Feng Zhang · Holger Puchta
James G. Thomson *Editors*

Advances in New Technology for Targeted Modification of Plant Genomes

Springer

Advances in New Technology for Targeted Modification of Plant Genomes

Feng Zhang • Holger Puchta • James G. Thomson
Editors

Advances in New Technology for Targeted Modification of Plant Genomes

Springer

Editors
Feng Zhang
Cellectis Plant Sciences
New Brighton, MN, USA

James G. Thomson
Crop Improvement and Genetics Unit
USDA Agricultural Research Service
Western Regional Research Center
Albany, CA, USA

Holger Puchta
Botanical Institute
Karlsruhe Institute of Technology
Karlsruhe, Germany

ISBN 978-1-4939-2555-1 ISBN 978-1-4939-2556-8 (eBook)
DOI 10.1007/978-1-4939-2556-8

Library of Congress Control Number: 2015936732

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Contents

Contributors

Peter R. Beetham, Ph.D. Cibus, San Diego, CA, USA

Ann Blechl, Ph.D. Crop Improvement and Genetics Unit, USDA Agricultural Research Service, Albany, CA, USA

Kunling Chen, Ph.D. State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, P. R. China

Michelle Christian, Ph.D. Department of Genetics, Cell Biology and Development, University of Minnesota, Saint Paul, MN, USA

Fayza Daboussi Cellectis Plant Sciences Inc., New Brighton, MN, USA

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

Caixia Gao, Ph.D. State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, P. R. China

Greg F.W. Gocal, M.Sc., Ph.D. Cibus, San Diego, CA, USA

Shigeru Iida, Ph.D. Graduate School of Nutritional and Environmental Sciences, University of Shizuoka, Suruga-Ku, Japan

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

Yiping Qi, Ph.D. Department of Biology, East Carolina University, Greenville, NC, USA

Christian Schöpke, M.Sc., Ph.D. Cibus, San Diego, CA, USA

Thomas J. Stoddard, B.Sc. Cellectis Plant Sciences Inc., New Brighton, MN, USA

James G. Thomson, Ph.D. Crop Improvement and Genetics Unit, USDA Agricultural Research Service, Western Regional Research Center, Albany, CA, USA

Daniel F. Voytas, Ph.D. Department of Genetics, Cell Biology and Development, Center for Genome Engineering, University of Minnesota, Saint Paul, MN, USA

Takaki Yamauchi, Ph.D. Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan

Feng Zhang, Ph.D. Cellectis Plant Sciences Inc., New Brighton, MN, USA

Chapter 1
Double-Strand Break Repair and Its Application to Genome Engineering in Plants

Holger Puchta and Friedrich Fauser

Abstract The induction of double-strand breaks (DSBs) is the basis for the targeted modification of plant genomes. At the same time, the efficient repair of DSBs is important for the survival of all organisms. To efficiently employ DSB repair for genome manipulation using synthetic nucleases, detailed knowledge of the repair process is required. Many aspects of the mechanisms and factors involved in DSB repair have been elucidated in plants over the last two decades. Here, we seek to summarize our current knowledge about the process of DSB repair via non-homologous end joining (NHEJ) as well as homologous recombination (HR) and place this knowledge in the context of strategies applied for genome engineering in plants. While the induction of a unique DSB is generally sufficient for editing single genes, the induction of multiple DSBs can be applied for the engineering of genomes. There is no question that the controlled induction of DSBs exhibits great potential for restructuring plant genomes.

Over the last several years, genome manipulation has been revolutionized by the development of different types of site-specific nucleases for the controlled induction of double-strand breaks (DSBs). This development has especially significant consequences for organisms in which gene targeting (GT) has not been established as a feasible technique. While in many species, such as bacteria or yeast, homologous sequences must be included in a DNA sequence to achieve integration into the genome, random integration of any type of DNA takes place in most multicellular eukaryotes. The establishment of GT in mouse embryonic stem cells represented a major breakthrough (Doetschman et al. 1987; Thomas and Capecchi 1987), allowing not only the knockout of any gene of interest but also various types of controlled genome rearrangements by means of site-specific recombinases (for a recent review, see Turan and Bode 2011). Unfortunately, in many other eukaryotes, GT has not been established at a feasible integration frequency despite decades of attempts (for a review of the situation in plants, see Puchta and Fauser 2013).

H. Puchta (✉) • F. Fauser
Botanical Institute II, Karlsruhe Institute of Technology,
Hertzstr. 16, Karlsruhe 76187, Germany
e-mail: holger.puchta@kit.edu; friedrich.fauser@kit.edu

F. Zhang et al. (eds.), *Advances in New Technology for Targeted Modification of Plant Genomes*, DOI 10.1007/978-1-4939-2556-8_1

1 DSB Induction as a Tool for Genome Manipulation in Plants

It has long been established in yeast that homologous recombination (HR) can be enhanced by the induction of DSBs at specific positions in the genome, which occurs in a controlled manner during mating type switching (for a review, see Paques and Haber 1999). The first indications that DSB induction might also lead to enhanced HR frequencies in plants came from studies involving DNA transposons in maize (Athma and Peterson 1991). An important step in the analysis of DSB repair in plants was the use of site-specific endonucleases, such as the homing endonuclease I-SceI (Perrin et al. 1993). I-SceI contains an 18-mer recognition site and was originally isolated from yeast mitochondria. The I-SceI ORF integrates itself into the mitochondrial 21S rRNA gene by inducing a DSB. This DSB is repaired via HR using a copy of the 21S rRNA gene as a matrix that already includes the I-SceI ORF as an intron. Thus, I-SceI is able to disseminate in the mitochondrial DNA pool (Jacquier and Dujon 1985). The demonstration that I-SceI can induce DSBs in plant cells in vivo (Puchta et al. 1993) showed that a tool was available to the plant community that not only enables various analyses of the basic mechanism of DSB repair but can also be used for genome manipulations. Although other site-specific nucleases have been applied for this purpose (Chilton and Que 2003; Chiurazzi et al. 1996), I-SceI remained the gold standard for studies on DSB repair in plants for quite some time. Transgenic plant lines were produced for homing endonuclease-induced repair experiments that contained an artificial I-SceI site coupled to marker gene sequences within a transgene. Experiments were designed in such a way that the marker was restored due to DSB repair after DSB induction (in some cases, a negative marker gene was also destroyed). As marker restoration was coupled mostly to only one specific type of repair reaction, the respective repair efficiencies could be determined. Moreover, when selection markers were used, the recombinant cells could be isolated and propagated, and the genetic changes could be analyzed in detail at the genome level. Using I-SceI, a proof of principle experiment could be performed showing that by inducing DSBs at a specific site in the genome, GT can be enhanced by orders of magnitude (Puchta et al. 1996). Interestingly, it was also demonstrated that even without the use of homologous sequences, a T-DNA can integrate into an endonuclease-induced DSB (Salomon and Puchta 1998; Chilton and Que 2003; Tzfira et al. 2003).

As discussed in detail in other chapters of this volume, different types of artificial nucleases have been developed over the years and applied more recently for genome engineering in plants. We will not discuss the characteristics of these different classes of enzymes in any detail here but will refer the reader to the respective chapters on meganucleases (MNs, Chap. 2), zinc finger nucleases (ZFNs, Chap. 3), transcription activator-like effector nucleases (TALENs, Chap. 4), and the CRISPR/Cas system (Chap. 5) as well as to a series of recent reviews in the literature (Gaj et al. 2013; Tzfira et al. 2012; Voytas 2013; Puchta and Fauser 2014). In 2005, the group led by Dan Voytas was able to demonstrate that an artificial locus in the tobacco

genome could be targeted using a ZFN with high efficiency (Wright et al. 2005). Four years later, the first reports related to targeting natural genes were published by Dan Voytas for tobacco (Townsend et al. 2009) and by scientists at Dow AgroScience for maize (Shukla et al. 2009). Additionally, using ZFNs, it could be demonstrated by inducing DSBs—first in a transgenic sequence (Lloyd et al. 2005) and later in natural genes (Osakabe et al. 2010; Zhang et al. 2010)—that error-prone repair often results in a loss of gene function. Subsequently, it was also demonstrated that TALENs (Shan et al. 2013a; Mahfouz et al. 2011; Zhang et al. 2013) and engineered homing endonucleases (D'Halluin et al. 2013) as well as the CRISPR/Cas system (Li et al. 2013; Nekrasov et al. 2013; Shan et al. 2013b; Fauser et al. 2014; Feng et al. 2014; Zhang et al. 2014) can be used in a similar manner for these types of genome manipulations in plants.

To better understand the basis of DSB-induced genome engineering in plants, it is important to understand the mechanism of DSB repair in detail. There are two principle ways in which such repair occurs: via HR or via nonhomologous end joining (NHEJ). While in the former pathway, the reaction takes place between sequences that are totally or at least nearly identical to each other, in the latter, the sequence information does not play a major role in the rejoining of the two double strands. Interestingly, there are several mechanisms of HR and NHEJ that can be differentiated. Below, we seek to describe these mechanisms in detail and to summarize current knowledge regarding factors that are involved in these pathways (for reviews, see also Waterworth et al. 2011; Lieberman-Lazarovich and Levy 2011; Puchta 2005).

2 Mechanisms of DSB Repair Involving Homologous Sequences

In principle, there are three main mechanisms of DSB repair involving the use of a homologous template: single-strand annealing (SSA), synthesis-dependent strand annealing (SDSA), and the so-called double-strand break repair (DSBR) model. DSBR is a prominent mechanism for meiotic recombination (Osman et al. 2011). In yeast, there are indications that DSBR also operates in somatic cells, although at a low frequency (Bzymek et al. 2010). While in SDSA and DSBR, the homologous repair template can be supplied in *cis* or *trans*, in the case of the SSA mechanism, recombination only occurs between two directly repeated homologous sequences. These homologies are generally in close proximity on the same chromosome. All three pathways are depicted in Fig. 1.1.

Following DSB induction in all pathways, single-stranded overhangs are produced via exonuclease-catalyzed resection. In SSA, both ends of the break must carry complementary sequences. These molecules can then directly anneal to one another, and a chimeric DNA molecule is formed. If this molecule contains 3′-overhangs, the respective parts of the sequence will be trimmed; otherwise, single-stranded

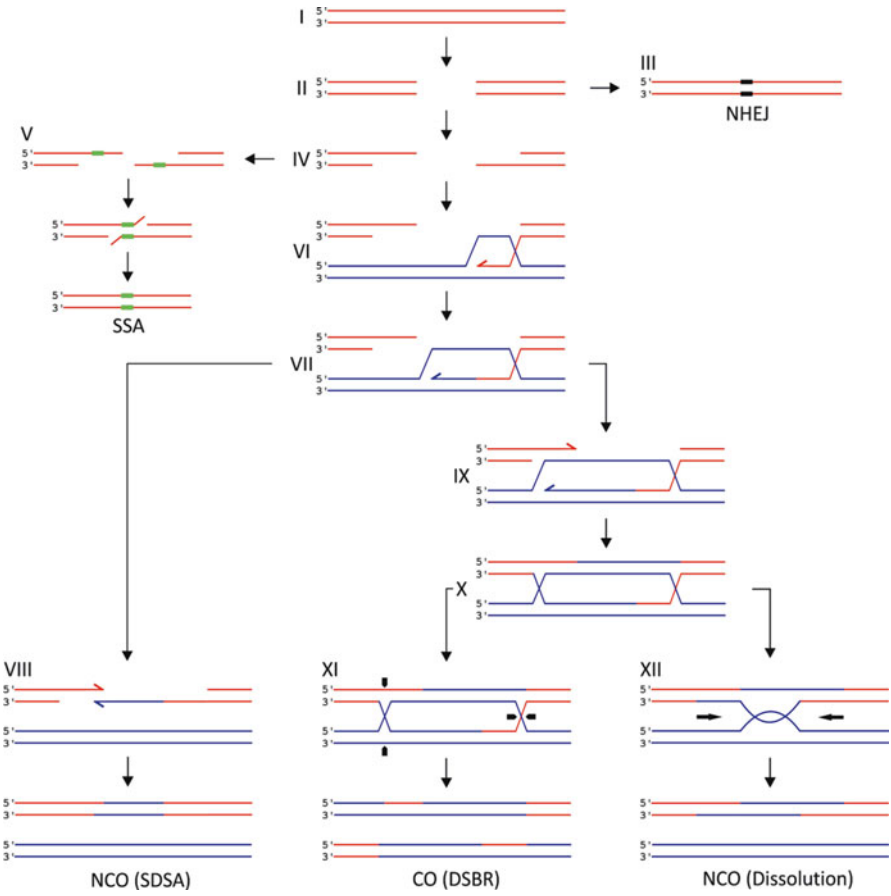

Fig. 1.1 Double-strand break (DSB) repair mechanisms. DSBs can be repaired either via homologous recombination (HR) or nonhomologous end joining (NHEJ). In principle, DSB repair via HR occurs via one of three main mechanisms: the single-strand annealing (SSA), synthesis-dependent strand annealing (SDSA), and double-strand break repair (DSBR) model. A model to explain the dissolution of a double Holliday junction (dHJ) was also added. Following the induction of a DSB (*II*), the cell must choose between NHEJ (*III*) and HR (*IV–XII*). Figure 1.4 depicts the NHEJ pathway in detail. In the case of HR-mediated DSB repair, single-stranded overhangs are produced via exonuclease-catalyzed resection (*IV*). In line with the SSA pathway (*V*), direct annealing of two single-stranded molecules can occur when complementary sequences are present at both ends of the break. After annealing, 3′-overhangs are trimmed, and single-stranded gaps are filled through DNA synthesis. Sequence information that is flanked by the homologies is lost. Therefore, SSA is also classified as a nonconservative HR-mediated DSB repair mechanism. In the case of the conservative HR pathways, a 3′-end invasion into a homologous double strand takes place, resulting in a D-loop (*VI*). Repair synthesis is initiated using the homologous strand as a matrix (*VII*). In contrast to DSBR and dHJ dissolution, SDSA only copies genetic information from the homologous matrix to the invaded single strand, generally leading to noncrossover (NCO) events following re-hybridization with the other end of the break (*VIII*). Once DNA synthesis occurs at the other broken end (*IX*, *X*), a dHJ is formed. This dHJ can either be resolved (in the case of DSBR) or dissolved (in the case of dHJ dissolution). In the case of DSBR, HJ-processing endonucleases are required to make symmetrical cuts. Depending on the orientation of the cuts, DSBR leads to CO or NCO events (*XI*). Dissolution occurs via the action of a RECQ DNA helicase that forms a hemicatenate and a class 1 topoisomerase to carry out strand separation, resulting in an NCO event

regions will be filled in via DNA synthesis. As a consequence, all information shared between the formerly repeated sequences will be lost (Fig. 1.1 V). SSA is a quite efficient mechanism: a study performed in tobacco using I-SceI indicated that up to one out of three DSBs is repaired via SSA under these circumstances (Siebert and Puchta 2002). SSA can, in principle, also occur between two DNA molecules that are not linked. These molecules could be transfected plasmid DNAs (Puchta and Hohn 1991) or T-DNAs (Tinland et al. 1994) as well as broken chromosomes (Pacher et al. 2007).

In the case of DSBR and SDSA, 3′-end invasion of a single strand into a homologous double strand occurs, resulting in a D-loop (Fig. 1.1 VII). Reparative synthesis is initiated using the newly paired strand as a template. From this point onward, the two pathways deviate: whereas in the case of SDSA, the genetic information of the matrix is only copied to one strand (Fig. 1.1 VIII), in DSBR, DNA synthesis also occurs at the other broken end, so that information is copied from both strands of the matrix (Fig. 1.1 XI). This second-end synthesis has major consequences for the further processing of the recombination intermediates. While in SDSA, the extended strand hybridizes with a single strand resulting from the resection of the other end of the break, in the case of DSBR, a double Holliday junction (dHJ) is formed. This structure can be either resolved or dissolved. For resolution, HJ-processing endonucleases are required (Fig. 1.1 XI). Dissolution occurs via the action of a RECQ DNA helicase that forms a hemicatenate and a class 1 topoisomerase that carries out strand separation, resulting in gene conversion (Fig. 1.1 XII). While in SDSA, the reaction always results in gene conversion, in the case of DSBR, the resolution pathway can end in a crossover event. Hence, larger parts of the recombining chromosomes are exchanged, which is of the utmost biological significance, as in meiosis, it is required for mixing parental genomes (Osman et al. 2011).

An interesting question is, of course, whether the SDSA or the DSBR mechanism is mainly used for the repair of DSBs in somatic cells. To discriminate between the two pathways, one must create an experimental situation in which the two mechanisms predict different outcomes. While in DSBR, homologous interactions between the two ends of the DSB are required, in the case of SDSA, the copying process is restricted to one end. Thus, *Agrobacterium*-mediated experiments were performed in tobacco using a T-DNA with homology to only one end of the break. The frequency of the repair of an I-SceI-induced DSB was compared with the results of experiments in which a T-DNA harboring homologous sequences to both ends of the break in the target locus was used. The recombination frequencies obtained with the T-DNA showing homology to both ends of the DSB were only one-third higher than those obtained using the one-ended construct (Puchta 1998). This small difference could be easily explained by one-sided invasion of the unique second homologous end of the respective T-DNA. Thus, homology to only one end of the DSB is sufficient for an efficient HR reaction to occur in plant cells. Somatic homologous DSB repair is initiated by a one-sided initiation event and does not require the other end of the break. These results are only in accordance with the SDSA model, as it is based on one-sided initiation and is therefore able to describe recombination events due to HR as well as to a combination of HR and NHEJ events.

Additionally, general considerations argue against an efficient DSBR mechanism in somatic plant cells. For example, if DSBR operated efficiently in somatic plant cells, crossovers would occur regularly. Additionally, repeated sequences are found at multiple ectopic positions in plant chromosomes, and crossovers between such repeated ectopic sequences would result in di- and acentric chromosomes. Thus, an efficient DSBR pathway would endanger genome stability in somatic cells.

Using the SDSA model, the results of GT experiments in somatic plant cells can be explained. Here, not only were perfect HR-mediated integration events found at the target locus, but recombinants in which only one end of the targeting vector was integrated via homology were often characterized. Alternatively, after copying sequences from the transgenic locus, the vector was observed to be integrated elsewhere in the genome (ectopic targeting). Depending on the experimental system employed, one-sided events could represent the major class of GT events (for review, see Puchta and Fauser 2013). An efficient strategy for reducing one-sided events in addition to random integration events is the use of negative selectable markers for GT (for details, see Chap. 6 of this volume). Additionally, in several studies involving DSB-induced targeting in tobacco, one-sided events were found in a fraction or even in the majority of cases (Puchta et al. 1996; Reiss et al. 2000; Wright et al. 2005).

3 The Chromosomal Site of the Template Makes a Difference

For SSA, the homologous template for the repair reaction of course has to be located in close proximity in direct orientation. This restriction does not apply for the SDSA and DSBR models. During meiosis, controlled recombination between homologs is the main repair mechanism. There are also indications that a portion of the DSBs are repaired during meiosis using the sister chromatid as template—at least in yeast (Goldfarb and Lichten 2010). A prerequisite for efficient allelic recombination in meiosis is the alignment of homologs and the formation of the synaptonemal complex (for a review on meiotic recombination in plants, see Osman et al. 2011).

DSBs must be induced in somatic plant cells for gene and genome engineering. In the case of HR, several different types of homologous sequences might be employed if an SDSA-like repair mechanism operates (Fig. 1.2). In G2 and S phase of the cell cycle, homology to the sister chromatid could be exploited (Fig. 1.2a). However, as this mechanism does not lead to any sequence changes, the frequencies of such events could not be determined for plants. Nevertheless, it must be assumed that this type of repair is likely the most efficient means of repair. Apart from a sequence on the same chromosome (Fig. 1.2b), an allelic sequence on the homologous chromosome (Fig. 1.2c) or an ectopic sequence on a different chromosome (Fig. 1.2d) could be employed. Using I-SceI, the efficiency of different types of templates could be determined based on the restoration of marker genes.

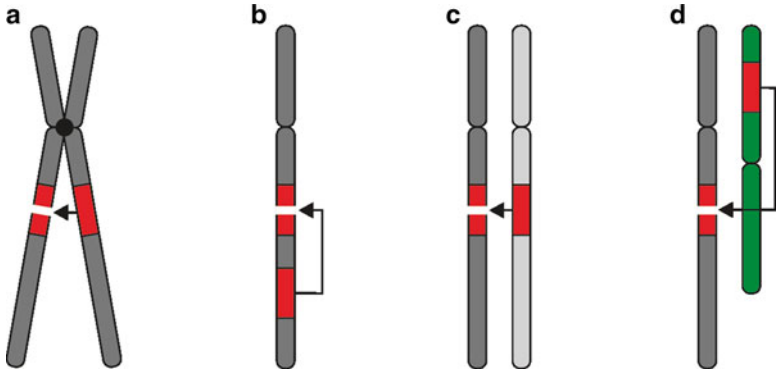

Fig. 1.2 Templates for homologous recombination. Several different types of homologous sequences can be used during the operation of an SDSA-like repair mechanism. (**a**) Homologous sequences on the sister chromatid appear to be used efficiently in G2 and S phases of the cell cycle. (**b**) Intrachromosomal homologous sequences are other potential templates for DSB repair. They are used at frequencies up to the percent range (**c**). Allelic sequences are available in diploid cells, but in contrast to meiotic recombination, they are used in very few cases in somatic cells. (**d**) Ectopic sequences on a different chromosome are also employed as a template, but at least as rarely as allelic sites

To test the efficiency of ectopic DSB-induced HR, a donor and an acceptor locus containing an I-SceI site and carrying partly homologous sequences of a nonfunctional kanamycin resistance gene were transformed independently into tobacco and combined by crossing. DSBs were induced via *Agrobacterium*-mediated transient expression of I-SceI in cell culture, and recombinants were selected for by kanamycin. The study revealed that approximately 1 out of 10,000 DSBs is repaired by the use of an ectopic sequence. Detailed molecular analysis of the recombinants indicated that HR did not occur in all cases at both ends of the DSB, and a combination of HR and NHEJ also took place (Puchta 1999). This finding is a clear indication of the operation of an SDSA-like mechanism (see above). In an independent study conducted in *Arabidopsis* based on DSB induction by a transposon, a similar efficiency of DSB-induced ectopic HR was reported (Shalev and Levy 1997).

As during meiosis, DSBs are mainly repaired using allelic sequences from the homolog, how efficiently such a template can be exploited in somatic plant cells is an interesting question. To answer this question, transgenic tobacco cell culture was used. The experimental setup was such that two transgenes in allelic positions were combined. One transgene carried the negative selectable marker cytosine deaminase (codA), together with an I-SceI site that was originally produced for the analysis of NHEJ (Salomon and Puchta 1998). The other transgene was derived from the former transgenic line. Here, the I-SceI site was destroyed following DSB induction, and the selection marker became nonfunctional due to a deletion associated with the repair event. After combining the two alleles via crossing, DSBs were induced through the transient expression of I-SceI. Selection was performed based on the loss of the marker. Using this approach, it was indeed possible to isolate

DSB-induced allelic recombination events. However, the vast majority of events resulted in a loss of function of the marker gene due to NHEJ. The frequency of allelic recombination was estimated to be approximately 10^{-4}, similar to that of ectopic recombination following DSB repair in the same species (Gisler et al. 2002). Thus, in contrast to meiosis, allelic HR is not a significant DSB repair pathway in somatic cells. It appears that there is not a great difference when homology is present in an allelic versus an ectopic position: as long as it is on a different chromosome it is hardly accessible for DSB-induced HR. With respect to genome engineering, these results indicate that ectopic and allelic HR pathways are not sufficiently efficient to successfully induce controlled genomic changes. Nevertheless, there is also no risk that these pathways will interfere with other applications of DSB-induced manipulation and lead to unwanted genome rearrangements.

