to improve GT frequencies in the long term. Moreover, we will only discuss the molecular mechanisms of DNA recombination insofar as they are important for the understanding of the GT reaction itself and will primarily refer to recent reviews for further details (Lieberman-Lazarovich and Levy, 2011; Waterworth et al., 2011). Genome editing in plants was discussed recently in an excellent review (Voytas, 2013), and two chapters on specific aspects of genome editing in plants are also included in this issue (see Marton et al., 2013; D’Halluin and Reuter, 2013). Therefore, we will not go into too much detail on the architecture and set-up of the different types of custom, artificial nucleases that are used for GT as well as their application for targeted mutagenesis.

The production of transgenic plants relies heavily on DNA recombination, which is required in somatic cells to repair DSBs within nuclear DNA. The prevailing mechanism of DSB repair in higher plants is nonhomologous end joining (NHEJ), which is also required for the random integration of foreign DNA into the plant genome. On the other hand, DSB repair can occur by the use of homologous sequences, which depending on the circumstances, occurs with a much lower efficiency than NHEJ (Puchta, 2005).

Mechanisms of homologous double-strand break repair in somatic plant cells

In principle, it is possible to discriminate between the two different mechanisms of homologous DSB repair, single-strand annealing (SSA) and synthesis-dependent strand annealing (SDSA), in somatic plant cells (Fig. 1). For SSA, a break between two direct repeats is repaired by removing the internal sequences and is thus a non-conservative pathway. SSA appears to be quite

Abbreviations used in this paper: DSB, double-strand break; Cas, CRISPR-associated; CRISPR, clustered regularly interspaced short palindromic repeats; GT, gene targeting; HR, homologous recombination; NHEJ, non-homologous end joining; SDSA, synthesis-dependent strand annealing; SSA, single-strand annealing; TALEN, transcription-activator like effector nuclease; ZFN, zinc finger nuclease.

*Address correspondence to: Holger Puchta, Botanical Institute II, Karlsruhe Institute of Technology, Kaiserstr. 12, 76131 Karlsruhe, Germany. Tel.: +49 721 608-43833; Fax.: +49 721 608-44874; e-mail: holger.puchta@kit.edu

Final, author-corrected PDF published online: 16 October 2013.
efficient in genomic regions with tandemly arranged duplication; in these regions, up to one out of every three DSBs is repaired by this mechanism (Siebert and Puchta, 2002). In contrast, SDSA is conservative. A homologous sequence is copied into the break site without altering the donor. SDSA appears to be approximately five to ten times less efficient than SSA under comparable conditions (Orel et al., 2003). In both pathways, single-stranded overhangs are produced via exonuclease-catalyzed resection after DSB induction. In the case of SSA, overhangs at both ends of the break carry complementary sequences, and the two single-strands can directly anneal with one another to form a chimeric DNA molecule. If the molecule contains 3' overhangs, these will be trimmed; otherwise, single-stranded regions would be filled in by repair synthesis. In the case of SDSA, a 3' end invades a homologous double-strand forming a D-loop. Repair synthesis begins using the newly paired strand as a template. After elongation, the strand is displaced from the D-loop structure and anneals with the 3' homologous strand that becomes available due to resection of the second end of the DSB. Thus, gene conversion without a loss of sequence information is the final result of the reaction (Fig. 1). Not surprisingly, the involvement of DNA repair proteins differs considerably between these two pathways. Proteins involved in the strand exchange reaction, such as AtRAD51 or AtRAD54, are required for SDSA but not for SSA (Roth et al., 2012). In contrast, certain helicases such as AtRECO4A and nucleases such as AtMUS81 that play a role in SDSA might also play a minor role in SSA (Mannuss et al., 2010), although no factor that is specific for SSA has been characterized yet.

**Gene targeting in plants - the first experiments**

Whereas DNA integration in bacteria and yeast occurs primarily via HR and genes can be readily targeted with homologous sequences, the situation in most multicellular organisms is quite different. GT in higher plants and animals was difficult to accomplish for a long time. Due to the pioneering and Noble Prize-winning work of Smithies (Doetschman et al., 1987) and Capecchi (Thomas and Capecchi, 1987), GT in mouse embryonic stem cells was accomplished successfully more than two decades ago. Nevertheless, similar success was not achieved in other higher eukaryotes for many years.