The situation is clearly different when homologies are supplied in close proximity to the break: such sequences, either on the same chromosome or on the sister chromatid, are more efficient matrixes for repair. Due to experimental limitations, it has not been possible to address sister chromatid recombination in plant cells in connection with DSB induction for a long time. Very recently, by applying I-SceI it could be demonstrated for barley that a DSB is predominantly repaired using the sister chromatid as a template during S- and G_2-phase (Vu et al. 2014). Thereby, reciprocal chromatid exchanges occur but its molecular mechanisms have not been elucidated yet. However, a series of studies has been published on endonuclease-induced intrachromosomal HR in plants (Chiurazzi et al. 1996; Orel et al. 2003; Roth et al. 2012; Siebert and Puchta 2002). SDSA appears to be approximately five to ten times less efficient than SSA under comparable conditions (Orel et al. 2003).

4 Extrachromosomal Templates

An efficient way to supply the cell with a matrix for HR-mediated DSB repair is to use an incoming T-DNA from *Agrobacterium tumefaciens* or transfected plasmid DNA. A T-DNA that is directly transformed into cells was found to be a better template for homologous DSB repair by several orders of magnitude than an ectopic chromosomal site carrying the same DNA template (Puchta 1999; Puchta et al. 1996). This situation most likely arises because the incoming T-DNA is more readily accessible for the copying processes than a chromosomal site, which can potentially be attributed to steric hindrances related to the use of an ectopic site as template. It might also occur because the ectopic site is chromatin packed or because the incoming DNA is recognized in some way by the cell as damages and, thus, actively recruits DNA repair factors that make recombination reactions more efficient.

Of great interest in this respect is the recently developed *in planta* GT technique (Fauser et al. 2012). Here, a linear DNA molecule is excised from the plant genome using a site-specific endonuclease at the same time as a break is induced in the target locus. Hence, the GT reaction can occur in vivo during plant development. If the induced targeting event is transferred to the progeny, genetically modified seeds can

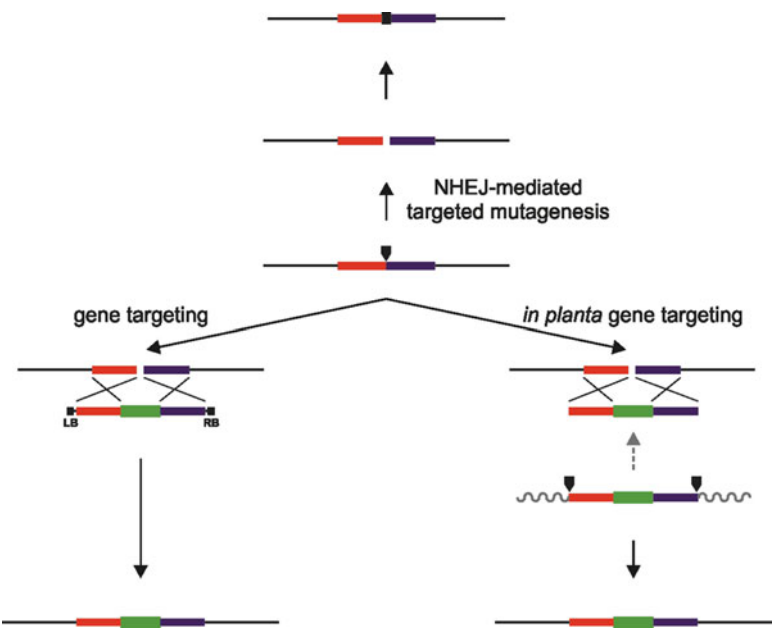

Fig. 1.3 Strategies of gene engineering. Endonuclease-mediated DSB induction can be employed for either targeted mutagenesis or GT. In the case of targeted mutagenesis, NHEJ leads to mutagenic religation of the broken ends. Typically, deletions occur, but insertions as well as nucleotide exchanges are also observed. Two different NHEJ pathways are depicted in Fig. 1.4. However, DSBs can also be used to activate a target locus for GT. DNA molecules transferred into cells carrying homologous sequences to the target locus integrate into the activated target site. On the other hand, the *in planta* GT system allows the simultaneous release of a linear GT vector and activation of a target site. Under both GT strategies, the GT vector can be designed for the site-specific integration of transgenes or to modify the target locus in a predefined manner (e.g., via AS exchange)

be harvested directly. This system was developed in *Arabidopsis* using the scorable marker ß-glucuronidase (GUS, uidA) and I-SceI, and this method should be applicable to any endogenous locus and custom-made nuclease. Its principle is depicted in Fig. 1.3. It is based on a transgene carrying sequences homologous to the target locus that is flanked by two recognition sites for a custom-made endonuclease that also cuts the locus of interest. Expression of the enzyme, which can be modulated by inducible, organ-specific, or constitutive promoters, should lead to the simultaneous release of a linear GT vector and the induction of a DSB at the target locus. Using a constitutive promoter for different target/donor combinations, up to one GT event per 100 seeds could be recovered. Hundreds of seeds exhibiting GT events could therefore be obtained. Molecular analysis of recombinant lines indicated that HR occurred at both ends of the DSB in almost all of these lines (19 out of 20 tested). Additionally, no extra copies of the vector were integrated elsewhere in the genome because in the case of a hemizygous, single-copy transgene, only one copy of the target vector can be set free per genome. Thus, the number of unwanted random integration events is minimized, thus differing from classical GT approaches,

in which multiple copies of a vector are often transferred into a single cell. Moreover, in contrast to classical GT approaches that rely on the generation of a greater number of transformation events, *in planta* GT requires only a single transformation event in principle. Hence, the GT vector as well as the ORF of a synthetic nuclease can be provided simultaneously. As *in planta* GT is more efficient than DSB-induced ectopic recombination, it appears that neither the number of copies nor the chromatin package itself are factors that hinder the DNA from acting as an efficient template. It seems more likely that the steric accessibility of the template is enhanced by the excision and that the excision might recruit repair factors to the template that could enhance the recombination reaction.

5 Factors Involved in Homologous Recombination

Intensive research has been taking place in recent decades, especially in yeast and mammals to identify the key factors involved in HR and define their roles in detail. Here, we wish to exclusively discuss the role of the most prominent factors in relation to what is known about their function in DSB repair in plant cells. The most direct way to characterize the role of individual factors in the different repair pathways is to use recombination traps in which a DSB can be induced by the expression of a site-specific endonuclease, and a marker is restored either via the SSA or SDSA pathway (Orel et al. 2003).

Interestingly, knowledge regarding the factors involved in SSA is quite limited. Early work by Charles White's group indicated that the RAD1/RAD10 heterodimer is as structure-specific flap-like endonuclease that trims the complementary strand prior to ligation (Dubest et al. 2002). Very recently the same group was also able to demonstrate that *Arabidopsis* RAD51 paralogues XRCC2, RAD51B, and RAD51D are involved independently of RAD51 in SSA (Serra et al. 2013). Other results indicate that the nuclease MUS81 and the FANCM helicase play some minor role in SSA (Mannuss et al. 2010). It might well be that different helicases and nucleases can substitute for each other under this quite robust and simple mechanism. SSA and SDSA share the feature of 3′-resection of the double-stranded ends, so that single strands become available for the recombination reaction to proceed. Work in mammals and yeast indicates the existence of two sub-pathways, one involving the DNA exonuclease exo1 and the other involving DNA replication helicase/nuclease 2 (DNA2) and a RECQ helicase (BLM in mammals and SGS1 in yeast) (for a recent review, see Blackwood et al. 2013).

Although it has been shown that the two EXO1 *Arabidopsis* homologs, Exo1A and Exo1B, are involved in the resection of telomeres (Kazda et al. 2012), direct evidence of their involvement in HR is still missing. The fact that overexpression of the rice protein OsRecQl4 (BLM counterpart) and/or OsExo1 (Exo1 homolog) can enhance intrachromosomal HR was taken as an indication that these proteins might, in fact, be involved in end resection in plants (Kwon et al. 2012). Indeed, *Arabidopsis* plants with a deficit of RECQ4A show some deficiency in both the SSA and SDSA

pathways. However, at this time, we cannot exclude the possibility that one of the other six RECQ-like helicases present in the *Arabidopsis* genome (Knoll and Puchta 2011) might also show some function in resection.

Invasion of the resected single strand in the double-stranded matrix is only required for SDSA. In all eukaryotes, RAD51, the homolog of the bacterial recombinase RECA, is involved in this key reaction. In the strand exchange reaction, a set of other factors is also required, which has been characterized in great detail in eukaryotes (for a recent review, see Suwaki et al. 2011). Indeed, recent experiments using pathway-specific recombination traps have demonstrated that the eukaryotic strand exchange protein RAD51 is extremely important for SDSA in *Arabidopsis*, but not for SSA. The same holds true for the SWI2/SNF2 chromatin remodeler AtRAD54 (Roth et al. 2012). BRCA2 is involved in the formation of RAD51 filaments. This function appears to be conserved in plants (e.g., Seeliger et al. 2011). Similar to AtRAD51 itself, its paralogs RAD51C and XXRC3 are essential for SDSA, but not for SSA. It is tempting to speculate that the same holds true for the other paralogs, RAD51B, RAD51D and XRCC2, as general deficiencies of HR have been reported (Da Ines et al. 2013; Durrant et al. 2007). Apart from RAD54, other ATPases appear to play a role in SDSA, including AtRAD5A, which is involved in "post-replicative" strand switching (Mannuss et al. 2010), and AtFANCM (Roth et al. 2012), which is involved in the control of the recombination reaction (Knoll et al. 2012). The nuclease MUS81 is also required for the processing of recombination intermediates within the SDSA pathway (Mannuss et al. 2010).

With respect to genome manipulation, knowledge of the roles of the factors involved in HR is a prerequisite for regulating DSB-induced DNA repair via modulation of the repair machinery of the cell. A classic example of the improvement of GT is provided by the expression of the yeast RAD54 ATPase in *Arabidopsis*, whereby Avi Levy's group was able to demonstrate that GT can be enhanced by the expression of this heterologous chromatin remodeler (Even-Faitelson et al. 2011; Shalev and Levy 1997). An early attempt to enhance DSB-induced GT involved expression of the bacterial strand exchange protein RecA. Unfortunately, this approach did not result in higher targeting frequencies. However, RECA overexpression shifted the obtained product classes towards more recombination events where both ends of the break were repaired via HR. Thus, although the efficiency of the reaction was not altered, its quality was enhanced (Reiss et al. 2000).

As NHEJ, in contrast to HR, often results in genomic change at the break site (see below), when performing NHEJ-mediated targeted mutagenesis, it is attractive to reduce the efficiency of HR pathways. The group led by Dan Voytas was able to achieve this result using a mutant deficient in the "structural maintenance of chromosomes" gene SMC6B. SMC6A and SMC6B are closely related and partly redundant factors required for DNA repair and HR in *Arabidopsis* that appear to be involved in sister chromatid interactions. While it has been reported that SMC6B mutants show a dramatic defect under conditions where sister chromatid recombination can be used for the restoration of a marker (Watanabe et al. 2009), the mutant shows only mild impacts on intrachromatid recombination (Roth et al. 2012). Interestingly, NHEJ-mediated targeted mutagenesis and GT was found to be enhanced at three

different loci of the *Arabidopsis* genome in an smc6b mutant (Qi et al. 2013). The easiest explanation for this phenomenon is that if the sister chromatid is not available to carry out the repair mechanism, an extrachromosomal copy can be used more efficiently as an alternative template. Alternatively, in the case of a lack of template, the DSB is repaired not via HR, but via NHEJ. This study provides a nice demonstration of the importance of template availability in the choice of pathways.

6 DSB Repair via Nonhomologous End Joining

NHEJ is the main mechanism of DSB repair in somatic plant cells and is also required for the random integration of DNA into plant genomes. DSB repair via NHEJ exhibits the characteristic that no extensive stretches of homology are required during the repair reaction. Ends are rejoined more or less directly, which can result in small deletions and, in some cases, also insertions (see below). Such a mechanism ensures that breaks can be efficiently repaired in an absence of homology with little genetic information being lost. However, it also poses the risk that genomic rearrangements might occur if several DSBs are repaired at the same time. This mechanism further provides new options for the control of genome engineering (see below).

Interestingly, at least two different pathways of NHEJ operate in plant cells that are known from other eukaryotes as well. These two pathways can be distinguished by the pattern of the resulting repair junctions as well as by the factors that are involved (Fig. 1.4). The classical NHEJ (cNHEJ) pathway involves minimal processing of broken ends before ligation occurs. This pathway is characterized by the involvement of the Ku70/Ku80 heterodimer. The function of this protein complex is to protect dsDNA from degradation by binding at the broken ends. If necessary, the ends are subsequently processed to become ligateable. For rejoining to occur, a specific ligase (Ligase 4) is necessary (Fig. 1.4). Thus, very little genetic information is lost, and small numbers of identical base pairs ("microhomologies") are found at the junctions only on rare occasion. Nevertheless, if a DSB occurs within an ORF, deletions of one or two nucleotides result in a frameshift and, thus, depending on the position of the break in the ORF, can often lead to complete knockout of gene function.

The main characteristic of the alternative NHEJ pathway (aNHEJ) (Mladenov and Iliakis 2011) is the regular occurrence of microhomologies, combined with the deletion of some nucleotides. In the case of the less well-characterized aNHEJ pathway, a certain amount of 3′-resection of the broken ends occurs, and a junction is formed by annealing of the two single strands involving a few complementary nucleotides. Following end trimming, religation occurs, and microhomologies can be observed at the junction site (Fig. 1.4). This process leads to the occurrence of deletions. Thus, more genetic information is lost more frequently when aNHEJ is used for DSB repair, rather than cNHEJ. Therefore, aNHEJ can be regarded as an extremely mutagenic means of DSB repair. Both NHEJ pathways appear to be con-

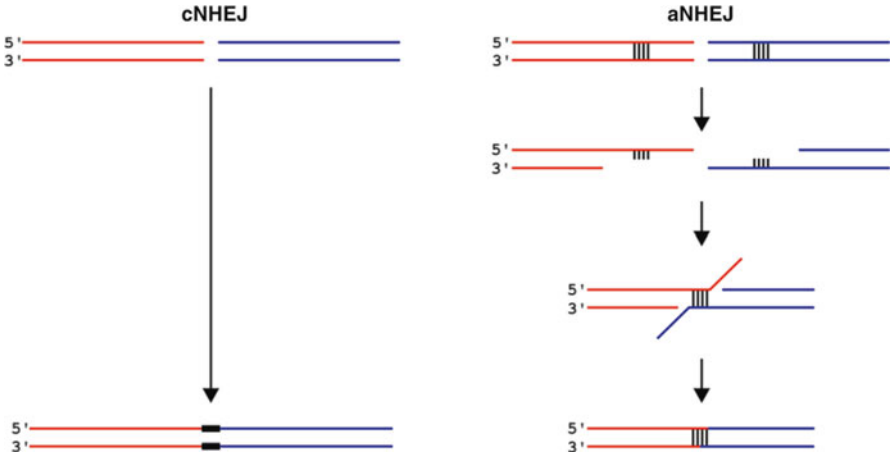

Fig. 1.4 DSB repair via nonhomologous end joining (NHEJ). At least two different NHEJ pathways operate in plant cells. Both are depicted here: the canonical, or classical NHEJ (cNHEJ) pathway and the alternative NHEJ (aNHEJ) pathway. The Ku70/Ku80 heterodimer and Ligase 4 are characteristic of cNHEJ leading to minimal processing of the broken ends prior to ligation. Typically, very little information is lost, and microhomologies are not involved. In contrast, broken ends undergo more processing in aNHEJ, and microhomology-mediated religation occurs. Therefore, larger deletions are often observed following aNHEJ-mediated DSB repair

served among eukaryotes. Recent experiments demonstrated that the poly ADP ribose polymerase I (PARPI) and XRCC1 proteins are actors in this pathway in plants, as previously shown for mammals (Jia et al. 2013; Charbonnel et al. 2011). There is increasing evidence that the two NHEJ pathways compete with each other: in a Ku80 *Arabidopsis* mutant, a 2.6-fold increase in the error-prone rejoining frequency, associated with end-degradation, was documented (Osakabe et al. 2010). Additionally, HR and NHEJ compete for DSBs. The group led by Dan Voytas achieved a fivefold to 16-fold enhancement of DSB-induced GT in a ku70 mutant and a threefold to fourfold enhancement of GT in the lig4 mutant (Qi et al. 2013). It is an important open question whether the different DSB pathways function at the same efficiency in different plant species or cell types during development (Kirik et al. 2000; Lloyd et al. 2005). Moreover, evidence was recently provided that there might be a third NHEJ pathway that is responsible for the joining of at least some DNA ends when cNHEJ and aNHEJ are knocked out (Charbonnel et al. 2011).

The fact that DSB repair via NHEJ can also be associated with insertions is important both for understanding genome evolution as well as for the application of this mechanism. This statement holds true for genomic sequences that are located elsewhere in the genome as well as for incoming T-DNAs (Salomon and Puchta 1998). Interestingly, in most of these cases, microhomologies were found at the junctions between break sites and inserts, which can be taken as a hint that either an aNHEJ mechanism or a mode of copying similar to the SDSA model described

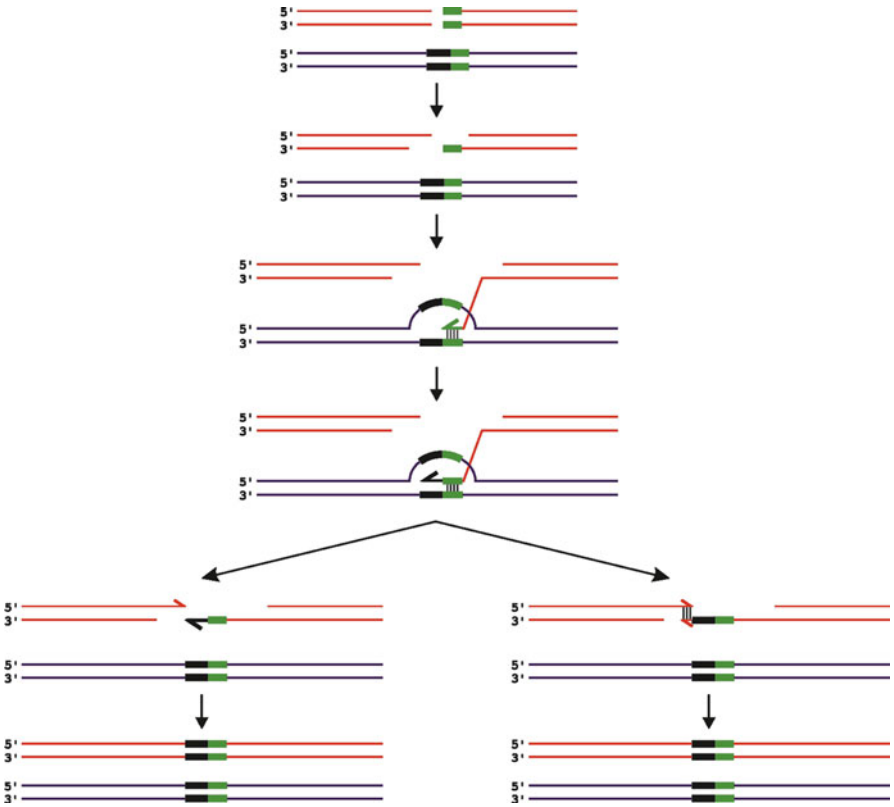

Fig. 1.5 SDSA-like insertions. Single-stranded overhangs are produced after DSB induction and the invasion of the 3′-end via microhomologies initiates a SDSA-like repair mechanism leading to insertions within the original DSB site. Any genomic sequence as well as extrachromosomal DNA can be used as template for the repair reaction. Microhomologies may also mediate second end capture

above for homologous DSB repair might be responsible for this phenomenon. As unique genomic sequences can still be found next to those inserted in the break site at their original location in the genome following DSB repair, a copying mechanism appears to be the most prevalent mode of repair (Salomon and Puchta 1998) as it is shown in Fig. 1.5. Thus, NHEJ can be applied similarly to HR for DSB-induced gene stacking (Chilton and Que 2003; Tzfira et al. 2003; Weinthal et al. 2013).

The potential of NHEJ for genome engineering is especially promising, as more than one DSB can be induced at a time in a plant genome. The principle underlying such applications is simple: in a certain number of cases, not the original linked DSB ends are rejoined, but instead, joining occurs between ends that were not linked previously. Thus, new combinations of genetic information are obtained. This phenomenon can be applied to achieve various types of planned genome rearrangements.

7 From Gene Engineering to Genome Engineering: Inducing More Than One DSB at a Time

The introduction of more than one DSB is a prerequisite for moving from the engineering of single genes to that of genomes. This step can be achieved by targeting either multiple identical sites within the genome using a unique synthetic enzyme or multiple different sites via the application of more than one nuclease. In particular, using the CRISPR/Cas system, it is possible to induce multiple DSBs at different sites simultaneously (Mali et al. 2013; Cong et al. 2013; Wang et al. 2013; Li et al. 2013).

Indeed, the repair of more than one DSB at a time might represent a challenge that cells have to face regularly. Multiple types of DNA damage might arise concurrently, particularly if a cell has to cope with genotoxic stress, but also during DNA replication. Cells are adapted to simultaneously repair several breaks through mechanisms that hold the correct broken ends together (Williams et al. 2010). Thus, the likelihood of complex genome rearrangements due to misrepair or misjoining of the broken ends is reduced. Nevertheless, as these mechanisms are not especially efficient, creating such a situation artificially provides a unique opportunity for genome engineering. It therefore becomes possible to achieve a tremendous number of different types of changes, including deletions, inversions, and exchanges of genomic sequences, and even of chromosome arms (Fig. 1.6).

A series of reports have been published that can be regarded as proof of concept experiments for genome manipulations based on the induction of a number of site-specific DSBs. The most prominent, and simplest, example is the programmed deletion

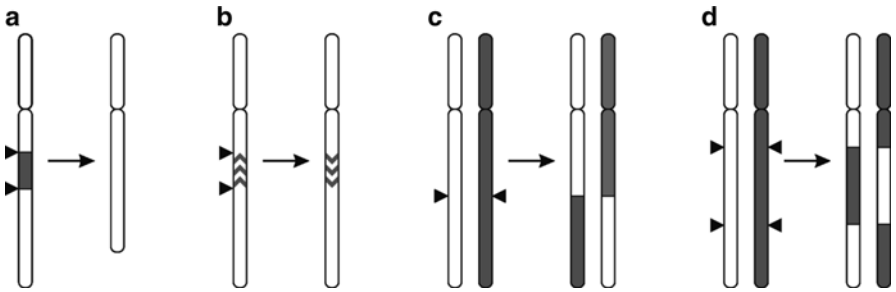

Fig. 1.6 Strategies of genome engineering. Future applications of genome engineering in plants will focus on larger deletions, the inversion of specific genomic regions, the exchange of chromosome arms or the exchange of sequence information between homologs of different accessions or cultivars as well as different chromosomes. Deletions can be induced by inducing two DSBs at a defined distance, whose repair can lead to the elimination of all sequence information between the respective recognition sites (**a**). If this sequence is not eliminated but is reintegrated the "wrong way round" an inversion takes place (**b**). If two DSBs are induced on two different chromosomes, chromosome arms may be exchanged (**c**). By inducing four DSBs (two per chromatid), sequences can also be exchanged between chromosomes (**d**)

of sequences such as marker genes following transformation. Through the induction of two DSBs in relatively close proximity, the sequences between the respective sites can be deleted from the plant genome (Petolino et al. 2010; Siebert and Puchta 2002). In principle, to achieve the controlled deletion of sequences, two differing types of repair reactions can be employed. The most direct way is to join broken ends following elimination of the internal sequence via NHEJ. However, in the presence of direct repeats, the direct annealing of repeated sequences via the SSA mechanism is also possible to obtain a deletion with a junction that can be predicted. Thus, if genomic sequences are organized in tandem repeats, as observed in certain gene clusters, the number of repeats is reduced. The deletion of single genes from the genome should become routine using artificial nucleases. An interesting option for genome engineering is the deletion of larger genomic regions. The size of these deletions might be limited in that genes that are essential for obtaining viable progeny cannot be eliminated. Through the induction of two DSBs, an inversion of the intervening sequence can also be achieved, as demonstrated in mammalian cells (Lee et al. 2012).

In addition to excising or inverting sequences, the controlled induction of more than one DSB can be applied to exchange sequences within a plant genome. This has been demonstrated in a proof of concept experiment involving tobacco chromosome arms, in which two unlinked transgenes, each carrying a restriction site for an endonuclease and parts of an intron containing kanamycin resistance genes were combined via crossing. The transgenes were constructed in such a way that the kanamycin resistance gene could be restored by joining two previously unlinked broken ends, either via SSA or via NHEJ (Pacher et al. 2007). The frequencies obtained using this approach indicate that DSB-induced translocation is up to two orders of magnitude more frequent in somatic cells than DSB-induced ectopic gene conversion (Puchta 1999). Indeed, both SSA and NHEJ events were recovered. Despite the fact that no selection was applied for the joining of the two other ends, the respective linkages could be detected in most of the tested cases, demonstrating that the respective exchanges were indeed reciprocal, as expected.

There are also applications in genome engineering that require the induction of more than two DSBs. One example discussed above is the *in planta* GT technique, which requires the induction of at least three DSBs. By inducing four chromosomal DSBs, it should be possible to exchange chromosomal segments. Here, DSBs should be induced at both ends of each sequence to be exchanged. Such an experiment has not been previously reported. However, the group led by Tzvi Tzfira described an exchange reaction between a chromosomal marker flanked by two ZFN recognition sites that could be exchanged with a DNA sequence in a T-DNA flanked by the same recognition sites (Weinthal et al. 2013).

The ways in which synthetic nucleases can be applied to achieve genome modifications are increasing. It is tempting to speculate that in the long run, DSB-induced genome engineering will result in synthetic plant genomes constructed from the most attractive alleles of a gene pool within, or even beyond, species boundaries.

References

Athma P, Peterson T (1991) Ac induces homologous recombination at the maize P locus. Genetics 128:163–173

Blackwood JK, Rzechorzek NJ, Bray SM, Maman JD, Pellegrini L, Robinson NP (2013) End-resection at DNA double-strand breaks in the three domains of life. Biochem Soc Trans 41:314–320

Bzymek M, Thayer NH, Oh SD, Kleckner N, Hunter N (2010) Double Holliday junctions are intermediates of DNA break repair. Nature 464:937–941

Charbonnel C, Allain E, Gallego ME, White CI (2011) Kinetic analysis of DNA double-strand break repair pathways in Arabidopsis. DNA Repair (Amst) 10:611–619

Chilton MD, 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

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

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823

D'Halluin K, Vanderstraeten C, Van Hulle J, Rosolowska J, Van Den Brande I, Pennewaert A, D'Hont K, Bossut M, Jantz D, Ruiter R, Broadhvest J (2013) Targeted molecular trait stacking in cotton through targeted double-strand break induction. Plant Biotechnol J 11(8):933–941

Doetschman T, Gregg RG, Maeda N, Hooper ML, Melton DW, Thompson S, Smithies O (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature 330:576–578

Dubest S, Gallego ME, White CI (2002) Role of the AtRad1p endonuclease in homologous recombination in plants. EMBO Rep 3:1049–1054

Durrant WE, Wang S, Dong X (2007) Arabidopsis SNI1 and RAD51D regulate both gene transcription and DNA recombination during the defense response. Proc Natl Acad Sci U S A 104:7307

Even-Faitelson L, Samach A, Melamed-Bessudo C, Avivi-Ragolsky N, Levy AA (2011) Localized egg-cell expression of effector proteins for targeted modification of the Arabidopsis genome. Plant J 68:929–937

Fauser F, Roth N, Pacher M, Ilg G, Sanchez-Fernandez R, Biesgen C, Puchta H (2012) In planta gene targeting. Proc Natl Acad Sci 109:7535–7540

Fauser F, Schiml S, Puchta H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant J 79:348–359

Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang D-L, Wang Z, Zhang Z, Zheng R, Yang L, Zeng L, Liu X, Zhu J-K (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. Proc Natl Acad Sci U S A 111:4632–4637

Gaj T, Gersbach CA, Barbas CF 3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31:397–405

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

Goldfarb T, Lichten M (2010) Frequent and efficient use of the sister chromatid for DNA double-strand break repair during budding yeast meiosis. PLoS Biol 8:12

Ines OD, Degroote F, Amiard S, Goubely C, Gallego ME, White CI (2013) Effects of XRCC2 and RAD51B mutations on somatic and meiotic recombination in Arabidopsis thaliana. Plant J 74:959–970

Jacquier A, Dujon B (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41:383–394

Jia Q, Dulk-Ras A, Shen H, Hooykaas PJ, de Pater S (2013) Poly(ADP-ribose)polymerases are involved in microhomology mediated back-up non-homologous end joining in Arabidopsis thaliana. Plant Mol Biol 82:339–351

Kazda A, Zellinger B, Rössler M, Derboven E, Kusenda B, Riha K (2012) Chromosome end protection by blunt-ended telomeres. Genes Dev 26:1703–1713

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

Knoll A, Puchta H (2011) The role of DNA helicases and their interaction partners in genome stability and meiotic recombination in plants. J Exp Bot 62:1565–1579

Knoll A, Higgins JD, Seeliger K, Reha SJ, Dangel NJ, Bauknecht M, Schröpfer S, Franklin FCH, Puchta H (2012) The Fanconi anemia ortholog FANCM ensures ordered homologous recombination in both somatic and meiotic cells in Arabidopsis. Plant Cell 1:1–18

Kwon Y-I, Abe K, Osakabe K, Endo M, Nishizawa-Yokoi A, Saika H, Shimada H, Toki S (2012) Overexpression of OsRecQl4 and/or OsExo1 enhances DSB-induced homologous recombination in rice. Plant Cell Physiol 53:2142–2152

Lee HJ, Kweon J, Kim E, Kim S, Kim JS (2012) Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. Genome Res 22:539–548

Li J-FF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol 31:688–691

Lieberman-Lazarovich M, Levy AA (2011) Homologous recombination in plants: an antireview. Methods Mol Biol 701:51–65

Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci U S A 102:2232–2237

Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu JK (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A 108:2623–2628

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. Science 339:823–826

Mannuss A, Dukowic-Schulze S, Suer S, Hartung F, Pacher M, Puchta H (2010) RAD5A, RECQ4A, and MUS81 have specific functions in homologous recombination and define different pathways of DNA repair in Arabidopsis thaliana. Plant Cell 22:3318–3330

Mladenov E, Iliakis G (2011) Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. Mutat Res 711:61–72

Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol 31:691–693

Orel N, Kyryk A, 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

Osakabe K, Osakabe Y, Toki S (2010) Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. Proc Natl Acad Sci U S A 107:12034–12039

Osman K, Higgins JD, Sanchez-Moran E, Armstrong SJ, Franklin FC (2011) Pathways to meiotic recombination in Arabidopsis thaliana. New Phytol 190:523–544

Pacher M, Schmidt-Puchta W, Puchta H (2007) Two unlinked double-strand breaks can induce reciprocal exchanges in plant genomes via homologous recombination and nonhomologous end joining. Genetics 175:21–29

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

Perrin A, Buckle M, Dujon B (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron–exon junctions. EMBO J 12:2939–2947

Petolino JF, Worden A, Curlee K, Connell J, Strange Moynahan TL, Larsen C, Russell S (2010) Zinc finger nuclease-mediated transgene deletion. Plant Mol Biol 73:617–628

Puchta H (1998) Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. Plant J 13:331–340

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

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56:1–14

Puchta H, Fauser F (2013) Gene targeting in plants: 25 years later. Int J Dev Biol 57:629–637

Puchta H, Fauser F (2014) Synthetic nucleases for genome engineering in plants: prospects for a bright future. Plant J 78:727–741

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

Puchta H, Dujon B, 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 B, 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 U S A 93:5055–5060

Qi Y, Zhang Y, Zhang F, Baller JA, Cleland SC, Ryu Y, Starker CG, Voytas DF (2013) Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. Genome Res 23:547–554

Reiss B, Schubert I, Köpchen K, Wendeler E, Schell J, 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 U S A 97:3358–3363

Roth N, Klimesch J, Dukowic-Schulze S, Pacher M, Mannuss A, Puchta H (2012) The requirement for recombination factors differs considerably between different pathways of homologous double-strand break repair in somatic plant cells. Plant J 72:781–790

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

Seeliger K, Dukowic-Schulze S, Wurz-Wildersinn R, Pacher M, Puchta H (2011) BRCA2 is a mediator of RAD51- and DMC1-facilitated recombination in Arabidopsis thaliana. New Phytol 193:364–375

Serra H, Ines OD, Degroote F, Gallego ME, White CI (2013) Roles of XRCC2, RAD51B and RAD51D in RAD51-independent SSA recombination. PLoS Genet 9(11):E1003971

Shalev G, Levy AA (1997) The Maize transposable element Ac induces recombination between the donor site and an homologous ectopic sequence. Genetics 146:1143–1151

Shan Q, Wang Y, Chen K, Liang Z, Li J, Zhang Y, Zhang K, Liu J, Voytas DF, Zheng X, Gao C (2013a) Rapid and efficient gene modification in rice and Brachypodium using TALENs. Mol Plant 6(4):1365–1368

Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu J-L, Gao C (2013b) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31:686–688

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, Choi VM, Rock JM, Wu YY, Katibah GE, Zhifang G, McCaskill D, Simpson MA, Blakeslee B, Greenwalt SA, Butler HJ, Hinkley SJ, Zhang L, Rebar EJ, Gregory PD, Urnov FD (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature 459:437–441

Siebert R, 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

Suwaki N, Klare K, Tarsounas M (2011) RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis. Semin Cell Dev Biol 22:898–905

Thomas KR, Capecchi MR (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell 51:503–512

Tinland B, Hohn B, Puchta H (1994) Agrobacterium tumefaciens transfers single-stranded transferred DNA (T-DNA) into the plant cell nucleus. Proc Natl Acad Sci U S A 91:8000–8004

Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK, Voytas DF (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459: 442–445

Turan S, Bode J (2011) Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. FASEB J 25:4088–4107

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

Tzfira T, Weinthal D, Marton I, Zeevi V, Zuker A, Vainstein A (2012) Genome modifications in plant cells by custom-made restriction enzymes. Plant Biotechnol J 10:373–389

Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. Annu Rev Plant Biol 64:327–350

Vu GTH, Cao HX, Watanabe K, Hensel G, Blattner FR, Kumlehn J, Schubert I (2014) Repair of site-specific DNA double-strand breaks in barley occurs via diverse pathways primarily involving the sister chromatid. Plant Cell 26:2156–2167

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153:910–918

Watanabe K, Pacher M, Dukowic S, Schubert V, Puchta H, Schubert I (2009) The STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6 complex promotes sister chromatid alignment and homologous recombination after DNA damage in Arabidopsis thaliana. Plant Cell 21:2688–2699

Waterworth WM, Drury GE, Bray CM, West CE (2011) Repairing breaks in the plant genome: the importance of keeping it together. New Phytol 192:805–822

Weinthal DM, Taylor RA, Tzfira T (2013) Nonhomologous end joining-mediated gene replacement in plant cells. Plant Physiol 162:390–400

Williams GJ, Lees-Miller SP, Tainer JA (2010) Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks. DNA Repair (Amst) 9:1299–1306

Wright DA, Townsend JA, Winfrey RJ Jr, Irwin PA, Rajagopal J, Lonosky PM, Hall BD, Jondle MD, Voytas DF (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J 44:693–705

Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T, Joung JK, Voytas DF (2010) High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proc Natl Acad Sci U S A 107:12028–12033

Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, Bogdanove AJ, Voytas DF (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol 161:20–27

Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, Mao Y, Yang L, Zhang H, Xu N, Zhu J-K (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol J 12(6):797–807

Chapter 2
Engineering Meganuclease for Precise Plant Genome Modification

Fayza Daboussi, Thomas J. Stoddard, and Feng Zhang

Abstract Meganucleases, also termed homing endonucleases, are rare-cutting enzymes that are encoded within the genome of nearly all microbes. These enzymes recognize and cleave long DNA sequences (typically 18–30 base pairs) generating double-strand DNA breaks (DSBs). The resulting DSBs can be repaired by different pathways leading to a variety of site-specific DNA modifications, such as insertions, deletions, or point mutations. Over the past 15 years tremendous efforts have been made to engineer a number of variant meganucleases that cleave novel DNA targets. Engineered meganucleases are now being used to generate targeted genomic modifications for a variety of basic and biotechnology applications, including creating valuable traits in crop species.

1 Introduction

The burgeoning demand for plant-derived products, such as food, feed, fuel, and fiber, underlies the importance of developing methods to continuously improve crop varieties with higher yields, lower input costs, and better nutritional value. A variety of technologies have been developed to enable creation of genetic variations with desirable traits. Approaches using mutagens such as ethyl methanesulfonate, gamma rays, fast neutron, and transfer DNA (T-DNA) involves the random introduction of mutations into the genome followed by large-scale screens of mutagenized populations. These methods largely depend on chance and have nearly no control over location and types of mutations recovered from the population. It has been a long sought-after goal to develop methods that could generate specific and targeted genetic mutations in plants. With more than 50 plant genomes being sequenced, genome modification approaches with improved precision will greatly facilitate the dissection of gene function, as well as creating new traits in both basic and applied research (Michael and Jackson 2013; Voytas 2013).

One of the most powerful means to introduce specific changes into genomes is through double-strand DNA break (DSB) repair pathways. Two major pathways,

F. Daboussi • T.J. Stoddard, B.Sc. • F. Zhang, Ph.D. (✉)
Cellectis Plant Sciences Inc., 600 County Road D, Suite 8, New Brighton, MN 55112, USA
e-mail: thomas.stoddard@cellectis.com; fzhang@cellectis.com

F. Zhang et al. (eds.), *Advances in New Technology for Targeted Modification of Plant Genomes*, DOI 10.1007/978-1-4939-2556-8_2

homologous recombination (HR) and nonhomologous end joining (NHEJ), have been employed to incorporate mutations at the break site (Voytas 2013; Puchta and Fauser 2013). In order to utilize DNA repair pathways, targeted DSBs first have to be created in the desired location. Efforts have been focused on developing sequence-specific nucleases that could be engineered to create DSBs in a target locus. To date, four classes of sequence-specific nucleases, meganucleases (homing endonuclease), zinc finger nucleases (ZFNs), transcriptional activator-like effector nucleases (TALENs), and the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9), can be chosen to cleave almost any sequence in any species. While ZFNs, TALENs, and CRISPR/Cas9 technologies are reviewed in other chapters, this review focuses on the state of the art in meganuclease engineering and its applications in plant genome engineering (Voytas 2013; Puchta and Fauser 2013; Small and Puchta 2014).

Meganucleases were the first class of sequence-specific nucleases employed to create targeted DSBs in eukaryote genomes. The very first meganuclease, *I-SecI*, was discovered and characterized from yeast in the 1970s and 1980s (Faye et al. 1979; Bos et al. 1978; Jacquier and Dujon 1985). Encoded by an intron in the mitochondrial large ribosomal RNA subunit (LsrRNA), this enzyme recognizes and cleaves an 18 bp unsymmetrical sequence in an intron-free allele of the LsrRNA gene (Jacquier and Dujon 1985; Stoddard 2014). The resulting break is repaired via HR using the intron-harboring allele as a repair template leading to insertion of the intron into the target site, this process is known as intron homing. Because of its long DNA recognition sequence and high specificity, *I-SceI* can be readily expressed in higher organisms without cleaving host genomes and inducing cytotoxicity. For these reasons it has been used to introduce DSBs at defined locations in studies of DNA repair mechanisms in many eukaryote genomes (Choulika et al. 1995; Rouet et al. 1994a, b). In 1993, Puchta et al. first demonstrated that DSBs induced by *I-SceI* enhance HR in plant species, namely *Nicotiana tabacum* (Puchta 1999; Puchta et al. 1993). This landmark research heralded the arrival of precise genome engineering in plants using sequence-specific nucleases.

Use of meganucleases in precise genome modification requires the ability to engineer nucleases with new sequence specificity. One challenge of meganuclease engineering is that cleavage and DNA-binding domains overlap (Stoddard 2011). When an amino acid sequence is altered to achieve new DNA sequence specificity, the catalytic activity of the enzyme is often compromised. Despite these difficulties, advances in high-throughput screening and protein modeling have made meganucleases easier to engineer and more accessible. This review will discuss the latest developments in customized meganucleases, the strategies to increase meganuclease activity, and the recent applications of engineered meganucleases in plant genome modification.

2 Engineering of Meganucleases for New Sequence Specificity

Although the discovery of the meganuclease dates back to the 1970s, widespread use as tools for genome editing has been hampered by the lack of methods allowing for efficient protein engineering of the protein–DNA interface (Stoddard 2014).

In the 2000s, a semi-rational approach based on computational analysis of structural features in the protein–DNA interface has emerged as a promising strategy to generate meganucleases with new specificities (Chevalier et al. 2002; Ashworth et al. 2006). This approach allowed for mutating specific residues that led to a significant decrease in the complexity of the mutant library to be processed. Using *I-CreI* as a model scaffold, several groups have been able to alter its DNA recognition properties on both a small and large scale, in terms of total proteins reengineered. Initial successes with the *I-CreI* (Seligman et al. 2002; Sussman et al. 2004; Rosen et al. 2006) and *I-SceI* scaffolds (Gimble et al. 2003; Chen and Zhao 2005) that used small-scale libraries resulted in the identification of only a few mutants. Later, Seligman and co-workers demonstrated that substitution of specific individual residues in the *I-CreI* αββαββα fold induces substantial cleavage of novel targets (Seligman et al. 2002). Then Gimble and collaborators modified the DNA-binding domain of *PI-SceI* and selected variant proteins with altered binding specificity using a double-hybrid binding assay (Gimble et al. 2003). While being able to generate a few meganucleases with novel DNA binding specificity, all attempts were limited to modification of a small subset of amino acid. Recently several large-scale screening approaches, including the one developed by our group, has been reported to modulate the DNA-binding and catalytic properties of meganucleases (Jarjour et al. 2009; Volna et al. 2007; Baxter et al. 2013).

The method implemented in our group, termed the two-step semi-rational approach, enables the complete redesign of endonuclease specificity. In this method, the first step consists of randomizing specific residues in the DNA-binding domain of the protein and identifying collections of variants with locally altered specificity. The second step, namely the combinatorial step, is to combine and assemble sets of mutants from different locally engineered variants to create globally engineered proteins with predictable specificity. Using this method, thousands of *I-CreI* mutants with locally altered specificities have been identified in our databank (Epinat et al. 2003; Arnould et al. 2006; Smith et al. 2006).

2.1 Yeast High-Throughput Screening Assay

In order to screen a large number of nuclease variants in the first step, an automated high-throughput screening method has been developed in *Saccharomyces cerevisiae* (Epinat et al. 2003; Chames et al. 2005). As shown in Fig. 2.1, two strains of *S. cerevisiae* were used for the screening of meganuclease activity: the first strain, *MAT*α drives expression of the meganuclease with a galactose-inducible promoter; while the second strain, *MAT*a, carries the substrate composed of an inactive *LacZ* gene under the control of a constitutive promoter. The *LacZ* gene is interrupted by an intervening sequence that is flanked by two direct repeats containing the target of interest and an *I-SceI* target site as an internal control. Upon mating of the two strains, diploid cells were selected with the help of the auxotrophic markers present in both meganuclease and *LacZ* expression vectors. These strains were then transferred onto galactose-rich medium, inducing meganuclease expression. The expressed meganucleases will introduce a DSB at their target leading to the

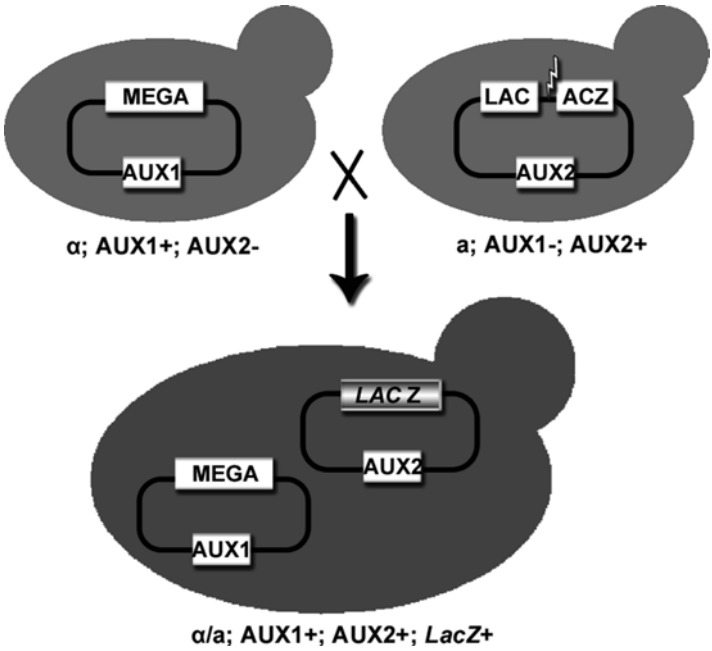

Fig. 2.1 Yeast high-throughput screening assay. One yeast strain, *MAT*α, contains the meganuclease whose expression is driven by a galactose-inducible promoter. Another strain, *MATa*, contains a single-strand annealing (SSA) reporter vector composed of the *LacZ* gene interrupted by the target of interest and an *I-SceI* target site as an internal control, flanked by two direct repeat. After mating and transformed diploid selection, the expression of the enzyme is induced on galactose-rich medium. If the enzyme recognizes and cleaves its target sequence, the *LacZ* reporter recombines by an SSA event. The restoration of a functional *LacZ* gene is measured by X-gal staining, revealing the beta-galactosidase activity

restoration of a functional *LacZ* gene through single-strand annealing mechanisms, allowing for the quantification of the meganuclease activity. This entire process was done on filters that were scanned to rank mutants. Each spot was quantified using the median values of pixels constituting the spot of mated yeast (Epinat et al. 2003).