At the Friedrich Miescher Institute in Basel, Switzerland, Jurek Paszkowski was the first to demonstrate in his pioneering study that genes can indeed be targeted in plant cells (Paszkowski et al., 1988). By direct gene transfer to tobacco protoplasts, he demonstrated that plasmid DNA containing part of the kanamycin gene could be integrated into the genome at a low frequency via HR such that a complementary gene fragment was restored. In addition to direct gene transfer, Paul Hooykaas and his group were able to demonstrate that GT in plant cells by Agrobacterium-mediated T-DNA transformation worked as well (Offringa et al., 1990). However, in these and a number of subsequent studies, the observed GT frequencies remained disappointingly low, at approximately one targeting event per $10^4$ to $10^5$ transformation events using different transformation methods and plant species (for a review of the first 15 years of GT experiments, see (Puchta, 2002)).

Interestingly, different groups analyzed the few targeting events that were isolated with selectable markers as model substrates and revealed that HR did not occur in all cases at both end of the targeting vector during the recombinational reaction [e.g. (Halin et al., 2001)]. In principle, two different classes of events could be identified, in which one end of the vector recombined with the target via HR and the other was repaired via NHEJ (Fig. 2). In one class, the targeting vector was inserted at a site of homology (“one-sided events”). Thus, the vector was inserted at the target locus with one junction formed via HR and the other via NHEJ. However, there was a second class of events in which the selectable marker gene was restored by copying parts of the genomic sequence onto the targeting vector, which was then integrated elsewhere in the genome via NHEJ (“ectopic GT events”). Thus, although the marker gene was restored via HR using the target locus as template, the original target was not changed and was still present in the genome. Although the outcome of these reactions improved our understanding of the mechanisms of recombination, they fell short of the biotechnological expectation of knocking out a gene in a controlled manner.

**Manipulation of the enzyme machinery can help (a bit)**

If GT is not efficient in plants, why not transfer into plants com-
ponents of the HR enzyme machinery from an organism in which GT works well? Several groups have addressed this possibility, and we would like to note the two most prominent examples, the RecA gene of *Escherichia coli* and the RAD54 gene of *Saccharomyces cerevisiae*. The group of Bernd Reiss produced transgenic tobacco plants expressing RecA, the strand exchange protein and key HR factor from bacteria. As expected, the rate of intrachromosomal HR was enhanced by one order of magnitude in plants expressing the *E. coli* enzyme. Unfortunately, no significant enhancement of the GT frequency could be achieved with Agrobacterium-mediated transformation (Reiss *et al.*, 2000). In a different approach, the group of Avi Levy overexpressed the RAD54 gene of baker's yeast in Arabidopsis. ScRAD54 is required for efficient strand invasion of HR intermediates. Indeed, the ScRAD54 protein substantially enhanced GT frequency. The authors used a promoterless GFP ORF inserted in an Arabidopsis cruciferin gene as assay system. As Cruciferin is a seed storage protein, integration of the GFP-containing gene into the genomic target site produced fluorescent seeds, making it a convenient and non-destructive marker for GT (Shaked *et al.*, 2005). The same authors expanded on their work in a later study with egg-cell specific expression of RAD54 to enhance GT using the Agrobacterium infiltration technique for transformation (Even-Faitelson *et al.*, 2011).

In addition to using enzymes from other organisms, the manipulation of the intrinsic recombination machinery of plants could also enhance GT efficiency. It is possible that NHEJ and HR function in equilibrium and that GT could be enhanced drastically by blocking NHEJ. However, no convincing study applying this approach to plants has been published thus far. It has been shown recently that these difficulties may be caused by the existence of independent end-joining pathways in plant cells (Charbonnel *et al.*, 2011). A mild reduction in T-DNA transformation efficiencies was reported when single NHEJ factors were knocked out, but no GT experiments have yet been published concerning this issue (Friesner and Britt, 2003; Nishizawa-Yokoi *et al.*, 2012).