2.2 *Principle of the Two-Step Semi-rational Approach*

In this method, the meganuclease *I-CreI* was chosen as the molecular scaffold to produce novel meganucleases. As indicated in Fig. 2.2, *I-CreI* works as a homodimer on a pseudo-palindromic target. The DNA–protein interaction is accomplished via a region involving 40 amino acids. As previously reported, each monomer can be divided into sub-domains. For instance, we described two sub-domains referred to as distal (10NNN) and proximal (5NNN) to the catalytic center (Fig. 2.2). These sub-domains interact with the bases in positions ±8, ±9, ±10 and ±3, ±4, ±5 within the

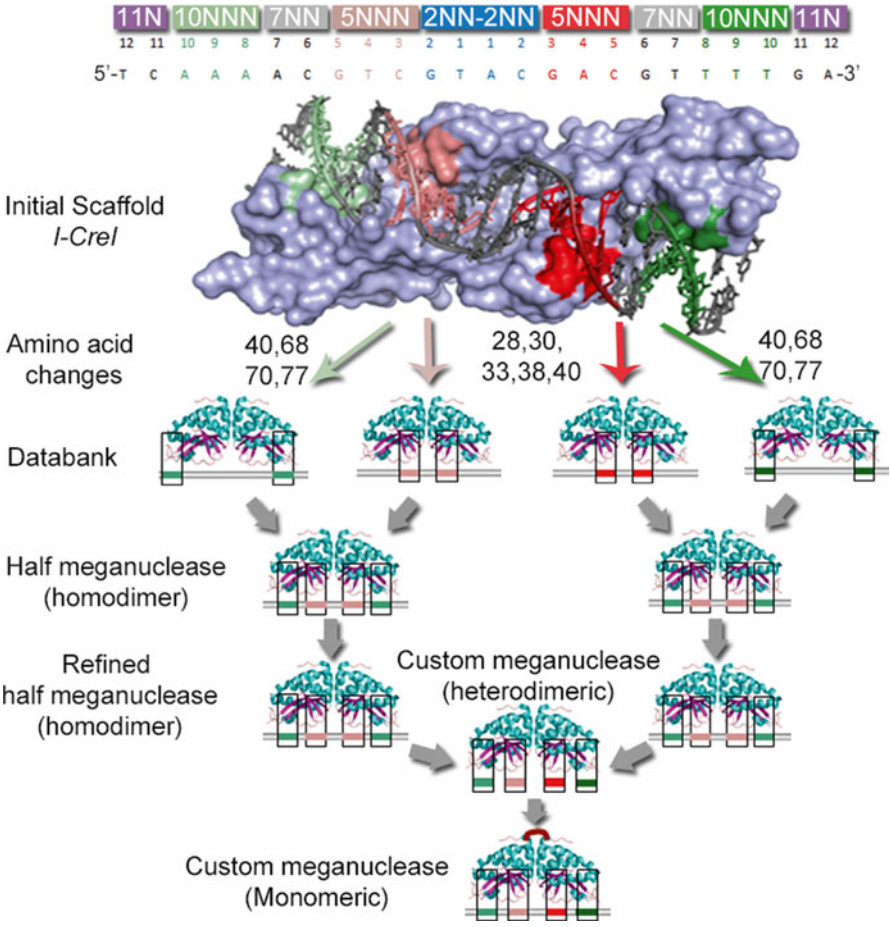

Fig. 2.2 Combinatorial approach to engineer meganuclease. Schematic representation of the 24 bp *I-CreI* target divided into sub-regions: 5NNN internal palindrome corresponding to the base position at +/−5-4-3; 10NNN external palindrome corresponding to the base position at +/−10-9-8; 7NN median cluster corresponding to the base position +/−7-6. The four central base pairs are called 2NN-2NN. Randomization of amino acids interacting with the bases 5NNN and 10NNN are shown in *red* and *green*, respectively. Variants are combined together to cleave palindromic targets as dimers

I-CreI target, respectively. Thus, the production of one meganuclease with a modified DNA-binding interface to recognize a chosen sequence requires the generation of four sets of mutations that affect amino acids interacting with domains 5NNN and 10NNN for each of the two monomers. Key positions of the *I-CreI* protein that are involved in the interaction with sub-domains 5NNN (amino acids 44, 68, 70, and 77) and 10NNN (amino acids 28, 30, 33, 38, and 40) are mutated (Fig. 2.2). Each mutant variant (homodimers containing one set of mutations) is screened using the yeast high-throughput screening assay against the corresponding 64 palindromic degenerate targets. The variants with the highest nuclease activities in the yeast assay are

selected, and then different sets of *I-CreI* mutations are combined in the same protein, resulting in entirely redesigned meganucleases cleaving one chosen sequence. In addition to the mutation of these key residues, the mutation D75N was introduced to allow more diversity in the final combinatorial mutants. This mutation did not affect the protein structure but has been shown to alter the substrate specificity of the 5NNN and the 10NNN sub-domains (Smith et al. 2006).

The two-step combinatorial approach has enabled us to generate several thousand engineered sequence-specific meganucleases leading to the production of a vast collection of variants stored as plasmids in a proprietary databank. To date, the exploitation of this databank allows for the identification of a potential cleavable target every 250 bp within a genome, with a success rate of 40 % at the protein engineering step (Galetto et al. 2009; Paques and Duchateau 2007). The non-negligible failure rate (~60 %) may suggest that sub-domains 5NNN and 10NNN are not always independent although the two sub-domains are far from each other and do not overlap structurally. Recently it has been reported that this context dependence could be overcome by modulating the 7NN region between these two sub-domains (Grizot et al. 2011). Combinatorial tests were performed using 7NNNN (covering regions 5NNC and 7NN) large libraries and confirmed the validity of this strategy to improve nuclease design. Taken together, the two-step combinatorial approach has enabled the generation of customized meganucleases with a greater range of target sites, thus making designer meganucleases applicable for many applications.

3 Meganuclease Scaffold Optimization

Although this combinatorial approach has shown its power to generate meganucleases with new specificity, the binding and/or cleavage properties of the engineered meganucleases can sometimes be suboptimal. In these cases, further scaffold optimization may be required to further improve nuclease activity and specificity. Several strategies have been implemented to achieve this goal.

3.1 Improvement of the Meganuclease Activity

A random mutagenesis approach has been employed to identify mutations in the *I-CreI* scaffold that could enhance nuclease activity (Arnould et al. 2007; Redondo et al. 2008; Grizot et al. 2009). This approach enabled identification of several specific mutations, such as mutations from glycine 19 to serine (G19S) or alanine (G19A), phenylalanine 54 to leucine (F54L), phenylalanine 87 to leucine (F87L), Serine 79 to glycine (S79G), valine 105 to alanine (V105A), and isoleucine 132 to valine (I132V). These mutations, acting alone or in combination, have shown to

significantly increase the activity of both natural and engineered meganucleases. Interestingly, the mutation G19S even displays a dual function by both improving activity and specificity of engineered meganucleases.

3.2 Improvement of the Meganuclease Specificity

In addition to nuclease activity, specificity is another important factor that has to be evaluated in many applications, especially for therapeutic purposes. Most meganucleases are heterodimers that are composed of two separately engineered monomers, each recognizing one half of the DNA target (Stoddard 2011). Heterodimer formation is obtained by co-expression of the two monomers within the same cells. As a result, three combinations of meganuclease molecules are produced within the cell, the heterodimer and both homodimer by-products. The homodimer by-products could lead to off-site cleavage jeopardizing overall meganuclease specificity. Suppression of homodimer formation can be achieved using three strategies. The first route is to generate obligatory heterodimers by modifying the protein–protein interface between the two monomers to facilitate heterodimerization, while preventing the formation of homodimers. The second strategy comprises the fusion of two monomers into a single molecule and the third approach involves the fusion of a meganuclease with a specific DNA-binding domain. Details of these strategies, examples, and potential utility are discussed below.

3.2.1 Obligatory Heterodimer

Engineering of the protein–protein interface between monomers is a common strategy to control protein dimerization. This approach has been applied to heterodimeric meganucleases such as *I-CreI*. Analysis of the *I-CreI* structure allowed the identification of residues Lys7, Glu8, Glu61, Lys96, Arg51, and Asp137 as amino acids involved in electrostatic interactions between monomers (Fajardo-Sanchez et al. 2008). The mutations of K7E/E8K, E61R/K96E, and R51D/D137R in homodimeric *I-CreI* meganucleases have led to the generation of obligate *I-CreI*-derived heterodimeric nucleases (Grizot et al. 2009). In addition, the mutation of G19S has been shown to have a major impact on decreasing homodimer activity, with no decrease in heterodimer activity (Arnould et al. 2007; Redondo et al. 2008). An example of protein–protein interface engineering is the engineered obligate heterodimeric meganuclease targeting the human gene RAG1. These mutant proteins carry only a subset of compensatory mutations (K7E/E8K and E61R/K96E) that proved sufficient to destabilize homodimer formation without any loss of heterodimer activity. Consequently, the resulting meganuclease demonstrated decreased toxicity with reduced off-site cleavage events measured by the formation of phosphorylated H2AX histone foci (γ-H2AX) (Grizot et al. 2009).

3.2.2 Single-Chain Molecule

The second strategy resides in mimicking the natural evolution of meganucleases, such as *I-SceI*, by creating a single-chain molecule (Posey et al. 2004; Chevalier and Stoddard 2001). High resolution three-dimensional structures of several meganucleases, including *I-SceI*, *I-CreI*, *I-DmoI*, *PI-SceI*, and *PI-PfuI*, have made it possible to generate active single-chain meganucleases (Chevalier et al. 2001; Duan et al. 1997; Heath et al. 1997; Hu et al. 1999; Ichiyanagi et al. 2000; Jurica et al. 1998; Poland et al. 2000; Silva et al. 1999). Single-chain meganucleases consist of a simple linker connecting two engineered meganuclease variants (monomers) (Chevalier et al. 2002; Epinat et al. 2003; Grizot et al. 2010). The resulting architecture has proven to significantly increase the activity of the meganuclease and dramatically decrease the cellular toxicity due to off-site targeting, when compared to unlinked meganucleases (Grizot et al. 2009). The single chain *I-CreI*-based custom meganucleases have been successfully designed to target the human endogenous *RAG1* locus efficiently (Grizot et al. 2009).

3.2.3 Fusion HE-TALE

The third strategy to improve meganuclease specificity involves fusing the DNA-binding domain from transcription activator-like effectors (TALE) to the N-terminus or C-terminus of meganucleases. TALE DNA binding domains consist of a series of 33–35 amino acid long repeats. Each repeat recognizes one single nucleotide that is determined by the two amino acids at position 12 and 13, also known as the repeat-variable diresidue (RVD). The code between the RVD and each nucleotide has been deciphered; therefore, new TALE DNA binding domains can be easily engineered according to this simple code. A more detailed description about TALE proteins can be found in Chap. 3 of this book. Several groups have attempted to create a hybrid nuclease by fusing TALE DNA domains with monomeric meganucleases, such as *I-TevI*, *I-AniI*, and *I-OnuI*. The resulting nucleases, termed megaTAL or compact TALEN, have shown great potential with high specificity and activity (Beurdeley et al. 2013; Kleinstiver et al. 2014; Boissel et al. 2014). Because all the meganucleases fused in megaTALs are monomeric endonucleases, the new hybrid nuclease does not require dimerization to cleave target sites. Moreover, DNA binding specificity of megaTALs is determined by both the TALE and meganuclease, reducing the likelihood that more than one target site would be present in the targeted genome. Thus, the fusion nucleases could offer higher specificity than regular meganucleases. In a recent research study, Boissel et al. have compared cleavage activity between megaTALs and their derived meganucleases, *I-AniI* and *I-OnuI* (Boissel et al. 2014). Interestingly, megaTALs showed increased cleavage activity by at least 2–3-fold. Even more remarkably, when an *I-OnuI*-derived megaTAL was employed to knock-out the *TCR*α (T-cell receptor alpha chain) gene it exhibited approximately 20-fold increase in gene knockout frequency, compared to the stand-alone meganuclease.

4 Factors Influencing the Meganuclease Activity In Vivo

In addition to the efforts to increase meganuclease activity through protein engineering, it is also prudent to analyze the meganuclease substrate for opportunities of further optimization. Several studies have shown that the efficacy of meganucleases could be influenced by the chromosome context of the target sequences, such as DNA packaging into chromatin, the position of nucleosomal proteins with respect to the target site, and chemical DNA modifications (e.g., methylation). Below several experiments are reviewed that indicate further optimization is possible to increase meganuclease activity in vivo.

4.1 Meganucleases Are Sensitive to Chromatin Compaction

In order to understand the relationship between meganuclease activity and chromatin accessibility, Daboussi et al. engineered a collection of 18 different *I-CreI-derived* meganucleases that targeted various coding and noncoding sequences within the human genome (Daboussi et al. 2012). This study revealed a strong correlation ($r = 0.93$; $P = 2 \cdot 10^{-8}$) between the frequencies of two types of in vivo mutagenesis activities, NHEJ-meditated targeted mutagenesis (TM) and homologous gene targeting (HGT) (Fig. 2.3a). However, these frequencies are not well correlated ($r = 0.33$; $P = 0.18$) to the intrinsic nuclease activity as measured by extrachromosomal single-strand annealing (SSA) assays (Fig. 2.3b), suggesting that chromosome context of target sequences may play a role in determining meganuclease efficacy. By further using a meganuclease whose cognate target is present several times in the genome, the authors were able to show (a) TM was not equally efficient at different chromosomal loci and (b) that the discrepancy was directly correlated to the chromatin accessibility, as measured by a micrococcal nuclease accessibility assay (CHART-PCR) (Fig. 2.3c). This study was then extended to meganucleases targeting various loci that exhibited different degrees of chromatin compaction, revealing a strong correlation between meganuclease efficiency and chromatin accessibility. These results indicate that target accessibility is a major determinant of designer meganuclease efficacy.

4.2 Meganucleases Are Sensitive to CpG Methylation

In higher eukaryotes, DNA methylation is another key factor influencing chromatin structure and chromosome accessibility (Bell et al. 2011). Valton et al. employed a combination of biochemical and structural approaches to investigate the impact of DNA methylation on meganuclease activity (Valton et al. 2012). Depending on the position of methylated CpGs within the DNA target, DNA methylation could have

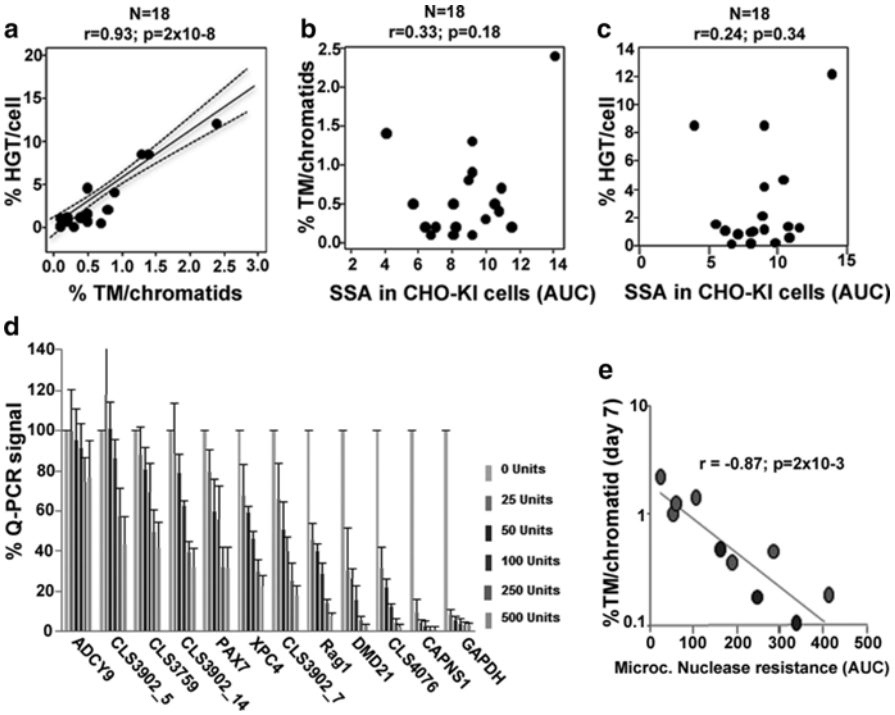

Fig. 2.3 Characterization of meganucleases properties using cell-based assays. (**a**) Correlation between the TM and HGT assays in 293-H cells. 18 MNs were characterized in both assays. (**b**) Correlation between the extrachromosomal SSA and the TM assay in 293-H cells. (**c**) Correlation between the extrachromosomal SSA assays and the HGT assay in 293-H cells. (**d**) CHART-PCR assay for micrococcal nuclease accessibility of the Rag1, DMD21, CLS3902_5, CLS3902_7, CLS3902_14, CLS4076, CAPNS1, ADCY9, CLS3759, and XPC4 loci. The GAPDH and PAX7 loci are used as controls for accessible and non-accessible chromatin, respectively. (**e**) Comparison between the micrococcal nuclease accessibility assay and the TM assay at day 7

a significant impact on the DNA binding affinity and cleavage activity of *I-Cre*. More specifically, methylation that occurs outside the central four bases of the cognate target showed little effect on *I-CreI* activity. However, methylation that occurs inside the central four bases, in particular at position +2b, dramatically decreases catalytic activity of *I-CreI*. This phenomenon was also observed for an engineered meganuclease targeting the human gene ADCY9. TM frequency at ADCY9 was hindered by the presence of one methylated CpG at position +2b. The reversal of low TM frequency by demethylation treatment with 5-aza-2-deoxycytidine (5-aza-dC) further confirmed that DNA methylation found in the core bases of the target sequence plays a critical role in the determination of meganuclease activity (Valton et al. 2012, 2014).

Crystal structure analysis of *I-CreI* with its target sequence indicated the possibility of engineering a meganuclease insensitive to DNA methylation. The structural and biochemical data revealed that the presence of the methyl group alters both the *I-CreI*

protein and the target DNA conformation, in particular the shift of valine 73 by about 1 Å. The substitution of the valine 73 with a smaller hydrophobic amino acid, such as alanine, has been shown to improve the frequency of TM by tenfold in the endogenous methylated locus. It will be interesting to further apply this strategy to engineered meganucleases or other types of designer nucleases (Valton et al. 2014).

4.3 Identification of Host Factors Regulating Homologous Gene Targeting

During the repair process initiated by nuclease-mediated DSBs, the NHEJ-based pathway appears more efficient than the HR-mediated pathway in most higher eukaryotes (Voytas 2013; Puchta and Fauser 2013). Highly efficient TM has been achieved through NHEJ using all types of engineered nucleases, including meganucleases, while precise gene editing through HR still requires a great deal of optimization (Voytas 2013; Puchta and Fauser 2013; Gaj et al. 2013). One strategy to promote the HR-based gene targeting is to modulate host factors in the DNA repair pathway. For instance, suppression of the host factors essential for the NHEJ pathway could shift DNA repair toward HR. In the model plant species *Arabidopsis*, Yiping et al. utilized engineered ZFNs to demonstrate that the frequency of HR-based gene targeting can be increased 3–16-fold by knocking out critical proteins in the NHEJ pathway, such as *Ku70* or *Ligase 4* (Qi et al. 2013). In order to identify novel factors enhancing meganuclease-mediated HGT, Delacote et al. developed a high-throughput siRNA screening system using an assay monitoring HGT induced by *I-SceI* (Delacote et al. 2011). By screening a library of siRNAs targeting 19,121 human genes, a total of 64 candidates were identified as increasing HGT by 1.2–3-fold. Two candidate siRNAs, targeting the *ATF71P* and *EP300* genes, showed the strongest stimulation (3.6 ± 1.0 and 3.0 ± 0.9-fold higher HGT than control, respectively) of HGT. Interestingly, these two genes have been described as transcription factors and are involved in chromatin modeling. Although the mechanisms of stimulating HR by knocking down these genes remain to be elucidated, the result from this study underlies the ongoing efforts to enhance precise genome modification by modulating chromatin accessibility (Segal and Meckler 2013).

4.4 Increased Targeted Mutagenesis Frequency by Co-factors

Compared to other designer nucleases (e.g., ZFNs, TALENs, and CRISPR/Cas9) meganucleases seem to be more proficient at inducing HGT, while frequency of meganuclease-mediated TM is generally lower than other nucleases (Daboussi et al. 2012). Recently, several groups demonstrated that co-transfection of plasmids encoding specific DNA processing enzymes and engineered meganucleases can increase the TM frequency in mammalian cells, without deleterious effects on

genome integrity and/or cell survival (Delacote et al. 2013; Certo et al. 2012). The efficiency can be enhanced up to 30-fold using the terminal deoxynucleotidyl transferase (Tdt) or human three prime repair exonuclease 2 (Trex2), a template-independent DNA polymerase and non-processive 3′ exonuclease, respectively. Upon cleavage, both enzymes process the liberated DNA ends by adding or removing a few nucleotides. The processed DNA ends are then rejoined with small deletions and/or insertions at the break site. These DNA processing enzymes make the NHEJ repair even more error prone and thus dramatically increase NHEJ mutagenesis frequency. Moreover, Tdt and Trex2 offer the possibility to "control" the type of mutagenic events induced, by favoring small insertions or deletions. While the use of Tdt leads to 70–90 % of events being small insertions (2–6 nucleotides), Trex2 usage results in 70–90 % of events yielding small deletions (2–4 nucleotides). This observation has also been made in plant species with Trex2, indicating general applicability of this strategy to enhance TM by coupling DNA end-processing enzymes with designer nucleases (unpublished data).

5 Precise Plant Genome Modification with Meganucleases

Since the 1990s, meganucleases have been employed by plant biologists to understand the basic mechanisms responsible for maintaining plant genome integrity, as well as to create precise modifications for plant genome manipulation. In 1999, Puchta et al. first introduced the meganuclease, *I-SceI*, to tobacco cells (Puchta 1999). Expression of *I-SceI* in tobacco cells created DSBs in a reporter gene that carries the *I-SceI* target site. DSBs were then repaired by HR leading to restoration of the reporter gene function. The resulting proteins from the reporter gene can be readily quantified as a means to measure the efficiency of HR. In general, without creating DSBs, HR in plants is not very efficient with frequencies ranging from 10^{-4} to 10^{-6}. By demonstrating that targeted DSBs can increase HR frequencies by two orders of magnitude in plants, this study opens up the possibility for enabling precise plant genome modification by creating targeted DSBs.

In recent years, efficient targeted genome modification in plants has been achieved with meganucleases. In 2008, D'Halluin et al. demonstrated that *I-SceI* facilitated precise gene integration at a pre-integrated target locus in maize cells (D'Halluin et al. 2008). The target locus contained a transgene sequence that carries an *I-SceI* site, followed by a promoterless selection marker gene (the *bar* gene) and a functional terminator. The donor construct was comprised of a 35S promoter flanked by sequences that share homology to the sequences flanking the *I-SceI* site of the target locus. HR between the donor sequence and the target locus resulted in the insertion of the 35S promoter at the *I-SceI* site, thereby driving expression of the *bar* gene which confers resistance to phosphinothricin (PPT). Positive HR events can be identified by selecting transformed cells on PPT media. As a result, a large number of the PPT-resistant events were recovered when the donor construct was co-transformed with *I-SceI*. Further characterization of these events confirmed the precise integration of the 35S promoter sequence at the *I-SceI* target site.

After the proof-of-concept study using the naturally existing meganuclease *I-SceI*, the same group demonstrated that highly efficient targeted gene insertion can also be achieved using engineered meganucleases (D'Halluin et al. 2013). The goal of this study was to create a gene stack in cotton by inserting two herbicide tolerance transgenes into a defined location adjacent to a pre-integrated insect-resistant transgenic locus. Transgenic cotton lines were created by co-delivering the customized meganucleases expression plasmids, designed to recognize and cleave the target locus, with a donor template plasmid using the biolistic method. The donor construct contains two transgenes, the 4-hydroxyphenylpyruvate dioxygenase gene (*hppd*) of *Pseudomonas fluorescens* and the modified enol-pyruvylshikimate-3-phosphate synthase gene (*epsps*) of *Zea mays*, which confer tolerance to the herbicides, *hppd* inhibitor and glyphosate, respectively. The expression cassettes of these two transgenes were flanked by the cotton genomic sequence, 1.5 kb upstream and 2 kb downstream of the meganuclease cleavage site, to facilitate HR upon creation of DSBs at the target site. After co-bombarded, with both nuclease and donor constructs, the cotton embryonic calli were selected for glyphosate tolerance. The glyphosate-tolerant calli were then selected and subjected to PCR screening to identify the targeted insertion events. As a result, 1,479 glyphosate-tolerant calli were obtained and 27 calli appeared to contain the targeted insertion. Furthermore, about 60–70 % of the targeted insertion events did not contain any additional transgene insertions in the genome, as demonstrated by Southern analysis. The gene targeting efficiency (~2 %) achieved in this study provides the opportunity to enable gene targeting and targeted trait stacking as a routine practice in major crops.

The efficient gene targeting method described above requires an efficient transformation system to generate a large population of transgenic plants. More recently, Puchta's group developed a new strategy to achieve highly efficient targeted insertion independent of an efficient transformation system (Fauser et al. 2012). The proof-of-concept experiment was set up in an *Arabidopsis* line that contains a pre-integrated transgene as the target locus. This target locus consists of an *I-SceI* site upstream of a promoterless β-glucuronidase (GUS) reporter gene. In this strategy, the donor construct, which contains a 35S promoter flanked by two *I-SceI* sites, was transformed and integrated into the plant genome first. The resulting transgenic line was then crossed with an *I-SceI*-expressing line. In the hybrid line, *I-SceI* created DSBs in both the target locus and the donor sequences leading to excision of the donor repair template from its integration site. HR occurring between the cleaved target sequence and the repair template results in the insertion of a 35S promoter in front of the GUS coding sequence. The HR events can then be readily identified by GUS staining. By screening F2 seedlings from the hybrid line, approximately 1 % of plants were identified and confirmed as the recombinant lines with precise HR events. The high efficiency observed in this study makes this strategy very attractive for gene targeting in plant species that don't have efficient transformation systems.

In addition to facilitating targeted gene insertions in plant and animal species, engineered meganucleases have also been employed to create targeted knockout mutations in genes of interest. In 2006 and 2013, the group led by Alexander Lyznik at Dupont/Pioneer Inc customized *I-CreI*-based meganucleases to target two maize loci, *liguless*1 and *ms*26 (Gao et al. 2010; Lyznik et al. 2012; Djukanovic et al. 2013).

When delivered into maize immature embryos, the engineered meganucleases induced indel (insertion/deletion) mutations in the targeted loci and created mutant T0 plants at a frequency of 3 % and 5.8 %, respectively. The majority of mutations were transmissible to the next generation (T1) and gave rise to the anticipated phenotypes. We also observed a similar mutagenesis rate using engineered meganucleases to target a transcription factor gene in rice (unpublished data). Taken together, these studies have proven the general utility of engineered meganucleases to create targeted genome modification in plant species.

6 Future Perspectives

In the past 5 years, sequence-specific nuclease technology has taken a great leap (Voytas 2013). Two new classes of nucleases, TALEN and CRISPR/Cas, have been discovered and utilized. Compared to meganucleases and ZFNs, these nucleases are easier to engineer and have fewer target site selection constraints. Consequently, researchers now have a great collection of tools available to conduct targeted genome modification based on their needs. While TALENs and CRISPR/Cas increasingly become the nuclease reagents of choice, meganucleases still offer unique advantages with the smallest nuclease architecture and high specificity that is particularly valuable for therapeutic applications. Compared to meganucleases, ZFNs and TALENs require dimerization to activate the FokI nuclease domain before DNA can be cleaved, requiring a pair of monomers to be constructed and delivered for a single target site. This architectural requirement limits the potential for ZFN and TALEN delivery into certain cell lines using viral vectors. CRISPR/Cas9 technology exhibits high off-target cleavage when used with single-guide RNAs, necessitating the use of nicking enzymes; therefore, two guide RNAs per target are required to achieve specificity equivalent to that of the other major platforms. The need to deliver both the Cas9 ORF plus two guides per target would appear to pose significant limitations for the use of CRISPR/Cas9 in therapeutic applications requiring viral vectorization. The compact architecture of meganucleases (only 1 kb in size) makes them well suited for therapeutic applications (Stoddard 2014; Boissel et al. 2014).

The most challenging limitation of meganuclease technology is the requirement for significant engineering efforts to produce enzymes with novel DNA specificity as cleavage and binding activity in meganucleases are tightly coupled. Two strategies have been developed to address the limitation of the present meganuclease platform. One is to characterize a large number of naturally occurring endonucleases in the LAGLIDADG meganuclease family (Taylor and Stoddard 2012). This approach helps expand the collection of meganucleases with distinct target site specificity for genome engineering applications. The other strategy is to create a hybrid nuclease architecture by combining the TALE DNA binding domain, due to ease of engineering and versatility, with the monomeric nuclease domain from meganucleases (Beurdeley et al. 2013; Kleinstiver et al. 2014; Boissel et al. 2014).

The resulting nucleases, termed as megaTAL or compact TALEN, have shown great potential with high activity and compact architectures.

Taken together, a large number of engineered meganucleases have been successfully developed for targeted genome modification in numerous organisms, including human, mouse, maize, cotton, tobacco, and rice (Arnould et al. 2007; Grizot et al. 2009; Daboussi et al. 2012; Valton et al. 2012; D'Halluin et al. 2008; D'Halluin et al. 2013; Gao et al. 2010; Menoret et al. 2013; Gouble et al. 2006; Rousseau et al. 2011; Munoz et al. 2011; Popplewell et al. 2013; Dupuy et al. 2013; Grosse et al. 2011). This list is expected to expand as nuclease efficacy and engineering capacity are constantly improved (Jarjour et al. 2009; Thyme et al. 2009). Coupled with the growing list of characterized, engineered, and naturally existing meganucleases, it can be anticipated that meganucleases will continue to provide valuable tools and resources for precise genome engineering in both plant and animal species.

References

Arnould S, Chames P, Perez C et al (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. J Mol Biol 355(3): 443–458

Arnould S, Perez C, Cabaniols JP et al (2007) Engineered *I-CreI* derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. J Mol Biol 371(1):49–65

Ashworth J, Havranek JJ, Duarte CM et al (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. Nature 441(7093):656–659

Baxter SK, Lambert AR, Scharenberg AM, Jarjour J (2013) Flow cytometric assays for interrogating LAGLIDADG homing endonuclease DNA-binding and cleavage properties. Methods Mol Biol 978:45–61

Bell O, Tiwari VK, Thoma NH, Schubeler D (2011) Determinants and dynamics of genome accessibility. Nat Rev Genet 12(8):554–564

Beurdeley M, Bietz F, Li J et al (2013) Compact designer TALENs for efficient genome engineering. Nat Commun 4:1762

Boissel S, Jarjour J, Astrakhan A et al (2014) megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. Nucleic Acids Res 42(4):2591–2601

Bos JL, Heyting C, Borst P, Arnberg AC, Van Bruggen EF (1978) An insert in the single gene for the large ribosomal RNA in yeast mitochondrial DNA. Nature 275(5678):336–338

Certo MT, Gwiazda KS, Kuhar R et al (2012) Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. Nat Methods 9(10):973–975

Chames P, Epinat JC, Guillier S, Patin A, Lacroix E, Paques F (2005) In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. Nucleic Acids Res 33(20):e178

Chen Z, Zhao H (2005) A highly sensitive selection method for directed evolution of homing endonucleases. Nucleic Acids Res 33(18):e154

Chevalier BS, Stoddard BL (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res 29(18):3757–3774

Chevalier BS, Monnat RJ Jr, Stoddard BL (2001) The homing endonuclease *I-CreI* uses three metals, one of which is shared between the two active sites. Nat Struct Biol 8(4):312–316

Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ, Stoddard BL (2002) Design, activity, and structure of a highly specific artificial endonuclease. Mol Cell 10(4):895–905

Choulika A, Perrin A, Dujon B, Nicolas JF (1995) Induction of homologous recombination in mammalian chromosomes by using the *I-SceI* system of Saccharomyces cerevisiae. Mol Cell Biol 15(4):1968–1973

Daboussi F, Zaslavskiy M, Poirot L et al (2012) Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. Nucleic Acids Res 40(13):6367–6379

Delacote F, Perez C, Guyot V et al (2011) Identification of genes regulating gene targeting by a high-throughput screening approach. J Nucleic Acids 2011:947212

Delacote F, Perez C, Guyot V et al (2013) High frequency targeted mutagenesis using engineered endonucleases and DNA-end processing enzymes. PLoS One 8(1):e53217

D'Halluin K, Vanderstraeten C, Stals E, Cornelissen M, Ruiter R (2008) Homologous recombination: a basis for targeted genome optimization in crop species such as maize. Plant Biotechnol J 6(1):93–102

D'Halluin K, Vanderstraeten C, Van Hulle J et al (2013) Targeted molecular trait stacking in cotton through targeted double-strand break induction. Plant Biotechnol J 11(8):933–941

Djukanovic V, Smith J, Lowe K et al (2013) Male-sterile maize plants produced by targeted mutagenesis of the cytochrome P450-like gene (MS26) using a re-designed *I-CreI* homing endonuclease. Plant J 76(5):888–899

Duan X, Gimble FS, Quiocho FA (1997) Crystal structure of P*I-SceI*, a homing endonuclease with protein splicing activity. Cell 89(4):555–564

Dupuy A, Valton J, Leduc S et al (2013) Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALEN. PLoS One 8(11):e78678

Epinat JC, Arnould S, Chames P et al (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. Nucleic Acids Res 31(11):2952–2962

Fajardo-Sanchez E, Stricher F, Paques F, Isalan M, Serrano L (2008) Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. Nucleic Acids Res 36(7):2163–2173

Fauser F, Roth N, Pacher M et al (2012) In planta gene targeting. Proc Natl Acad Sci U S A 109(19):7535–7540

Faye G, Dennebouy N, Kujawa C, Jacq C (1979) Inserted sequence in the mitochondrial 23S ribosomal RNA gene of the yeast Saccharomyces cerevisiae. Mol Gen Genet 168(1):101–109

Gaj T, Gersbach CA, Barbas CF 3rd, ZFN (2013) TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31(7):397–405

Galetto R, Duchateau P, Paques F (2009) Targeted approaches for gene therapy and the emergence of engineered meganucleases. Expert Opin Biol Ther 9(10):1289–1303

Gao H, Smith J, Yang M et al (2010) Heritable targeted mutagenesis in maize using a designed endonuclease. Plant J 61(1):176–187

Gimble FS, Moure CM, Posey KL (2003) Assessing the plasticity of DNA target site recognition of the P*I-SceI* homing endonuclease using a bacterial two-hybrid selection system. J Mol Biol 334(5):993–1008

Gouble A, Smith J, Bruneau S et al (2006) Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. J Gene Med 8(5):616–622

Grizot S, Smith J, Daboussi F et al (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Nucleic Acids Res 37(16):5405–5419

Grizot S, Epinat JC, Thomas S et al (2010) Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. Nucleic Acids Res 38(6):2006–2018

Grizot S, Duclert A, Thomas S, Duchateau P, Paques F (2011) Context dependence between subdomains in the DNA binding interface of the *I-CreI* homing endonuclease. Nucleic Acids Res 39(14):6124–6136

Grosse S, Huot N, Mahiet C et al (2011) Meganuclease-mediated Inhibition of HSV1 Infection in Cultured Cells. Mol Ther 19(4):694–702

Heath PJ, Stephens KM, Monnat RJ Jr, Stoddard BL (1997) The structure of I-Crel, a group I intron-encoded homing endonuclease. Nat Struct Biol 4(6):468–476

Hu D, Crist M, Duan X, Gimble FS (1999) Mapping of a DNA binding region of the P*I-SceI* homing endonuclease by affinity cleavage and alanine-scanning mutagenesis. Biochemistry 38(39):12621–12628

Ichiyanagi K, Ishino Y, Ariyoshi M, Komori K, Morikawa K (2000) Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. J Mol Biol 300(4):889–901

Jacquier A, Dujon B (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41(2):383–394

Jarjour J, West-Foyle H, Certo MT et al (2009) High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. Nucleic Acids Res 37(20): 6871–6880

Jurica MS, Monnat RJ Jr, Stoddard BL (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease *I-CreI*. Mol Cell 2(4):469–476

Kleinstiver BP, Wang L, Wolfs JM et al (2014) The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. G3 (Bethesda) 4(6):1155–1165

Lyznik LA, Djukanovic V, Yang M, Jones S (2012) Double-strand break-induced targeted mutagenesis in plants. Methods Mol Biol 847:399–416

Menoret S, Fontaniere S, Jantz D et al (2013) Generation of Rag1-knockout immunodeficient rats and mice using engineered meganucleases. FASEB J 27(2):703–711

Michael TP, Jackson S (2013) The first 50 plant genomes. Plant Genome 6(2). doi:10.3835/plantgenome2013.03.0001in

Munoz IG, Prieto J, Subramanian S et al (2011) Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. Nucleic Acids Res 39(2):729–743

Paques F, Duchateau P (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr Gene Ther 7(1):49–66

Poland BW, Xu MQ, Quiocho FA (2000) Structural insights into the protein splicing mechanism of P*I-SceI*. J Biol Chem 275(22):16408–16413

Popplewell L, Koo T, Leclerc X et al (2013) Gene correction of a duchenne muscular dystrophy mutation by meganuclease-enhanced exon knock-in. Hum Gene Ther 24(7):692–701

Posey KL, Koufopanou V, Burt A, Gimble FS (2004) Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. Nucleic Acids Res 32(13):3947–3956

Puchta H (1999) Use of I-Sce I to induce DNA double-strand breaks in Nicotiana. Methods Mol Biol 113:447–451

Puchta H, Fauser F (2013) Gene targeting in plants: 25 years later. Int J Dev Biol 57(6–8):629–637

Puchta H, Dujon B, 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(22):5034–5040

Qi Y, Zhang Y, Zhang F et al (2013) Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. Genome Res 23(3):547–554

Redondo P, Prieto J, Munoz IG et al (2008) Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. Nature 456(7218):107–111

Rosen LE, Morrison HA, Masri S et al (2006) Homing endonuclease *I-CreI* derivatives with novel DNA target specificities. Nucleic Acids Res 34(17):4791–4800

Rouet P, Smih F, Jasin M (1994a) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol 14(12):8096–8106

Rouet P, Smih F, Jasin M (1994b) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci U S A 91(13):6064–6068

Rousseau J, Chapdelaine P, Boisvert S et al (2011) Endonucleases: tools to correct the dystrophin gene. J Gene Med 13(10):522–537

Segal DJ, Meckler JF (2013) Genome engineering at the dawn of the golden age. Annu Rev Genomics Hum Genet 14:135–158

Seligman LM, Chisholm KM, Chevalier BS et al (2002) Mutations altering the cleavage specificity of a homing endonuclease. Nucleic Acids Res 30(17):3870–3879

Silva GH, Dalgaard JZ, Belfort M, Van Roey P (1999) Crystal structure of the thermostable archaeal intron-encoded endonuclease I-DmoI. J Mol Biol 286(4):1123–1136
Small I, Puchta H (2014) Emerging tools for synthetic biology in plants. Plant J 78(5):725–726
Smith J, Grizot S, Arnould S et al (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res 34(22):e149
Stoddard BL (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. Structure 19(1):7–15
Stoddard BL (2014) Homing endonucleases from mobile group I introns: discovery to genome engineering. Mob DNA 5(1):7
Sussman D, Chadsey M, Fauce S et al (2004) Isolation and characterization of new homing endonuclease specificities at individual target site positions. J Mol Biol 342(1):31–41
Taylor GK, Stoddard BL (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. Nucleic Acids Res 40(12):5189–5200
Thyme SB, Jarjour J, Takeuchi R et al (2009) Exploitation of binding energy for catalysis and design. Nature 461(7268):1300–1304
Valton J, Daboussi F, Leduc S et al (2012) 5'-Cytosine-phosphoguanine (CpG) methylation impacts the activity of natural and engineered meganucleases. J Biol Chem 287(36): 30139–30150
Valton J, Cabaniols JP, Galetto R et al (2014) Efficient strategies for TALEN-mediated genome editing in mammalian cell lines. Methods 69(2):151–170
Volna P, Jarjour J, Baxter S et al (2007) Flow cytometric analysis of DNA binding and cleavage by cell surface-displayed homing endonucleases. Nucleic Acids Res 35(8):2748–2758
Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. Annu Rev Plant Biol 64:327–350

Chapter 3
High Efficient Genome Modification by Designed Zinc Finger Nuclease

Yiping Qi

Abstract Zinc finger nucleases (ZFNs) are engineered artificial nucleases for inducing site-specific DNA double-strand breaks (DSBs) at the genomic loci of interest. There are two major DNA repair mechanisms in plant cells: nonhomologous end joining (NHEJ) and homologous recombination (HR). Repair of the DNA DSBs can lead to different genome modifications, such as targeted mutagenesis or chromosomal deletions by NHEJ, and gene targeting by HR. Thanks to the recent improvement on ZFN-engineering platforms, diverse ZFN-mediated genome modifications have been achieved in different plant species, such as Arabidopsis, tobacco, petunia, maize, and soybean. The frequencies of ZFN-mediated genome modifications are typically high in somatic cells, but quite low in germinal cells. In order to fully empower ZFNs in plant genome engineering, the challenges such as low germinal transmission frequencies and cellular cytotoxicity will need to be addressed in the near future.

1 Introduction

Precise genome modifications are facilitated by site-specific DNA double-strand breaks (DSBs). Upon induction of a DNA DSB, the broken chromosome has to be repaired to allow the cell cycle to proceed. Nonhomologous end joining (NHEJ) and homologous recombination (HR) are two major DNA repair pathways in plant cells (Puchta 2005; Voytas 2013). The NHEJ pathway is error-prone and can lead to mutations, while the HR pathway allows precise sequence modification due to its homology-based nature. Engineered site-specific nucleases can target DNA DSBs in a sequence-specific manner and thus will facilitate different outcomes of genome modification, depending on which DNA repair pathway is used. Zinc figure nuclease (ZFN) is one kind of site-specific nucleases that has been widely used in genome modifications in many higher organisms including both model and crop plants

Y. Qi, Ph.D. (✉)
Department of Biology, East Carolina University,
N108 Howell Science Complex, Mail Stop 551, Greenville, NC 27858, USA
e-mail: qiy@ecu.edu

F. Zhang et al. (eds.), *Advances in New Technology for Targeted Modification of Plant Genomes*, DOI 10.1007/978-1-4939-2556-8_3

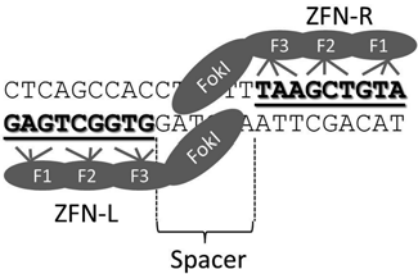

Fig. 3.1 A ZFN pair and its target site. Schematic of a ZFN pair binding to its target site. In this illustration, each ZFN monomer (ZFN-L or ZFN-R) contains a three-finger DNA-binding domain and a FokI nuclease domain. Each zinc finger recognizes a 3-bp sequence. Dimerization of the ZFN pair renders recognition of a specific 18-bp DNA sequence and allows the FokI nuclease to cut the DNA between the two recognition sites—a 5–7 bp region called "spacer"

(Voytas 2013; Carroll 2011; Weinthal et al. 2010; Tzfira et al. 2012). ZFNs functions as dimers while each monomer contains a zinc finger DNA-binding domain and a nonspecific FokI nuclease domain (Fig. 3.1) (Kim et al. 1996; Bibikova et al. 2001). The zinc finger domain typically consists of three to four zinc finger DNA-binding motifs, with each one recognizing 3-nt DNA sequence. Thus, a ZFN pair can recognize an 18 or 24 bp DNA sequence, which is highly specific in the genome. The modularity of DNA binding by zinc fingers makes engineering site-specific ZFNs possible. In this chapter, I will review how to engineer ZFNs, summarize what have been done with ZFNs in modifying plant genomes (Fig. 3.2 and Table 3.1), and briefly discuss how to further improve the use of ZFNs in plants.

2 Making and Testing ZFNs

In the past few years, multiple platforms for engineering ZFNs have been developed. Briefly, they can be grouped into two categories. The first category is modular assembly, which put together individual fingers with precharacterized DNA-binding specificities (Bae et al. 2003; Beerli and Barbas 2002; Liu et al. 2002; Segal et al. 2003; Mandell et al. 2006). Although straightforward, modular assembly generally has a low successful rate in generating active ZFNs due to low activities and/or high toxicities (Ramirez et al. 2008; Cornu et al. 2008). With a prescreen for enriching fingers that were often found in functional ZFNs, a higher successful rate (23 %) was reported for modular assembly (Kim et al. 2011). More recently, Bhakta et al. developed an extended modular assembly method with a much higher successful rate (71 %) when variant lengths of zinc finger arrays (ZFAs) were tested (Bhakta et al. 2013). The second category involves prescreening of multi-finger archives by taking account of context-dependency among neighboring fingers. This category includes the proprietary platform developed by Sangamo Bioscience (Doyon et al.

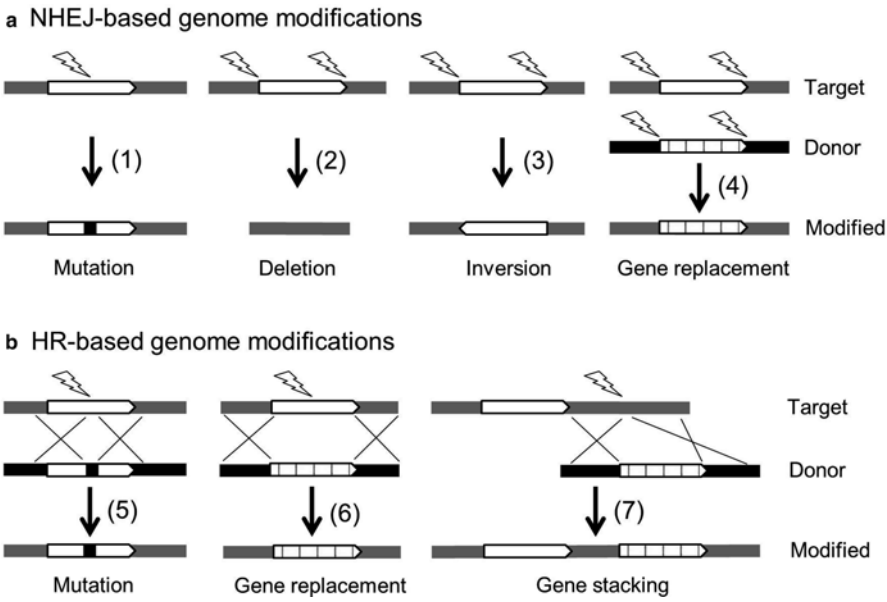

Fig. 3.2 ZFN-mediated plant genome modifications. (**a**) ZFN-mediated plant genome modifications based on NHEJ. (*1*) Imprecise repair of a DNA DSB can result in mutations which are typically small insertions and deletions (Indels). Two simultaneous DNA DSBs may result in chromosomal deletions (*2*) and inversions (*3*). Two simultaneous DNA DSBs at the target with two concurrent ones at the donor by the same ZFN can facilitate gene replacement (*4*). (**b**) ZFN-mediated plant genome modifications based on HR. Depending on the design of target sites and donors, repair of a DNA DSB can lead to different outcomes such as defined point mutations (*5*), gene replacement (*6*), and gene stacking (*7*). Detailed examples of all these ZFN-mediated genome modifications in plants (*1*–*7*) were summarized in Table 3.1

2008), which is commercially available with the brand name CompoZr® through Sigma-Aldrich®. This category also includes some methods that were developed in academics, such as OPEN (Oligomerized Pool ENgineering) (Maeder et al. 2008), CoDA (Cotext-Dependent Assembly) (Sander et al. 2010a), and an approach using an optimized two-finger archive (Gupta et al. 2012).