Indeed, mutation of certain host factors involved in DNA repair can also lead to a hyper-recombination phenotype, as measured by increased intra- or interchromosomal HR frequencies. This process has been confirmed for certain DNA helicases such as AtREC4QA or AtFANCM that are involved in the control of HR (Hartung *et al.*, 2007; Knoll *et al.*, 2012) and for chromatin assembly factors such as AtCAF1 (Endo *et al.*, 2006; Kirik *et al.*, 2006). Again, no GT experiments have been published using these mutant backgrounds.

The major problem with approaches relying on the manipulation of the recombinatorial enzyme machinery is that these strategies may destabilize the genome of an organism in a general way. One must bear in mind that if HR is more efficient, than enhanced recombination might occur not only between the targeting vector and the target locus but also between repetitive sequences all over the genome, leading to undesired and uncontrolled off-site effects. Attempts have been made to minimize the time window for GT using inducible or organ-specific promoters (Even-Faitelson *et al.*, 2011), but even with these approaches, exclusive activation of the site of interest has not been possible.

Gene targeting in *Physcomitrella patens* remains a mystery

In contrast to flowering plants, the moss *Physcomitrella patens* is able to integrate DNA efficiently via HR (see Strotbek *et al.*, 2013 in the current issue). This finding was first reported by Didier Schaefer and Jean-Pierre Zryd (Schaefer and Zryd, 1997). Soon thereafter, Ralf Reski’s group demonstrated that this technique could efficiently knock-out gene functions (Strepp *et al.*, 1998). This moss has become a valuable system for the investigation of the basic processes of plant biology. Nevertheless, no obvious peculiarities in the DNA repair and recombination machinery were found, despite the fact that the genome sequence of this moss was elucidated some years ago (Rensing *et al.*, 2008). Thus, transferring the efficient GT machinery of *Physcomitrella* to higher plants has not yet been accomplished because we do not currently understand why GT is so efficient in moss.

Negative selection can help (a bit)

Another strategy to eliminate surplus random integration events is to not block the NHEJ pathways or tune up the HR machinery but to simply select against random NHEJ events, so that these are eliminated from the transgenic pool. Negative selectable marker genes can be used for this approach. These genes have to be included in the targeting construct flanking the homologous regions so that they will not be integrated into the genome if HR occurs correctly (Fig. 3). By targeting the waxy gene of rice, Shigeru Iida’s group was able to demonstrate in a pioneering study from a decade ago that GT experiments can be performed in plants using a negative selectable marker (Terada *et al.*, 2002). They performed...
Agrobacterium-mediated transformation of highly proliferative callus material derived from the seeds of *Oryza sativa* with a T-DNA carrying genomic homologies of several kbs. The homologies were flanked by two genes for diphtheria toxin, which are lethal to plants when randomly integrated into the genome. Thus, recombinant calli could be produced and regenerated to fertile plants as well. The estimated ratio of targeted to random integration events was $6.5 \times 10^{-4}$, which makes this technique feasible but laborious. Since then, the group succeeded in targeting a few other natural genes in rice (Ono et al., 2012), indicating that the technique is a useful way to achieve GT in plants. Nevertheless, in the long term, the negative selection strategies might be outcompeted by less time-consuming approaches linked with DSB-induced GT in plants (see below).

**Double-strand break induction: a major breakthrough for site-specific integration**

In the end, what helped to solve the GT challenge in plants was to simply copy nature. An efficient way to induce recombination at specific genomic loci is to break both strands of the DNA at the site of interest in a genetically programmed way, as occurs during meiosis or during mating type switching. Under such circumstances, DSBs are often induced by specific cellular enzymes. Indeed, it was demonstrated some time ago in yeast that artificial sites can also be activated for HR by the induction of DSBs [for a review, see (Paques and Haber, 1999)].