Many web-based tools are available for designing and selecting ZFNs. Zinc Finger Tools allows users to search for ZFNs which can be made by modular assembly (Mandell et al. 2006). ZiFit can identify potential ZFN sites in a given DNA sequence to facilitate engineering of ZFNs by the OPEN or CoDA method (Sander et al. 2010b). Zinc Finger Database (ZiFDB) has an extensive collection of all published zinc finger arrays and the updated ZiFDB2.0 version has included all possible ZFAs that can be generated by the CoDA method (Fu et al. 2009; Fu and Voytas 2013). ZFNGenome is a genome browser-based tool which helps locate potential ZFN target sites in multiple model organisms including Arabidopsis and soybean (Reyon et al. 2011).

Table 3.1 ZFN-mediated genome modifications in plants

Organism	Target	Delivery and expression system	Outcome	Reference
Arabidopsis thaliana	Transgene with a QQR ZFN site	Floral dip transformation; ZFNs induced by heat-shock	Targeted mutagenesis	Lloyd et al. (2005)
Arabidopsis thaliana	Transgene with a PTF-E2C ZFN site	Floral dip transformation; ZFNs expressed by the Rps5a, the CaMV 35S and tamoxifen-inducible promoters	Gene targeting	de Pater et al. (2009)
Arabidopsis thaliana	*ADH1* and *TT4*	Floral dip transformation; ZFNs induced by estrogen	Targeted mutagenesis	Zhang et al. (2010)
Arabidopsis thaliana	*ABI4*	Floral dip transformation; ZFNs induced by heat-shock	Targeted mutagenesis	Osakabe et al. (2010)
Arabidopsis thaliana	*MPK8*, *MPK11*, *MKK9*, *MPK15*, *MAPKKK18*, and *GA3OX2*	Floral dip transformation; ZFNs induced by estrogen	Targeted mutagenesis	Sander et al. (2010a)
Arabidopsis thaliana	*PPO*	Floral dip transformation; ZFNs expressed by the CaMV 35S promoter	Gene targeting	de Pater et al. (2012)
Arabidopsis thaliana	*At1g53430*, *At1g53440*, *At1g70450*, *At1g70460*, *At4g16960*, *At4g16940*, and *At4g16860*	Floral dip transformation; ZFNs induced by estrogen	Targeted deletion and inversion	Qi et al. (2013c)
Arabidopsis thaliana	*ADH1*	Transient transformation of protoplasts with the donor; ZFNs induced by estrogen	Gene targeting	Qi et al. (2013b)
Arabidopsis thaliana	Transgene with two QQR ZFN sites	Floral dip transformation; ZFNs expressed by the CaMV 35S promoter	Gene replacement	Weinthal et al. (2013)
Nicotiana tabacum	*Disabled GUS*: *NPTII* transgene with a Zif268 target site	Transient transformation of protoplasts; ZFNs expressed by the CaMV 35S promoter	Gene targeting	Wright et al. (2005)

Nicotiana tabacum	Endochitinase gene *CHN50*	Agrobacterium-mediated transformation of BY2 cells; ZFNs expressed by the CsVMV promoter	Gene targeting	Cai et al. (2009)
Nicotiana tabacum	*SurA* and *SurB*	Transient transformation of protoplasts; ZFNs expressed by the CaMV 35S promoter	Gene targeting	Townsend et al. (2009)
Nicotiana tabacum	Transgene flanked with multiple CCR5-ZFN sites	Agrobacterium-mediated transformation of leaf disks; ZFNs expressed by the CsVMV promoter	Targeted deletion	Petolino et al. (2010)
Nicotiana tabacum	*mGUS* transgene with a QQR ZFN site	Tobacco rattle virus (TRV)-based expression system; ZFNs expressed by the virus subgenomic promoter	Targeted mutagenesis	Marton et al. (2010)
Nicotiana tabacum	Transgene with two QQR ZFN sites	Agrobacterium-mediated transformation of leaf disks; ZFNs expressed by the CaMV 35S promoter	Gene replacement	Weinthal et al. (2013)
Zea mays	*IPK1*	Whisker-mediated transformation of embryogenic callus; ZFNs expressed by the maize ubiquitin 1 promoter	Gene targeting	Shukla et al. (2009)
Zea mays	Transgene with a “trait landing pad”	Whisker-mediated transformation of embryogenic callus; ZFNs expressed by the maize ubiquitin 1 promoter	Gene stacking	Ainley et al. (2013)
Petunia sp.	*mGUS* transgene with a QQR ZFN site	Tobacco rattle virus (TRV)-based expression system; ZFNs expressed by the virus subgenomic promoter	Targeted mutagenesis	Marton et al. (2010)
Glycine max	*GFP* transgene; *DCL2a*, *DCL1b*, *DCL4a*, *DCL4b*, *RDR6a*, *RDR6b*, and *HEN1a*	Hairy-root and whole-plant transformation; ZFNs induced by estrogen	Targeted mutagenesis	Sander et al. (2010a), Curtin et al. (2011)

Modifying the plant genomes with ZFNs is not a trivial process. It will be helpful if activities and/or toxicities of ZFNs can be tested and measured by an easy reporter system. Multiple such systems have been developed. For example, a bacterial two-hybrid reporter assay was successfully implemented for rapid assessment of ZF DNA-binding affinity in *E. coli* (Wright et al. 2006). A yeast-based assay was used for testing cleavage activities of ZFNs in vivo, which predicted *in planta* activities very well when ZFNs were expressed in Arabidopsis (Zhang et al. 2010). The yeast assay described by Zhang et al. employed a single strand annealing (SSA) reporter system (Zhang et al. 2010), which relies on HR. Because NHEJ is the dominant pathway for DNA repair in plants, assays measuring NHEJ activity may be more sensitive and efficient. To this end, a gene correction assay with a mutated β-glucuronidase (GUS) reporter was developed in plants, and the readout depended on restoration of a functional open reading frame of the reporter gene by NHEJ-based repair (Tovkach et al. 2009). Similarly, a firefly luciferase (LUC) transient expression system was established for quantifying relative activities of ZFNs and other site-specific nucleases in *Nicotiana benthamiana* (Johnson et al. 2013).

3 Targeted Mutagenesis by ZFNs

ZFNs can be excellent tools for targeted mutagenesis in plants. Such a reverse genetic application was pioneered in the model plant Arabidopsis. In a proof-of-concept study, a pre-integrated transgene was successfully targeted and mutated by a precharacterized ZFN pair (Lloyd et al. 2005). To avoid potential toxicity, a heat-inducible promoter was used for driving ZFN expression. About 10 % of induced T1 plants gave rise to mutants in the next generation, suggesting it is feasible to make mutants with ZFNs (Lloyd et al. 2005). In another study, mutations were also induced in somatic cells on a transgene by a ZFN pair under constitutive or inducible promoters, but recovery of germinal mutants was not pursued (de Pater et al. 2009). In 2010, two papers reported the successful use of ZFNs to mutate Arabidopsis endogenous genes where ZFNs were driven by different inducible promoters (Zhang et al. 2010; Osakabe et al. 2010). In one study, ZFNs were engineered by the OPEN method to target the *ADH1* and *TT4* genes. Upon induction by estrogen, ZFNs induced high frequency targeted mutagenesis and germinal transmitted mutations were recovered for both genes (Maeder et al. 2008; Zhang et al. 2010). In the other report, a modularly assembled ZFN pair made targeted mutations in the *ABI4* gene when expressed under a heat-inducible promoter (Osakabe et al. 2010). Streamlined protocols for targeted mutagenesis in Arabidopsis using ZFNs are available (Qi et al. 2013a; Zhang and Voytas 2011). The success in Arabidopsis encouraged the application of ZFNs to other plants. For example, Curtin et al. targeted multiple genes in soybean with ZFNs and recovered a heritable mutation in the *DCL4b* gene, demonstrating that soybean is also amenable to ZFN-mediated genome modification (Curtin et al. 2011).

4 Targeted Chromosomal Deletions and Inversions by ZFNs

ZFN-induced DSBs typically result in small deletions no longer than 50 bp (Qi et al. 2013b). In order to achieve large chromosomal deletions, two DSBs need to be created simultaneously at both ends. Such an idea was demonstrated in human cells where different lengths of chromosomes were deleted by two ZFN pairs (Lee et al. 2010). In plants, targeted chromosomal deletion can be a valuable strategy for removing marker genes or reducing genetic redundancy of tandemly arrayed genes (TAGs). Petolino et al. sought to demonstrate the use of ZFNs for the deletion of a transgene in tobacco (Petolino et al. 2010). In their strategy, two different transgenic tobacco lines were generated. The first line expresses a highly active ZFN pair that was previously used to target the human *CCR5* gene. The second line contains a *GUS* transgene cassette which is flanked by 4 tandemly arrayed CCR5-ZFN sites at both ends. These two transgenic lines were crossed and F2 plants with the deletion of the *GUS* gene were successfully identified and confirmed (Petolino et al. 2010). Genetic redundancy of TAGs poses a challenge for reverse genetics. In a recent study, we made multiple ZFNs for targeted deletions of TAGs at different chromosomes in Arabidopsis (Qi et al. 2013c). Since ZFNs were engineered to recognize identical sequences in multiple genes, only one pair of ZFN is needed for making certain TAG deletions. Chromosomal deletions ranging from a few kb to over 9 Mb were found in somatic cells. The estimated deletion frequencies were quite low, which hampered the recovery of germinal events. Upon two DSBs, the released chromosomal DNA could be inverted and ligated back, resulting in chromosomal inversion. Interestingly, we were able to detect such inversions in Arabidopsis. However, inversions appear to happen less frequently than deletions (Qi et al. 2013c), an observation that was also found in human cells (Lee et al. 2011).

5 Gene Replacement and Gene Stacking by ZFNs

Two decades ago, enhanced HR through induction of DSBs by a site-specific endonuclease was demonstrated in plant cells (Puchta et al. 1993). However, applying this strategy to a gene of interest cannot be realized unless site-specific nucleases are readily to be made. As ZFNs can be engineered to target virtually any gene, it is now possible to stimulate HR-based gene targeting (GT) in plants by using ZFNs. A proof-of-concept study was demonstrated in tobacco cells where ZFN-facilitated GT was achieved at a high frequency (Wright et al. 2005). In this study, repair of a defective *GUS*::*NPTII* reporter by HR enabled an efficient screen for GT events (Wright et al. 2005). In another study, a herbicide resistance gene was inserted into the endogenous *CHN50* locus in the tobacco BY2 culture cell line (Cai et al. 2009). Significant breakthrough was made when two groups independently reported the successful generation of heritable recombinant plants through ZFN-mediated GT (Townsend et al. 2009; Shukla et al. 2009). In the first study, tobacco acetolactate

synthase genes (*ALS SuRA* and *SuRB*) were modified to confer resistance to imidazolinone and sulphonylurea herbicides (Townsend et al. 2009). In the second study, targeted gene addition at the *IPK1* locus was made in maize (Shukla et al. 2009). Gene stacking is another very useful and efficient technique in crop breeding. Recently, Ainley et al. demonstrated gene stacking could be achieved by HR through targeting a "trait landing pad" with ZFNs (Ainley et al. 2013). The above-mentioned gene replacement and gene stacking were all achieved by HR. Interestingly, a recent study in Arabidopsis and tobacco demonstrated NHEJ could also be used for gene replacement when the donor sequences were flanked by the ZFN sites (Weinthal et al. 2013). Since NHEJ is the dominant DNA repair pathway in plants, such a strategy may work well for gene replacement and gene stacking.

In contrast to targeted mutagenesis, GT mediated by ZFNs has achieved limited success in Arabidopsis. For example, ZFN-mediated GT at a pre-integrated transgenic locus in Arabidopsis was pursued by de Pater et al. and ~0.1 % GT frequency was reported (de Pater et al. 2009). The same group also attempted to target the endogenous *PPO* gene with a ZFN by introducing two mutations which will result in recombinant plants resistant to the herbicide butafenacil (de Pater et al. 2012). They achieved a GT frequency of 0.31 %, which was roughly four times higher than no ZFN was applied (de Pater et al. 2012). The moderate improvement of GT by the ZFN could be due to its low activity and/or a low germinal transmission frequency.

6 Donor and/or ZFN Delivery

Both ZFNs and donors can be delivered into plant cells as transgenes, where agrobacterium-mediated transformation was most frequently used (Table 3.1). Occasionally, non-agrobacterium-mediated transformation was used for either transient expression in protoplasts or generation of stable transformants (Shukla et al. 2009; Ainley et al. 2013; Zhang et al. 2013). As ZFNs work in pairs, equal amounts of both ZFN monomers can be generated from a single messenger RNA separated by "self-cleaving" 2A peptide sequence (Doyon et al. 2008; Sander et al. 2010a; Zhang et al. 2010; Szymczak et al. 2004; Sollu et al. 2010). A Tobacco rattle virus (TRV)-based expression system was successfully used to express ZFNs for targeted mutagenesis in both tobacco and petunia (Marton et al. 2010). Presumably, the RNA virus vector DNA would not integrate into the genome and it thus represents a non-transgenic ZFN-delivery method. It will be interesting to test whether GT can be made by simultaneous delivery of donors with this viral vector system. RNA delivery of ZFNs was routinely done in mammalian cells, zebrafish eggs, and insects (Gupta et al. 2012; Cui et al. 2011; Watanabe et al. 2012; Chen et al. 2011; Beumer et al. 2008). In addition, protein delivery of ZFNs can be achieved in mammalian cells through receptor-mediated endocytosis or even without any facilitation due to the intrinsic cell-penetrating capabilities of the ZFN architecture (Chen et al. 2013; Gaj et al. 2012). Also in mammalian cells, single strand DNA oligonucleotides were successfully used as donor molecules to facilitate GT (Chen et al. 2011).

Analogous to the mammalian culture cell system, the plant protoplast system may be a good platform for testing the feasibility of using RNA or protein to deliver ZFNs and using oligonucleotides to deliver donors.

7 Cytotoxicity and Off-Targeting by ZFNs

ZFNs are often expressed in plants under inducible promoters due to worries of potential cytotoxicity (Zhang et al. 2010; Lloyd et al. 2005; Osakabe et al. 2010; Qi et al. 2013c). One plausible explanation for cytotoxicity of ZFNs is off-target cutting. Although ZFs are engineered to recognize specific DNA sequences, it is likely they bind to other homologous sequences as well. A web-based tool was developed to search for potential off-target sites by ZFNs in many non-plant genomes (Cradick et al. 2011). In plants, only limited work has been done for analyzing ZFN-induced off-target cutting. In Arabidopsis, potential off-target sites were identified by a BLAST search and they were subsequently tested for evidence of ZFN cutting (Zhang et al. 2010). In maize, a high-throughput approach was used for measuring off-targeting by ZFNs (Shukla et al. 2009). Neither study has found evidence of mutagenesis at the potential off-target sites (Zhang et al. 2010; Shukla et al. 2009). Recently, more sophisticated approaches were reported for identifying off-target sites in the human genome. Collectively, these studies showed there indeed are many potential off-target sites for ZFNs (Gabriel et al. 2011; Pattanayak et al. 2011; Sander et al. 2013). Importantly, these off-target sites can only be partially predicted in silico and have to be verified empirically. The wild-type FokI nuclease domains can form homodimers and thus also contribute to potential off-target cutting. To alleviate this problem, obligate FokI heterodimers were engineered which drastically reduced cellular toxicity with minor or no compromise in activity (Miller et al. 2007; Szczepek et al. 2007; Doyon et al. 2011). For targeted deletions involving two pairs of ZFNs, an autonomous heterodimer FokI architecture combination was proposed for use (Sollu et al. 2010). Another FokI-engineering outcome is zinc finger nickases which only break one strand of DNAs without introducing DSBs (Ramirez et al. 2012). As nicking on DNA is thought to largely trigger HR but not NHEJ repair, the nickases can facilitate HR-based GT without undesired mutagenic effects. Thus, zinc finger nickases should be quite useful for HR-based genome modification in crops where off-target mutagenesis is a major concern.

8 Somatic Versus Germinal Modifications by ZFNs

Successful plant genome modifications can only be guaranteed if such modifications will be transmitted to the next generation. In Arabidopsis, we have found that germinal transmission frequencies on targeted mutagenesis and deletions were

much lower than somatic frequencies (Qi et al. 2013c), and this phenomenon also seemed to be true for other site-specific nucleases (Christian et al. 2013). In these studies, an estrogen-inducible CaMV 35S promoter or a constitutive CaMV 35S promoter was used. One possibility is that ZFNs couldn't be expressed at a high level under the CaMV 35S promoter in germinal cells. Other promoters such as Arabidopsis ubiquitin 10 promoter (Grefen et al. 2010) and an egg-specific promoter (Yang et al. 2005; Even-Faitelson et al. 2011) may allow high or localized ZFN expression in germinal cells. The second possibility is that there are much fewer germinal cells than somatic cells in each plant. In order to get a higher chance for obtaining germinal events, it was recommended to screen progenies from multiple individual T1 transgenic plants (Qi et al. 2013a; Zhang and Voytas 2011). We have also come up a "threshold" hypothesis, where a certain high somatic mutagenesis frequency (such as 7 %) needs to be reached in order to obtain germinal events (Zhang et al. 2010; Qi et al. 2013c). The third possibility is that DNA repair mechanism may be different in plant stem cells when compared to somatic cells. It has been reported that the cell niches in Arabidopsis are hypersensitive to DNA damage (Fulcher and Sablowski 2009). Interestingly, the study also found the hypersensitivity of stem cells to DNA damage could be suppressed when the key DNA damage response gene *ATM* was knocked out. In the future, it will be worth testing whether higher germinal transmission frequency of genome modifications can be achieved in the *atm* mutant background. The challenge of low germinal transmission frequency of genome modification may be pertained in other plants as well. If it is true, an alternative way to overcome this challenge is to regenerate whole plants from somatic tissues as far as the plants are amenable to cell culture and regeneration.

9 Genetic Approaches to Facilitate High Frequency Genome Modifications

The outcomes of genome modifications rely on DNA repair pathways. By manipulating DNA repair pathways, one can potentially enhance the frequencies of ZFN-mediated genome modifications. Such a strategy was demonstrated in Arabidopsis with ZFNs that target endogenous genes (Qi et al. 2013b). When the sister-chromatid-based HR pathway gene *SMC6B* was knocked out, the DNA repair pathway shifted to NHEJ and resulted in higher frequency of targeted mutagenesis. When the classic NHEJ pathway genes *KU70* or *LIG4* was mutated, microhomology-based alternative NHEJ and HR were both promoted, leading to higher frequencies of larger deletions and gene targeting (Qi et al. 2013b). Enhanced gene targeting was also reported by overexpressing the yeast HR gene *RAD54* (Even-Faitelson et al. 2011; Shaked et al. 2005). Thus, genome modifications can be improved by manipulating DNA repair pathways, which may indirectly contribute to the recovery of germinal transmitted events.

10 Future Perspective

Besides ZFN, three other site-specific nuclease systems are available: meganuclease (Smith et al. 2006; Paques and Duchateau 2007), TALEN (transcription activator-like Effector nuclease) (Christian et al. 2010; Miller et al. 2010; Li et al. 2010; Bogdanove and Voytas 2011), and CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/Cas9) (Cong et al. 2013; Mali et al. 2013; Jinek et al. 2012). The emerging TALEN and CRISPR/Cas9 systems are both quite promising due to their flexibility on target site selection and easiness of engineering. TALENs have been used in genome engineering in Arabidopsis (Christian et al. 2013; Cermak et al. 2011), tobacco (Zhang et al. 2013; Mahfouz et al. 2011), rice (Li et al. 2012; Shan et al. 2013a), brachypodium (Shan et al. 2013a), and wheat (Wendt et al. 2013). The CRISPR/Cas9 system has also worked in the cells of many plants including Arabidopsis, tobacco, sorghum, and rice (Miao et al. 2013; Jiang et al. 2013; Mao et al. 2013; Feng et al. 2013; Xie and Yang 2013; Shan et al. 2013b; Li et al. 2013; Nekrasov et al. 2013). However, the issues faced by ZFNs such as germinal transmission, off-target cutting, and chromatin accessibility will likely all apply to other nucleases as well. Compared to other nuclease systems, the ZFN technology has its own advantages. For example, direct protein delivery of ZFNs into cells was possible (Gaj et al. 2012), which is unlikely to work for TALENs and Cas9 proteins due to their larger sizes and protein characteristics. Unlike ZFNs, both the TALEN and CRISPR/Cas9 systems utilized proteins of bacterial origins and this feature may render genetically modified organisms (GMOs) generated by either system to face a tougher regulation. The ZFN technology has already helped achieve a variety of genome modifications in plants (Fig. 3.2 and Table 3.1). It will likely continue to play a significant role in plant genome engineering in the future.

References

Ainley WM, Sastry-Dent L, Welter ME et al (2013) Trait stacking via targeted genome editing. Plant Biotechnol J 11(9):1126–1134

Bae KH, Kwon YD, Shin HC et al (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. Nat Biotechnol 21(3):275–280

Beerli RR, Barbas CF 3rd (2002) Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol 20(2):135–141

Beumer KJ, Trautman JK, Bozas A et al (2008) Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci U S A 105(50): 19821–19826

Bhakta MS, Henry IM, Ousterout DG et al (2013) Highly active zinc-finger nucleases by extended modular assembly. Genome Res 23(3):530–538

Bibikova M, Carroll D, Segal DJ et al (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol 21(1):289–297

Bogdanove AJ, Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. Science 333(6051):1843–1846

Cai CQ, Doyon Y, Ainley WM et al (2009) Targeted transgene integration in plant cells using designed zinc finger nucleases. Plant Mol Biol 69(6):699–709
Carroll D (2011) Genome engineering with zinc-finger nucleases. Genetics 188(4):773–782
Cermak T, Doyle EL, Christian M et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39(12):e82
Chen F, Pruett-Miller SM, Huang Y et al (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods 8(9):753–755
Chen Z, Jaafar L, Agyekum DG et al (2013) Receptor-mediated delivery of engineered nucleases for genome modification. Nucleic Acids Res 41(19):e182
Christian M, Cermak T, Doyle EL et al (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186(2):757–761
Christian M, Qi Y, Zhang Y, Voytas DF (2013) Targeted mutagenesis of Arabidopsis thaliana using engineered TAL effector nucleases (TALENs). G3 (Bethesda) 3(10):1697–1705
Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823
Cornu TI, Thibodeau-Beganny S, Guhl E et al (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. Mol Ther 16(2):352–358
Cradick TJ, Ambrosini G, Iseli C, Bucher P, McCaffrey AP (2011) ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites. BMC Bioinformatics 12:152
Cui X, Ji D, Fisher DA, Wu Y, Briner DM, Weinstein EJ (2011) Targeted integration in rat and mouse embryos with zinc-finger nucleases. Nat Biotechnol 29(1):64–67
Curtin SJ, Zhang F, Sander JD et al (2011) Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiol 156(2):466–473
de Pater S, Neuteboom LW, Pinas JE, Hooykaas PJ, van der Zaal BJ (2009) ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. Plant Biotechnol J 7(8):821–835
de Pater S, Pinas JE, Hooykaas PJ, van der Zaal BJ (2012) ZFN-mediated gene targeting of the Arabidopsis protoporphyrinogen oxidase gene through Agrobacterium-mediated floral dip transformation. Plant Biotechnol J 11(4):510–515
Doyon Y, McCammon JM, Miller JC et al (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol 26(6):702–708
Doyon Y, Vo TD, Mendel MC et al (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods 8(1):74–79
Even-Faitelson L, Samach A, Melamed-Bessudo C, Avivi-Ragolsky N, Levy AA (2011) Localized egg-cell expression of effector proteins for targeted modification of the Arabidopsis genome. Plant J 68(5):929–937
Feng Z, Zhang B, Ding W et al (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell Res 23(10):1229–1232
Fu F, Voytas DF (2013) Zinc Finger Database (ZiFDB) v2.0: a comprehensive database of C2H2 zinc fingers and engineered zinc finger arrays. Nucleic Acids Res 41(Database issue):D452–D455
Fu F, Sander JD, Maeder M et al (2009) Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. Nucleic Acids Res 37(Database issue):D279–D283
Fulcher N, Sablowski R (2009) Hypersensitivity to DNA damage in plant stem cell niches. Proc Natl Acad Sci U S A 106(49):20984–20988
Gabriel R, Lombardo A, Arens A et al (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol 29(9):816–823
Gaj T, Guo J, Kato Y, Sirk SJ, Barbas CF 3rd (2012) Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. Nat Methods 9(8):805–807
Grefen C, Donald N, Hashimoto K, Kudla J, Schumacher K, Blatt MR (2010) A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. Plant J 64(2):355–365

Gupta A, Christensen RG, Rayla AL, Lakshmanan A, Stormo GD, Wolfe SA (2012) An optimized two-finger archive for ZFN-mediated gene targeting. Nat Methods 9(6):588–590

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41(20):e188

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096): 816–821

Johnson RA, Gurevich V, Levy AA (2013) A rapid assay to quantify the cleavage efficiency of custom-designed nucleases in planta. Plant Mol Biol 82(3):207–221

Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A 93(3):1156–1160

Kim S, Lee MJ, Kim H, Kang M, Kim JS (2011) Preassembled zinc-finger arrays for rapid construction of ZFNs. Nat Methods 8(1):7

Lee HJ, Kim E, Kim JS (2010) Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res 20(1):81–89

Lee HJ, Kweon J, Kim E, Kim S, Kim JS (2011) Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. Genome Res 22(3):539–548

Li T, Huang S, Jiang WZ et al (2010) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res 39(1):359–372

Li T, Liu B, Spalding MH, Weeks DP, Yang B (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol 30(5):390–392

Li JF, Norville JE, Aach J et al (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol 31(8):688–691

Liu Q, Xia Z, Zhong X, Case CC (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. J Biol Chem 277(6):3850–3856

Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci U S A 102(6):2232–2237

Maeder ML, Thibodeau-Beganny S, Osiak A et al (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31(2):294–301

Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu JK (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A 108(6):2623–2628

Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. Science 339(6121):823–826

Mandell JG, Barbas CF III (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* 34(Web Server issue):W516–523

Mao Y, Zhang H, Xu N, Zhang B, Gao F, Zhu JK (2013) Application of the CRISPR-Cas System for Efficient Genome Engineering in Plants. Mol Plant 6(6):2008–2011

Marton I, Zuker A, Shklarman E et al (2010) Nontransgenic genome modification in plant cells. Plant Physiol 154(3):1079–1087

Miao J, Guo D, Zhang J et al (2013) Targeted mutagenesis in rice using CRISPR-Cas system. Cell Res 23(10):1233–1236

Miller JC, Holmes MC, Wang J et al (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25(7):778–785

Miller JC, Tan S, Qiao G et al (2010) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29(2):143–148

Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol 31(8):691–693

Osakabe K, Osakabe Y, Toki S (2010) Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. Proc Natl Acad Sci U S A 107(26):12034–12039

Paques F, Duchateau P (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr Gene Ther 7(1):49–66

Pattanayak V, Ramirez CL, Joung JK, Liu DR (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat Methods 8(9):765–770

Petolino JF, Worden A, Curlee K et al (2010) Zinc finger nuclease-mediated transgene deletion. Plant Mol Biol 73(6):617–628

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56(409):1–14

Puchta H, Dujon B, 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(22):5034–5040

Qi Y, Starker CG, Zhang F, Baltes NJ, Voytas DF (2013a) Tailor-made mutations in Arabidopsis using zinc finger nucleases. Methods Mol Biol 1062:193–209

Qi Y, Zhang Y, Zhang F et al (2013b) Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. Genome Res 23(3):547–554

Qi Y, Li X, Zhang Y et al (2013c) Targeted deletion and inversion of tandemly arrayed genes in Arabidopsis thaliana using zinc finger nucleases. G3 (Bethesda) 3(10):1701–1715

Ramirez CL, Foley JE, Wright DA et al (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods 5(5):374–375

Ramirez CL, Certo MT, Mussolino C et al (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. Nucleic Acids Res 40(12): 5560–5568

Reyon D, Kirkpatrick JR, Sander JD et al (2011) ZFNGenome: a comprehensive resource for locating zinc finger nuclease target sites in model organisms. BMC Genomics 12:83

Sander JD, Dahlborg EJ, Goodwin MJ et al (2010a) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8(1):67–69

Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res* 38(Web Server issue):W462–468

Sander JD, Ramirez CL, Linder SJ et al (2013) In silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. Nucleic Acids Res 41(19):e181

Segal DJ, Beerli RR, Blancafort P et al (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. Biochemistry 42(7):2137–2148

Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. Proc Natl Acad Sci U S A 102(34):12265–12269

Shan Q, Wang Y, Chen K et al (2013a) Rapid and efficient gene modification in rice and brachypodium using TALENs. Mol Plant 6(4):1365–1368

Shan Q, Wang Y, Li J et al (2013b) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31(8):686–688

Shukla VK, Doyon Y, Miller JC et al (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature 459(7245):437–441

Smith J, Grizot S, Arnould S et al (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res 34(22):e149

Sollu C, Pars K, Cornu TI et al (2010) Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. Nucleic Acids Res 38(22):8269–8276

Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol 25(7):786–793

Szymczak AL, Workman CJ, Wang Y et al (2004) Correction of multi-gene deficiency in vivo using a single "self-cleaving" 2A peptide-based retroviral vector. Nat Biotechnol 22(5):589–594

Tovkach A, Zeevi V, Tzfira T (2009) A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells. Plant J 57(4):747–757

Townsend JA, Wright DA, Winfrey RJ et al (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459(7245):442–445

Tzfira T, Weinthal D, Marton I, Zeevi V, Zuker A, Vainstein A (2012) Genome modifications in plant cells by custom-made restriction enzymes. Plant Biotechnol J 10(4):373–389

Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. Annu Rev Plant Biol 64:327–350

Watanabe T, Ochiai H, Sakuma T et al (2012) Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. Nat Commun 3:1017

Weinthal D, Tovkach A, Zeevi V, Tzfira T (2010) Genome editing in plant cells by zinc finger nucleases. Trends Plant Sci 15(6):308–321

Weinthal DM, Taylor RA, Tzfira T (2013) Nonhomologous end joining-mediated gene replacement in plant cells. Plant Physiol 162(1):390–400

Wendt T, Holm PB, Starker CG et al (2013) TAL effector nucleases induce mutations at a preselected location in the genome of primary barley transformants. Plant Mol Biol 83(3):279–285

Wright DA, Townsend JA, Winfrey RJ Jr et al (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J 44(4):693–705

Wright DA, Thibodeau-Beganny S, Sander JD et al (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nat Protoc 1(3):1637–1652

Xie K, Yang Y (2013) RNA-guided Genome Editing in Plants Using A CRISPR-Cas System. Mol Plant 6(6):1975–1983

Yang W, Jefferson RA, Huttner E, Moore JM, Gagliano WB, Grossniklaus U (2005) An egg apparatus-specific enhancer of Arabidopsis, identified by enhancer detection. Plant Physiol 139(3):1421–1432

Zhang F, Voytas DF (2011) Targeted mutagenesis in Arabidopsis using zinc-finger nucleases. Methods Mol Biol 701:167–177

Zhang F, Maeder ML, Unger-Wallace E et al (2010) High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proc Natl Acad Sci U S A 107(26): 12028–12033

Zhang Y, Zhang F, Li X et al (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol 161(1):20–27

Chapter 4
Engineered TAL Effector Proteins: Versatile Reagents for Manipulating Plant Genomes

Michelle Christian and Daniel F. Voytas

Abstract Transcription activator-like (TAL) effectors are proteins produced by plant pathogens of the genus *Xanthomonas*. They are delivered to plant cells during infection and bind to specific plant gene promoters to activate transcription and promote bacterial infection. DNA binding by TAL effectors is mediated by an array of typically 14–24 repeats; each repeat is 34 amino acids in length and folds into a hairpin-like structure that contacts a single base in the target DNA. The TAL effector DNA-binding motif has proven highly modular, and custom TAL effector arrays can be made to recognize virtually any site in a plant genome, thereby providing a valuable reagent for genome manipulation. In particular, when TAL effector arrays are fused to a nuclease, they can create targeted double-strand breaks at a locus of interest. The repair of the breaks can be directed to achieve a variety of targeted genome modifications, with applications ranging from understanding plant gene function to creating novel traits in agronomically important crop species.

Transcription activator-like (TAL) effectors are a class of DNA-binding proteins produced by plant pathogenic bacteria of the genus *Xanthomonas*. Pathovars of *Xanthomonas* infect a wide variety of plants, including rice, citrus, pepper, cotton, and soybean, and their TAL effectors are critical for pathogenesis (Mansfield et al. 2012). Upon infecting plant tissue, TAL effectors are secreted into cells via the type III secretion pathway. Once inside the cell, they move to the nucleus where they bind to promoters and transcriptionally activate host susceptibility (S) genes, such as *UPA20*, *Os8N3*, *OsTFIIAg1*, or *OsTFX1*, resulting in disease (Fig. 4.1) (Bogdanove et al. 2010). Expression of target genes promotes infection, for example, by producing sucrose transporters that feed the colonizing bacteria (Chen et al. 2010). Over time, plants have evolved mechanisms to resist infection: (1) S gene activation

M. Christian, Ph.D.
Department of Genetics, Cell Biology and Development, University of Minnesota, Saint Paul, MN, USA

D.F. Voytas, Ph.D. (✉)
Department of Genetics, Cell Biology and Development, Center for Genome Engineering, University of Minnesota, 328 Cargill Building, 1500 Gortner Avenue, Saint Paul, MN, USA
e-mail: voytas@umn.edu

F. Zhang et al. (eds.), *Advances in New Technology for Targeted Modification of Plant Genomes*, DOI 10.1007/978-1-4939-2556-8_4

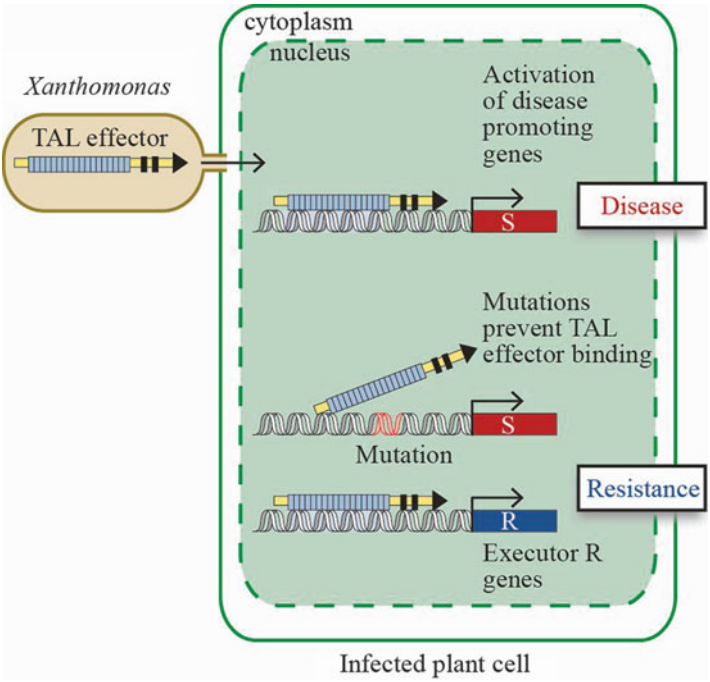

Fig. 4.1 Transcription activator-like (TAL) effectors. TAL effectors are proteins made by *Xanthomonas*, a genus of plant pathogenic bacteria. Upon infection, the bacteria inject TAL effectors into host plant cells via the type III secretion system; TAL effectors are then transported to the nucleus. Once in the nucleus, they bind the promoters of susceptibility (*S*) genes (*red*) that help facilitate colonization and promote disease. Mutations in plant gene promoters that affect DNA binding prevent disease progression. Plants have also evolved mechanisms for resistance, such as use of executor *R* genes (*blue*) that are under transcriptional control of TAL effector target sequences, which would normally be present only in *S* gene promoters. Figure adapted from (Bogdanove et al. 2010)

can be reduced by mutations in individual promoter target sites or (2) activation traps have evolved in which executor resistance (R) genes are under transcriptional control of TAL effectors and have target sequences in their promoters that would typically be present in S gene promoters (Fig. 4.1) (Bogdanove et al. 2010).

The first TAL effector discovered was AvrBs3 from *Xanthomonas campestris pv. Vesicatoria*, a pepper pathogen (Bonas et al. 1989). This TAL effector triggers an immune response only in plants carrying the disease resistance *Bs3* gene. Two studies addressed how AvrBs3 elicits either resistance or susceptibility phenotypes in related pepper varieties and led to the discovery of a conserved nucleotide sequence that is required for host gene activation by TAL effectors, subsequently named the "up-regulated by AvrBs3" (UPA) or "up-regulated by TAL effector" (UPT) box (Kay et al. 2007; Romer et al. 2007). These studies showed for the first time that sequence-specific DNA binding is required for TAL effector function. Thus, the hunt began to identify the mechanism by which TAL effector proteins recognize and bind their promoter targets.

Several features characterize TAL effectors: they have an N-terminus required for type III secretion and a C-terminus that contains several nuclear localization signals (NLS) and an acidic activation domain (AAD), which is common among transcription factors (Fig. 4.2a) (Boch and Bonas 2010). TAL effectors also contain

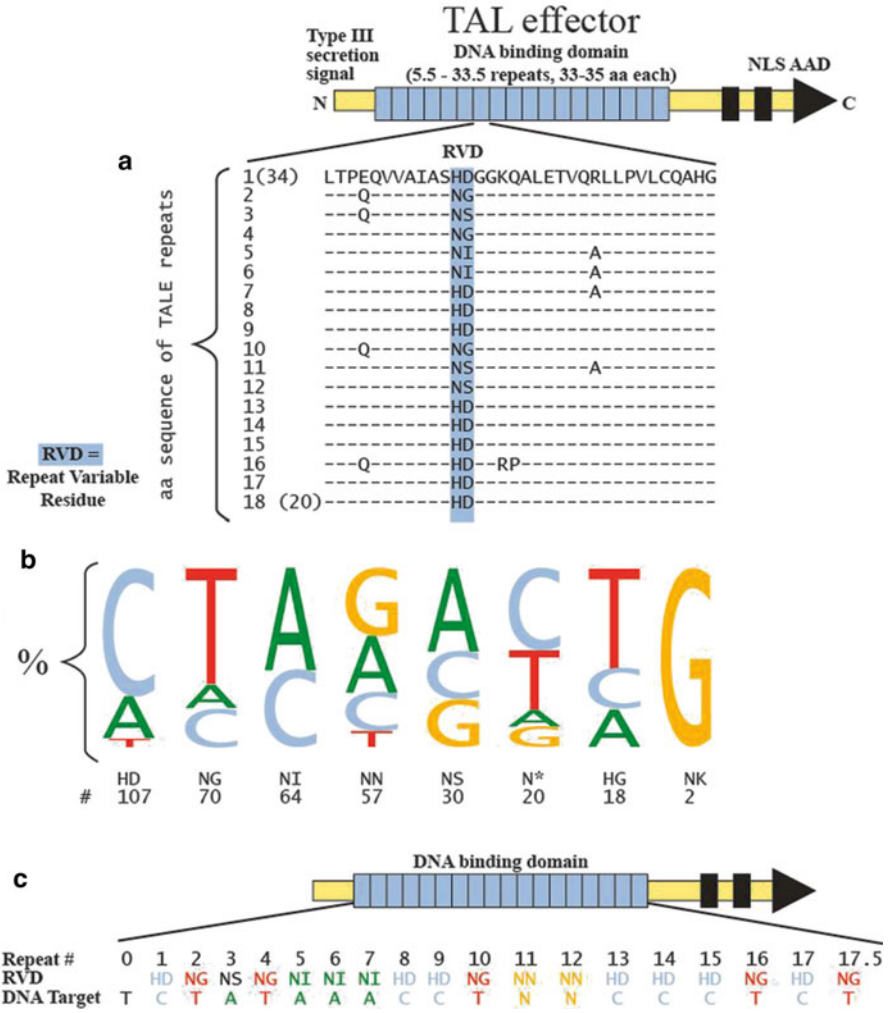

Fig. 4.2 Schematic of a TAL effector and the DNA recognition code. (**a**) Schematic of a TAL effector protein. The central region of repeats (*blue squares*) specifies DNA binding. The nuclear localization signals (NLS) and acidic activation domains (AAD) are also depicted. The aa sequences for individual repeats of a representative array are shown; repeat variable di-residues (RVDs) are highlighted in *blue* and *dashes* indicate conserved aa residues. (**b**) RVD-nucleotide association frequencies. The size of the *letter* in the sequence logo reflects the frequency with which RVDs are associated with particular bases (Moscou and Bogdanove 2009). (**c**) Schematic of a TAL effector repeat array depicting repeat position, RVD identity and the nucleotide target. A conserved "T" is shown at the "0" repeat position

a central region that is responsible for binding DNA, namely an array of near-perfect repeats. Each repeat consists of 34 aa (though variants exist), and nearly all repeat arrays terminate with a final, truncated repeat containing the first 20 aa of a canonical repeat, referred to as a half repeat. Additionally, 33 aa repeats contain a deletion of the residue in the 13th position. The number of repeats within a TAL effector array varies from protein to protein, ranging from 1.5 to 33.5, with an average of 17.5. There also exist sequence polymorphisms across TAL effector proteins, both in the N- and C-termini and within the repeat array. The most pronounced variability exists at residues 12 and 13 of the repeat, referred to as the repeat-variable diresidues (RVDs). Among naturally occurring TAL effectors, over 20 distinct RVDs have been observed; however, HD, NG, NI, and NN account for over 75 % of the total (Moscou and Bogdanove 2009).

Two landmark papers appeared in 2009 that described the basis for sequence-specific DNA recognition by TAL effectors (Moscou and Bogdanove 2009; Boch et al. 2009). Each repeat in the TAL effector DNA-binding domain recognizes one base in duplex DNA, and both studies derived a DNA-binding "code" for nucleotide recognition by the TAL effector repeats. The association frequencies for the most common or canonical RVDs in the repeats are as follows: HD binds C most frequently; NG associates with T; NI with A; and NN recognizes both G and A (Fig. 4.2b). Additional common RVDs include NS, which recognizes A, C, and G, whereas HG mimics NG in binding to T. N* (missing residue 13 in a 33 aa repeat) is the most promiscuous RVD, having been found to associate with all four nucleotides at some frequency. A key feature of nearly all sequences bound by TAL effectors is a conserved T at the −1 position (Fig. 4.2c). In functional studies of target gene activation using TAL effectors, permutation of the conserved T resulted in loss of gene up-regulation (Boch et al. 2009). Taken together, the one-to-one correspondence between TAL effector repeats and single DNA bases constitutes a simple code that can be used to predict TAL effector DNA-binding sites and to design TAL effector arrays that recognize unique sequences in any genome.

Recent work has provided insight into the structural basis for DNA recognition by the TAL effector repeat. One of the first structures solved was an artificial TAL effector, dHAX3, which has 11 repeats of 34 aa plus the half repeat, representing three common RVDs (HD, NG and NS) (Deng et al. 2012). The structure was solved to high resolution in both the unbound state (1.85 Å) and in complex with DNA (2.4 Å). When unbound, the dHAX3 TAL effector has an extended conformation that is nearly twice the size of its bound conformation. Concomitantly, the structure of the naturally occurring TAL effector, PthXo1 from the rice pathogen *Xanthomonas oryzae*, was solved (Mak et al. 2012). Although the resolution was lower (3 Å), PthXo1 has 23.5 repeats that bind two full turns of DNA and has six unique RVDs (HD, NN, NG, HG, NI, and N*). In addition, the structure of PthXo1 was solved bound to its natural DNA target. Both structural studies revealed that the amino acids comprising each TAL repeat form a left-handed, two-helix bundle connected by a loop with residues 12 and 13 at the apex. When situated in an array, the self-associating repeats form a right-handed superhelix that wraps around the major groove of DNA (Fig. 4.3a, b).

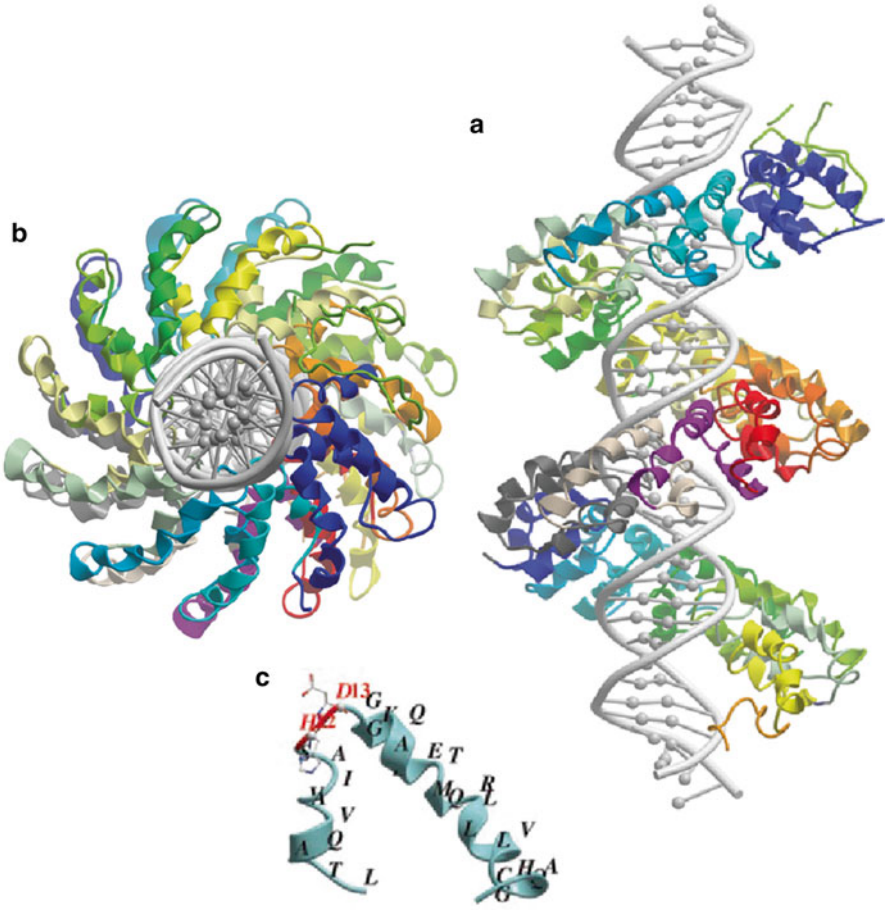

Fig. 4.3 Crystal structure of a TAL effector bound to DNA. The solved crystal structure of the PthXo1 TAL effector DNA-binding region in complex with its target site viewed both from the side (**a**) and top (**b**) (Mak et al. 2012). PthXo1 has 23.5 canonical repeats. Colors of individual TAL repeats differ. (**c**) The structure of a representative repeat. The "HD" RVD residues (aa positions 12 and 13) that recognize a cytosine are highlighted in *red*. All images were rendered using MolSoft ICM browser software, 2011

Sequence-specific contacts occur between each TAL effector repeat and the major groove of DNA (Deng et al. 2012; Mak et al. 2012). The second residue in the RVD loop (position 13) determines nucleotide specificity by making contacts with each base along a single, contiguous strand of the DNA target (Fig. 4.3c). The first residue in the loop (position 12) likely plays a stabilizing role in establishing a "pre-bound" conformation by forming a hydrogen bond between the side chain (of residue 12) and the backbone of the first helix in each repeat. The residue in position 8 of the same repeat also interacts with residue 12, thereby providing stability within the individual repeat units. The glutamine (Gln) at position 17 is the only repeat

residue that directly contacts the phosphate backbone. The structure of the PthXo1 TAL effector bound to its DNA target revealed two "cryptic" repeats (termed the −1 and 0 repeats) in the N-terminal region just upstream of the canonical repeats (Mak et al. 2012). These cryptic repeats each form a helix and loop similar to the helix and RVD loop of the canonical repeats. A tryptophan residue at the apex of the −1 repeat forms a nonpolar contact with the conserved thymidine (T) nucleotide at position 0, which is present in nearly all TAL effector DNA target sites.