Therefore, site-specific nucleases became an obvious possible strategy to increase the HR frequency in plants. Obviously, for this purpose, endonucleases are required with restriction sites that are complex enough that, statistically, no natural site should be present in the respective host genome to avoid unwanted off-target effects. A promising candidate is the homing endonuclease (“meganuclease”) I-SceI, which was originally isolated from yeast mitochondria. It has a 18-mer recognition site (Fig. 4A). The I-SceI ORF integrates itself by inducing a DSB into the mitochondrial 21S rRNA gene. The DSB is repaired with the aid of a copy of the 21S rRNA gene that already includes the I-SceI ORF as an intron. Thus I-SceI can be spread in the mitochondrial DNA pool (Jacquier and Dujon, 1985). Initially, the applicability of DSB-induced HR to plants was confirmed using I-SceI and transiently transformed plasmid molecules in Nicotiana protoplasts (Puchta et al., 1993). Most importantly, a DSB could also be introduced at a transgenic locus within the tobacco genome *in vivo* via transient expression of I-SceI, resulting in a homologous integration frequency of up to $10^{-2}$ at the transgenic locus using a T-DNA carrying homologies to this locus (Puchta et al., 1996). In most of the events, both ends of the target vector were integrated via HR, but combinations of HR and NHEJ were also found. Using the same system, it was also demonstrated that GT frequencies were approximately half as high if a T-DNA only carried homology to one end of the break (Puchta, 1998). This is a strong indication that HR is initiated at both ends independently, which is in accordance with the SDSA model. However, even without homologies to the genomic target site, it was found that T-DNAs could insert into meganuclease-induced DSBs via NHEJ (Chilton and Que, 2003; Salomon and Puchta, 1998; Tzfira et al., 2003). Moreover, meganuclease-mediated DSB induction could also be used for the excision of sequences that were flanked by the respective meganuclease recognition sites from the genome either via NHEJ or, in case of available homologies, by SSA (Siebert and Puchta, 2002). Additionally, the loss of gene function via NHEJ after DSB induction was also achieved (Kirik et al., 2000).

Although early studies with homing endonucleases demonstrated how efficiently sequence-specific meganucleases could be used for plant genome editing, the limitation of having only one particular sequence as a recognition site precluded any practical application for the knock-out of natural genes. It was not possible to target DSBs to sites of interest in the genome. For this purpose, nucleases that were prone to artificial manipulation of their binding sites were required. Therefore, more than a decade ago, attempts to develop meganucleases with artificially modulated binding sites were initiated (Fig. 4B). Although such an enzyme was recently used for the NHEJ-mediated targeted mutagenesis of maize (Gao et al., 2010), no successful GT by a modified meganuclease has been published for plants. In general, the problem with manipulating meganuclease binding sites is that the recognition specificity has to be changed without harming the endonuclease activity, although both reside in the same domain. Therefore, alternative approaches were taken relatively early to combine an endonuclease domain with a separate DNA binding domain of different biological origin that could be manipulated independently. An endonuclease domain from a partial ORF of the classic restriction enzyme FokI, which acts as dimer, has been used. With this system, the dimerization of two FokI domains through the action of two independent DNA binding domains is necessary to induce a DSB. As both DNA bindings domains can be manipulated independently to recognize
different sequence tracks in close proximity on opposite strands, the specificity of the artificial enzyme can be dramatically increased. Two types of DNA binding domains have been used successfully (Fig. 4 C,D). More than a decade ago, work on zinc finger nucleases (ZFNs) was initiated, and in the last few years, Transcription Activator-like Effector Nucleases (TALENs) have garnered interest. In the following sections, the progress with these systems and their application to GT in plants will be discussed in detail.

Zinc finger nucleases: fulfillment of the promise

The set-up of ZFNs was pioneered by Dana Carroll. ZFNs consist of the endonuclease domain of the restriction enzyme FokI fused to the zinc finger binding arrays found in transcription factors (Smith et al., 2000). A zinc finger binding array typically recognizes three consecutive nucleotides. Three to four arrays are combined as DNA binding sites so that, in total, 18 to 24 bases of genomic sequence can be recognized per dimer (Fig. 4C). Dan Voytas recognized early on the potential of these enzymes for genome editing in plants, and his group was the first to demonstrate that ZFNs could be used for GT in plants. Using the restoration of a defective marker gene in tobacco protoplasts for selection, they demonstrated that GT frequencies could be enhanced by up to $10^3$ in comparison with random integration (Wright et al., 2005). Indeed, this was the decisive breakthrough that demonstrated beyond a doubt that custom-made nucleases were the ultimate solution to problems troubling GT in plants.

Nevertheless, another four years passed before the group of Dan Voytas and scientists from Dow Chemicals were able to demonstrate in two independent studies that it was possible to target endogenous genes in tobacco and maize using ZFNs. In tobacco, the SuRA and SuRB loci were modified at the single nucleotide level by ZFN-mediated GT. Thus, resistances to different herbicides were established, and GT frequencies of up to several per cent were reported (Townsend et al., 2009). In maize, ZFN-mediated GT of the maize IPK1 gene was achieved. Again, the GT construct was made in such a way that expression of the recombinant IPK1 resulted in an herbicide-resistant phenotype. The authors reported GT frequencies of more than 10% in most experiments (Shukla et al., 2009). Very recently, ZFN-mediated GT was documented in Arabidopsis (de Pater et al., 2012; Qi et al., 2013).

ZFNs can also be used for site-directed mutagenesis via NHEJ in plants (Lloyd et al., 2005). The groups of Dan Voytas and Seichi Toki have demonstrated that ZFNs are efficient tools for the knock-out of natural genes in Arabidopsis (Osakabe et al., 2010; Zhang et al., 2010).

The work with ZFNs has demonstrated conclusively that different types of genome editing can be achieved with this class of enzyme. Still, some concerns remain...
Off-target effects are a source of some concern. ZFNs might also cut other sites in the genome that are similar to the target site, inducing unpredicted mutations that might result in uncontrolled secondary effects. Indeed, a collection of ZFNs had negative effects on cell proliferation, which may indicate that these nucleases could harm cells by creating unwanted DSBs at secondary sites. Initially, the construction of ZFNs was quite time consuming. Although kits for the construction of ZFNs are now available, construction is still quite laborious. A further uncertainty stems from the fact that the different binding modules in the zinc finger binding arrays influence each other. Therefore, the construction of domains for new genomic sites has not always been as predictable and efficient as expected.

Transcription activator-like effector nucleases make things easier

In a groundbreaking analysis, the plant pathologists Ulla Bonas and Jens Boch discovered that a protein delivered into its host by the bacterial pathogen Xanthomonas carries a DNA binding domain that binds to different plant promoters (Boch et al., 2009). The domain consists of 13 to 28 copies of highly conserved repeats spanning 34 amino acids each. Interestingly, each repeat is able to recognize a single base. Thus, the scientific community acquired an alternative and easily programmable DNA binding domain. Very quickly, several proofs of concept were reported by various groups, who documented that a powerful new class of nucleases was feasible with a set-up similar to ZFNs (Voytas, 2013) (Fig. 4D).

Due to the repetitive nature of the DNA binding motifs, cloning the domain to E. coli could potentially cause problems. However, the development of ligation-based methods made it possible to handle the assembly more easily. TALENs can now be assembled within a week in the lab using publically available kits (Cermak et al., 2011). Moreover, as every single base is addressed by a single repeat, the management of binding specificities is also much easier relative to ZFNs. Thus, TALENs to almost every possible sequence motif in the genome can be produced without experimental difficulty and with a high success rate. Due to the longer recognition sites of TALENs, these nucleases may cause fewer unwanted off-target effects than ZFNs.

As TALENS are a relatively recent development and plants have a relatively long generation time, few reports have been published thus far documenting the successful application of these enzymes for genome editing in plants. In one study, the authors demonstrated that the binding motif for a pathogen-based transcription factor could be destroyed by TALEN-mediated introduction of a mutation into the promoter of the OsSWEET14 gene, leading to enhanced disease resistance in rice (Li et al., 2012). The group of Dan Voytas introduced targeted mutations within the ALS gene in up to 30% of transformed tobacco protoplasts using TALENs. The same authors designed a TALEN-mediated GT experiment using a donor template that created an in-frame gene fusion between ALS and a YFP marker gene. GT efficiency was measured by quantifying YFP fluorescence using flow cytometry. Approximately 14% of protoplasts showed fluorescence, indicating a very high frequency of GT. In the same study, GT experiments were performed with a TALEN and a 322-bp donor molecule differing by 6-bp from the ALS coding sequence. Even without selection, 4% of the regenerated calli showed evidence of targeted gene replacement (Zhang et al., 2013). These results strongly indicate that TALENs are indeed the most convenient and most effective tools to perform DSB-induced GT in plants today.

RNA guided double-strand break induction: another breakthrough?

Just this year, a bacterial system based on sequence-specific recognition via an RNA molecule for the manipulation of eukaryotic genomes was described: the CRISPR/Cas System (clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR-associated) (for a review, see (Mussolin and Cathomen, 2013)). This system relies on the bacterial endonuclease Cas9, which can be redirected to different target sites simply by modifying the sequence of a single synthetic chimeric guide RNA (sgRNA or crRNA:tracr RNA) (see Fig. 4E). This new technique has been applied to genome editing in a variety of different eukaryotes including plants (Feng et al., 2013; Li et al., 2013; Nekrasov et al., 2013; Shan et al., 2013; Xie and Yang, 2013) and appears to be as efficient as TALENs. What makes the system especially attractive is the very simple design process. One needs only insert the desired DNA oligonucleotide into a vector construct for target site selection, as specificity is solely defined by base-complementarity to the guide RNA. The Cas9 protein does not require any reengineering and has worked well for all of the target sites that have been studied. Moreover, expressing multiple guide RNAs allows multiplexing, which reduces the costs and time needed to generate plants with multiple targeted mutations.

In planta gene targeting

The availability of various tools for site-specific DSB induction makes GT a feasible task, at least for plant species that are easy to transform and regenerate. Unfortunately, this is still not true for most cultivated plant species. For many crop plants, it is still difficult to obtain and regenerate to fertility even a single transgenic plant. Therefore, we recently developed a GT technique that should be applicable to all transformable plant species, even if the transformation efficiency is extremely low (Fauser et al., 2012). The basic idea behind the strategy is that, if one were able to perform the targeting reaction during plant development, then the progeny should carry targeted modifications. Initially, an efficient variation of in vivo GT was developed in Drosophila melanogaster by the group of Kent Golic more than a decade ago (Rong and Golic, 2000). Their method relied on the integration of the targeting vector into the host genome. By induced expression of a site-specific recombinase, a circular DNA was excised from the genome. This circle was then linearized by the simultaneous expression of the homing endonuclease. By induced expression of a site-specific recombinase, a circular DNA was excised from the genome. This circle was then linearized by the simultaneous expression of a homing endonuclease. Thus, GT could be achieved by means of a linearized DNA vector. The set-up of this system was quite complex, because in addition to the construction of the donor sequence with sites for the recombinase and the homing endonuclease, expression cassettes for both enzymes have to be included into the transgene construct or supplied in trans. No successful use of such a system has been reported in plants thus far. Nevertheless, an in vivo approach did not seem unattractive for plants, if it would be possible to simplify the procedure. Therefore, we developed an alternative method using Arabidopsis as a model. Although the scorable marker β-glucuronidase (GUS) and I-SceI were used, the
Integration

Target locus

GT vector for transgene insertion

Modification

Target locus

GT vector for AS exchange

Fig. 5. In planta gene targeting. In addition to the construction of a custom-made nuclease that is able to induce a DSB at a unique genomic site of interest, a transgene construct is produced that harbors the targeting vector as a cassette flanked by two restriction sites of the same nuclease. The targeting vector might contain a gene between the homologous regions that can be used for crop improvement (left) or just a defined modification of one or a few amino acids. After random integration in the plant genome, expression of the endonuclease is induced. This could be achieved by placing the nuclease ORF under the control of an inducible or organ-specific promoter in the same targeting vector outside regions of homology or by expression in trans either by transient or stable transformation. The GT vector is excised and the target locus is activated simultaneously by DSB induction. The vector integrates via HR into the target, leading to a stably modified endogenous locus. Each arrow depicts a recognition site for the nuclease.

The last few years saw a tremendous acceleration in the development of new tools for genome editing. Many options that would have been difficult to envision a few years ago now seem to be possible. Naturally, the application of genome editing tools is more advanced in the animal field as in plants due to its tremendous importance for plants and agriculture. On the one hand, we should not be too optimistic: GT is far from being established in many plant species, and the plant community will still need to spend a tremendous amount of time and expend effort to change this situation. It might well be that, under certain circumstances, only some combinations of custom-made endonucleases, the manipulation of the enzyme machinery and the use of negative selection will lead to feasible GT frequencies in some crop species.

Moreover, due to the recent developments, we not only have many more tools in our hands for genome editing but also must reevaluate our strategies to define which tool is optimal for a specific purpose. For years, GT seemed to be the only way to knock-out gene functions in a directed way. However, as was demonstrated with ZFNs and TALENs, DSB induction and repair via NHEJ in many cases leads to the complete knock-out of gene function. It is technically less demanding to just induce a DSB without simultaneously offering a repair template to the cell. The DSB inducing reagent can be supplied as DNA but also as RNA or even as protein. This type of method has been used successfully for mutant production not only in Arabidopsis but also in several important crop species. Therefore, we expect that future applications of DSB-induced GT
will be restricted to two main tasks: the site-specific integration of foreign genes and the introduction of sophisticated changes, such as the modification of single or a few amino acids in plant genes to modify their function in a controlled way (Fig. 5).

There are multiple ways in which site-specific integration of genetic information will improve our ability to address specific questions in basic science and biotechnology. New sequence information can be targeted to specific regions of the genome. Thus, it will be possible in a very systematic way to discriminate between sequence content and genomic position. Discriminating between these levels is important for different research areas such as gene expression, DNA repair, meiotic recombination and epigenetic modifications. However, the approach is also extremely valuable for agriculture: stacking improved traits at specific loci ("landing pads") in elite cultivars of crop plants is of central interest to plant breeding companies.

The same holds true for using DSB-induced GT to change single or small stretches of amino acids in proteins or single or a few base pairs in DNA control elements. Fine-tuning of genetic information in the original genomic context will allow us to identify connections between enzyme structure and function and modify enzyme expression or specificity. Techniques that create small changes of a few base pairs in the genome also raise the question of how such an organism should regarded in relation to our current understanding of "genetically modified organisms" (GMOs). GMOs are generally regarded as organisms that carry DNA from another species ("foreign" DNA). On the contrary, plants modified by chemical mutagens, gamma radiation or X-rays in breeding programs are regarded as "natural" varieties, although they harbor a heavy mutational load. Taking plant genome sizes into account, statistical basics indicate that a small stretch of deleted or modified DNA sequence at its endogenous locus cannot be regarded as "foreign". Therefore, there is no rational argument for considering plants that have small changes caused by DSB-induced GT or NHEJ to be GMOs. This example demonstrates how far-reaching the consequences of these new genome editing techniques are. Public concerns about the use of GMOs in agriculture might become obsolete in the long run.

Acknowledgements

HP wants to thank all of the colleagues with whom he has directly or indirectly shared efforts to develop genome editing in plants over the last two decades. We regret that the majority of publications dealing with DSB repair and GT in plant systems could not be cited due to space limitations. Moreover, we would like to thank Manfred Focke for his critical reading of the manuscript. Work on GT in our group has been funded over the years by the European Union and the Bundesministerium für Bildung und Forschung (BMBF).

References


Gene targeting in plants...
Further Related Reading, published previously in the Int. J. Dev. Biol.

Female gametophytic mutants of *Arabidopsis thaliana* identified in a gene trap insertion mutagenesis screen
Vladimir B. Brukhin, Miloslawa Jaciubek, Arturo Bolaños Carpio, Vera Kuzmina and Ueli Grossniklaus

Genetic, molecular and cellular approaches to the analysis of maize embryo development
Matilde José-Estanyol, Ignacio López-Ribera, Miriam Bastida, Torben Jarhmann, Nuria Sánchez-Pons, Cristian Becerra, Carlos M. Vicent and Pere Puigdomènech

Analysis of a homologue of the adducin head gene which is a potential target for the *Dictyostelium* STAT protein Dd-STATa
Ryota Aoshima, Rieko Hiraoka, Nao Shimada and Takefumi Kawata

Gametophyte interaction and sexual reproduction: how plants make a zygote
Leonor C. Boavida, Ana Maria Vieira, Jörg D. Becker and José A. Feijó

Plant microRNAs and development
Sara Jover-Gil, Héctor Candela and María-Rosa Ponce

5 yr ISI Impact Factor (2011) = 2.959