Both structural predictions and the solved crystal structures provide an explanation for observations made in previous functional and molecular studies regarding TAL effector/DNA interactions (Mak et al. 2012; Bradley 2012). Specificity-determining interactions can be classified by the type of contact made between RVD loop residues and the DNA base: first, strong hydrogen bonds provide recognition for RVDs such as HD to cytosine or NN to a purine (A or G); second, in the absence of hydrogen bonds, weaker van der Waals contacts and efficient packing between the backbone of a glycine and a thymidine (T) nucleotide provide the basis of NG and HG contacts; third, relaxed specificity due to the steric exclusion of alternate bases serves to explain the preference of NI for A nucleotides; and finally, N* RVDs, which lack a residue at position 13 and thus lack an appropriate side chain, are shown to accommodate any base, with little to no impact on overall affinity of the array (Mak et al. 2012).

1 Genome Editing Using TAL Effector Nuclease Technology

The development and implementation of zinc finger nucleases (ZFNs) set precedence for the general application of DNA-binding domains fused to nonspecific Type II restriction enzymes like FokI (Kim et al. 1996). However, prior to their testing, TAL effector–nuclease fusion proteins were not expected to function the same as ZFNs; data suggested that TAL effector proteins form dimers in the cytosol of host cells before being transported into the nucleus (Gurlebeck et al. 2005, 2006; Kay et al. 2005, 2009). Dimerization would have posed problems in engineering a TAL effector nuclease using an approach that was successful for ZFNs. The first demonstration of TAL effector arrays fused to a FokI nuclease indicated that the proteins were indeed able to cleave their predicted targets with efficiencies comparable to ZFNs (Christian et al. 2010). These TAL effector nucleases (TALENs) were designed to function as dimers, with each monomer composed of TAL effector DNA recognition repeats fused to the catalytic domain of FokI (Fig. 4.4a). TALEN monomers were engineered to bind to one of two DNA half-sites separated by a spacer sequence. This architecture allows the FokI monomers to dimerize and create a double-strand DNA break (DSB) in the spacer sequence between the half-sites.

Initial work to explore the potential of the TAL effector DNA recognition domain for creating targeted nucleases assessed whether native TAL effectors function as nucleases when fused with FokI (Christian et al. 2010). A yeast-based assay was

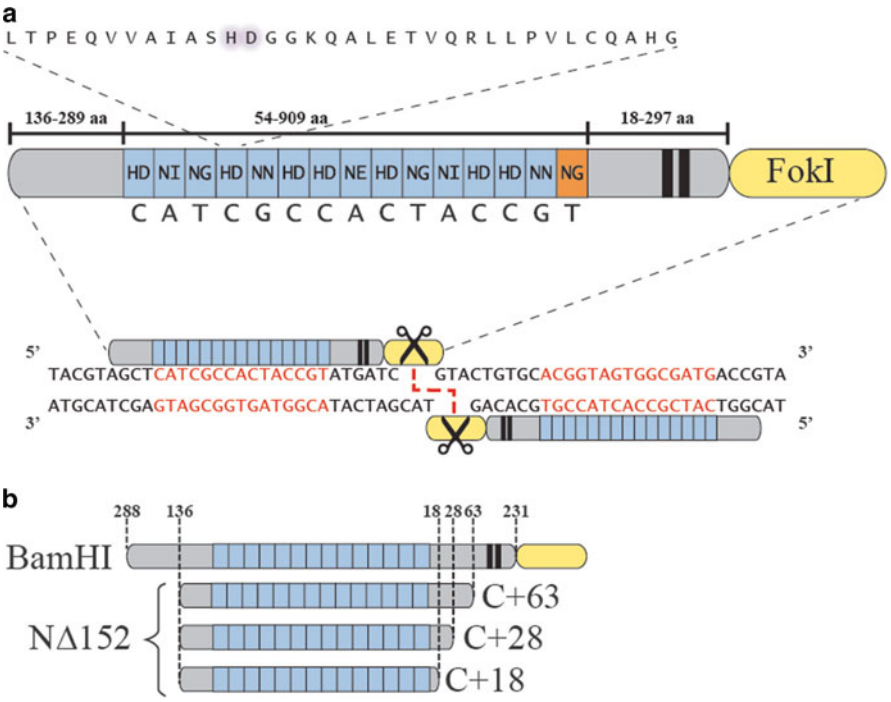

Fig. 4.4 Schematic of a TAL effector nuclease (TALEN) and various TALEN architectures. (**a**) In the expanded view of the TALEN, the lengths of various domains are indicated, along with a representative TAL effector array and its cognate nucleotide target. The orientation of a full TALEN pair is shown with the TALEN target sequence highlighted in *red*. (**b**) Truncated TALEN backbone architectures. A series of truncations were made to the N- and C-termini flanking the DNA-binding domain. N152 lacks aa up to residue 152 in the N-terminus, whereas truncations made to the C-terminus are indicated by the number of aa remaining (e.g., the C63 truncation has 63 aa after the last repeat)

carried out using a TALEN expression construct and a target reporter. The target reporter has a disrupted *lacZ* gene with a 125 bp duplication of coding sequence flanking the TALEN target sequence (composed of two half sites and a spacer sequence). If the TALENs bound and generated a DNA double-strand break at the target site, the break would be repaired by recombination between the duplicated coding sequences, thus reconstituting a functional *lacZ* gene. TALENs derived from two native TAL effectors, AvrBs3 and PthXo1, were found to generate LacZ activity significantly higher (15.8–30-fold higher) than the control yeast cells that contained only the targets. These data demonstrated that TALENs function as site-specific nucleases in vivo to cleave intended target sequences.

A second study demonstrating the development of TAL effector nucleases (termed TALNs) tested the fusion of the FokI nuclease domain to both the N- and C-termini of the full length AvrXa7 TAL effector (Fig. 4.4b) (Li et al. 2010). Fusions of FokI to the N-terminus were less active than those made to the C-terminus

(approximately 14-fold versus a 290-fold change over the negative control, respectively). Therefore, the currently adopted architecture of TALENs uses C-terminal fusions to FokI. Shortly thereafter, a third study was published that tested TALENs derived from the Hax3 effector from a mustard pathogen (Mahfouz et al. 2011). This work verified TALEN activity in vivo using a transient cleavage assay in tobacco leaves. Taken together, the initial publications on TALEN technology made several important conclusions: first, TAL effector–nuclease fusions retain the ability to bind DNA; second, TALENs cleave their predicted DNA targets both in vitro and in vivo with activities comparable to ZFNs; third, TAL effector arrays can be engineered to recognize novel, predicted DNA targets; and lastly, unlike ZFNs, TALENs retain cleavage activity over a range of target spacer lengths. Importantly, the modular nature of TAL effector repeat overcame a significant limitation in DNA targeting, and TAL effector-based reagents quickly proved to be very robust for targeting any sequence in a genome. A significant bottleneck in initial studies, however, was the difficulty to rapidly engineer novel TAL effector arrays to recognize a desired sequence. The response by researchers to overcome this limitation was remarkably swift, and multiple TAL effector engineering methods were quickly made available.

2 Methods to Engineer Novel TAL Effector Arrays

As a testament to the almost immediate interest in applications of TAL effector-based technology, a number of both publicly and commercially available platforms quickly became available for engineering novel TAL effector arrays. One of the first and most widely adopted assembly method is based on a highly efficient cloning system, termed Golden Gate, which allows for the ordered assembly of multiple DNA fragments in a single reaction (Engler et al. 2008, 2009). Golden Gate cloning capitalizes on properties of certain Type IIS restriction endonucleases, which cleave outside their recognition sequence to create unique 4 bp overhangs. The sequence of these "sticky ends" are such that after cycles of digestion and ligation, only fragments with corresponding overhangs will come together in the desired order (Fig. 4.5). Golden Gate-based methods avoid the need for PCR-based assembly, which can be error prone, particularly when amplifying the highly repetitive TAL effector arrays. Further, no gel purification steps are required, which are tedious and time consuming. Other groups have devised methods to scale up the production of engineered TAL effectors and are thus considered to be "high-throughput" (Reyon et al. 2012; Briggs et al. 2012; Schmid-Burgk et al. 2013). Finally, a number of companies have begun to offer TAL effector design and engineering services for a fee. These services, though based on proprietary methods, are becoming less costly and offer an alternative for academic researchers who wish to use the technology without implementing "in-house" assembly methods.

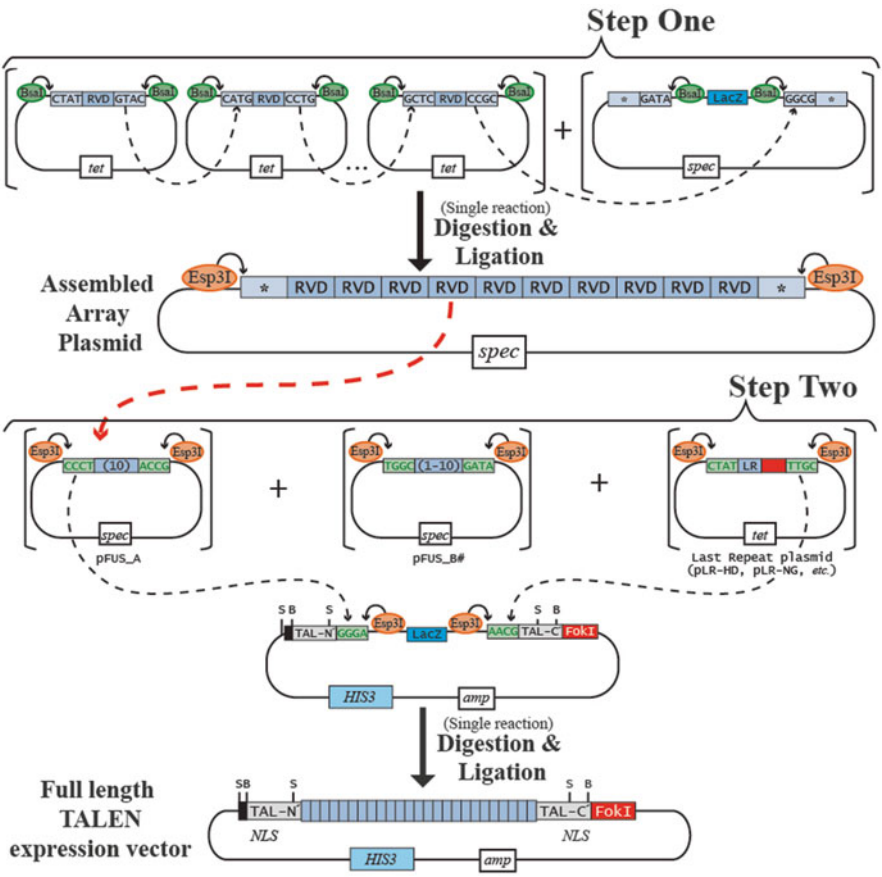

Fig. 4.5 Golden gate cloning method for assembling TAL effector arrays (Cermak et al. 2011). Type IIS restriction endonucleases (e.g., BsaI and Esp3I) allow users to rapidly clone modules encoding the desired RVDs (*light blue boxes*). RVDs are released with unique cohesive ends for ordered, single-reaction assembly into plasmids in a first step. The resulting arrays are subsequently released and assembled in sequential order in a second step into a backbone plasmid to create full length constructs with custom repeat arrays. *AD* transcriptional activation domain, *tet* tetracycline resistance, *spec* spectinomycin resistance, *amp* ampicillin resistance, *HIS3* histidine biosynthesis marker gene, *NLS* nuclear localization signal, *B* BamHI, *S* SphI

3 Modifying Genes Using TAL Effector Fusion Proteins

With the assembly of novel TAL effector arrays no longer a bottleneck, the application of the technology to genome engineering accelerated rapidly. In the two short years following the publication of the first proof-of-principle studies, over 80 published examples of TAL effector-mediated genome modification appeared, spanning 17 different model organisms. Many studies reported the use of TALEN

pairs to mediate the formation of small insertions and deletions in a gene of interest, thereby altering or inactivating the gene product. More complex chromosomal modifications can be generated when two TALENs pairs are used to create DSBs simultaneously, leading to the formation of large deletions and inversions (Ansai et al. 2013; Carlson et al. 2012; Ma et al. 2012). One landmark study demonstrated that TALENs could efficiently modify human genes, namely *NTF3* and *CCR5* (Miller et al. 2011). This study also made a critical advance by defining an optimal TALEN architecture for binding and cleavage, specifically preferred N- and C-terminal domains flanking the DNA-binding domain that give optimal TALEN activity (see also below).

Nuclease-mediated gene editing is advantageous for modifying the genomes of organisms that have longer life cycles and present DNA delivery challenges, such as large animals and agriculturally important crops. TALENs have been used to successfully knockout the low-density lipoprotein (LDL) receptor in miniature pigs to create animals that can be used to study cholesterol-related diseases (Carlson et al. 2012). TALENs have been used to edit genes in several plant species, with the first demonstration occurring in tobacco and Arabidopsis protoplasts (Mahfouz et al. 2011; Cermak et al. 2011; Zhang et al. 2013). TALENs have been successfully used to modify the genomes of whole plants, including barley, rice, and Brachypodium (Li et al. 2012a; Shan et al. 2013; Wendt et al. 2013). For example, a TALEN pair was used to disrupt the natural target of TAL effectors recognized by a strain of a rice *Xanthomonas* pathogen in order to render rice plants resistant to infection (Li et al. 2012a). Taken together, these studies demonstrate the potential that TALENs offer for modifying crop and livestock genomes to improve plant and animal agriculture.

Success in fusing the TAL effector DNA-binding domain to an endonuclease opened the door for additional applications in which activity of other functional domains were directed to a predetermined sequence. There have been several reports of successfully altering endogenous gene expression using TAL effector arrays fused to transcriptional activator, repressor, or chromatin-modifying domains. The AAD present in full-length TAL effector proteins has been harnessed to activate expression of endogenous plant genes (Boch et al. 2009; Morbitzer et al. 2010; Romer et al. 2009a; Scholze and Boch 2010). More recently, engineered TAL effector arrays have been fused to well-characterized activation domains (e.g., VP16 and VP64) or repressors (e.g., SID and KRAB) to create artificial transcription factors that can induce changes in gene expression in the range of twofold to 30-fold (Miller et al. 2011; Cermak et al. 2011; Bultmann et al. 2012; Cong et al. 2012; Geissler et al. 2011; Garg et al. 2012; Li et al. 2012b; Maeder et al. 2013; Mahfouz et al. 2012; Perez-Pinera et al. 2013; Tremblay et al. 2012; Zhang et al. 2011). Synergistic effects on endogenous gene expression have also been observed upon co-expression of multiple TAL effector activators targeting a single promoter in mammalian cells (Maeder et al. 2013; Perez-Pinera et al. 2013).

The ease of engineering TAL effector arrays has enabled large-scale studies never before seen with other nucleases (Reyon et al. 2012; Schmid-Burgk et al. 2013; Kim et al. 2013; Chen et al. 2013). One group assembled TALENs for 18,740

genes spanning the human genome, and showed that all TALENs tested in a pilot study (targeting 124 genes) were active and disrupted their target genes at high frequencies (Kim et al. 2013). In addition to the ability to efficiently scale up studies using TAL effector technology, fusions to yet other functional domains to create TAL effector recombinases, deaminases, and methylases have also been considered to extend the power and utility of TAL effector DNA-binding arrays (Mercer et al. 2012; Pennisi 2012).

Efforts have been put forth to create monomeric nucleases with a single DNA recognition domain (Beurdeley et al. 2013; Fonfara et al. 2012; Kleinstiver et al. 2012). One significant advantage of monomeric nucleases is the reduced size of the protein. TALENs are roughly three times larger than other engineered nucleases (~1,900 aa total), which can frustrate delivery to cells, whether as DNA, RNA, or protein. Furthermore, the highly repetitive nature of TAL effector arrays can complicate efforts to clone, deliver, and integrate a TALEN pair into the genome. Two recent proof-of-principle studies demonstrated that the fusion of either a zinc finger array or a TAL effector protein to the catalytic domain of the I-TevI homing endonuclease resulted in functional designer nucleases with efficiencies of mutagenesis similar to those reported for ZFNs and TALENs (Beurdeley et al. 2013; Kleinstiver et al. 2012).

4 Optimizing TAL Effector Architecture

After studies demonstrated the use of TAL effector fusions for genome modification, subsequent work focused on optimizing the technology, both to enhance targeting capacity and to eliminate undesirable effects. A potential drawback for the use of TAL effectors is their size. Compared to other nuclease platforms, TAL effectors are large (~950 amino acids for a TALEN monomer compared with ~310 amino acids for a ZFN monomer and ~165 amino acids for a meganuclease). Initial engineering efforts focused on determining the minimal TALEN architecture required to achieve maximum binding and nuclease activity. This included (1) optimizing both the N- and C-terminal regions surrounding the repeat array (termed the "backbone") and (2) determining the number of repeat arrays necessary for optimal nuclease activity with spacers of varying lengths. As stated previously, fusions of TAL effector proteins to the FokI nuclease were initially constructed using a full-length, or near full-length, naturally occurring TAL protein (Christian et al. 2010; Li et al. 2010). In the TAL effector PthXo1, for example, 288 and 295 amino acids at the N- and C-termini flank the repeat region, respectively. Features present within the N-terminal region include up to four cryptic repeats, and the C-terminus has two functional NLS along with an AAD (present in the full-length TALEN backbone only).

To delimit the DNA-binding domain, different research groups made and tested various N- and C-terminal truncations (Fig. 4.4b). For example, an N-terminal

truncation at residue 152 (retaining 136 aa, N152) and C-terminal truncations leaving just 18 amino acids after the repeat domain still allow for effective DNA binding (Miller et al. 2011; Mussolino et al. 2011; Sun et al. 2012). The structure of TAL effectors bound to DNA is consistent with these observations (Deng et al. 2012; Mak et al. 2012). An earlier study on TAL effector secretion found that the first 152 residues of the N-terminus mediate transport into plant cells, but are dispensable for other functions (Szurek et al. 2002). However, additional groups reported that further truncation of this region leads to a loss of TAL effector and TALEN function (Mussolino et al. 2011; Sun et al. 2012; Meckler et al. 2013). For example, both TAL effector activators and nucleases retaining 50–94 residues of the N-terminus directly upstream of the first repeat display little or no activity, suggesting that specific residues or structural features within this region influence DNA binding and likely the stability of TAL effector proteins (Sun et al. 2012; Meckler et al. 2013). Indeed, the reported structure of the N-terminal residues bound to DNA reveals cryptic repeats that might initiate critical stabilizing interactions, specifically between the tryptophan (W232) in the −1 repeat of the PthXo1 effector and the conserved T at the 0 position in the TAL effector target sites. N-terminal residues 148–288 autonomously bind DNA nonspecifically, suggesting that the region is responsible for initiating nucleation of the repeat super-helical structure that wraps around the target (Mak et al. 2012; Gao et al. 2012).

The length of the C-terminus immediately after the last half-repeat of a TAL effector DNA-binding domain has a direct impact on the activity of TALENs. In contrast to the N-terminus, it appears that little of the C-terminus is required for TALEN activity (Miller et al. 2011; Zhang et al. 2013; Christian et al. 2012). TAL effectors lacking all but the first 63 aa following the 20th residue of the half-repeat of an array are effective as both activators and nucleases (Miller et al. 2011). TALEN architectures with even further truncation of the C-terminus (termed C17 or C18, retaining 17 or 18 aa) are also active, and require shorter TALEN spacer lengths for optimal activity (12–14 bp) compared to the spacer length requirement for the C63 architecture (15–18 bp) (Mussolino et al. 2011; Christian et al. 2012). In plants, a 25-fold increase in mutagenesis was observed at endogenous targets with the truncated TALENs relative to those with the near full-length architecture (Zhang et al. 2013).

The length of the TAL effector DNA-binding domain also affects binding efficiency, and thus has been subject to optimization. TAL effectors found in nature vary in the number of repeats within a particular array, with an average of 17.5 repeats (Boch and Bonas 2010). A systematic study of TAL effector binding indicates that at least 6.5 repeats are required for activation of a downstream reporter gene, and activity greatly increases with 9.5 repeats and continues to rise as repeats are added thereafter, until a plateau is reached after 14.5 repeats (Boch et al. 2009). The general consensus among data from TALEN studies is that between 14 and 18 repeats is ideal to ensure specific binding.

5 Factors Affecting DNA Binding

Complexities in TAL effector DNA binding were suggested by observations that single base pair mutations in the AvrBs3 recognition site result in both qualitative and quantitative effects on transcriptional activation (Kay et al. 2009; Romer et al. 2009b). Although there is on average no influence of immediate neighbors on RVD–nucleotide associations (Moscou and Bogdanove 2009), individual RVD contributions, RVD composition of the array, and array length affect the affinity and specificity of DNA binding. One landmark study assessed the properties of individual RVDs using artificial TAL effector transcriptional activators (Streubel et al. 2012). Several interesting observations were made: TAL effectors containing strings of HD and NN RVDs were able to bind cognate polycytosine (C) and polyguanine (G) targets, respectively, and activate a downstream reporter gene; however, the same NN TAL effector was unable to activate a reporter with the cognate polyadenine (A) target, despite the known specificity of NN for both G and A nucleotides. Somewhat surprisingly, neither TAL effectors containing NI nor NG RVDs were able to activate reporters with their cognate upstream targets (strings of A or T, respectively), suggesting that there are both "strong" (HD, NN) and "weak" RVDs (NI, NG, NK).

In addition to the four most common or canonical RVDs (HD, NI, NN, and NG), several uncommon or rare RVDs have been shown to provide novel specificities (Miller et al. 2011; Cong et al. 2012; Streubel et al. 2012). The RVD NH has specificity for guanine only, unlike the standard NN RVD that recognizes both G and A nucleotides (although perhaps not with equal affinity). The RVD NP has specificity for three different bases (A, C, and T), and the RVD NS, which was previously designated as nonspecific, showed a clear preference for A and G, although activity of TAL effector arrays with NS were low. In vitro SELEX data first suggested that NK RVDs may have higher specificity for G than NN (Miller et al. 2011); however, modification of a locus in zebrafish was more effective when engineered TALENs were made using repeats with NN RVDs rather than NK (Huang et al. 2011). In vitro binding studies of TALENs containing NN or NK RVDs revealed that binding affinity was greatly reduced with the addition of multiple NK RVDs to an array (Christian et al. 2012).

The guanine-specific NH RVD was shown to be much more efficient at recognizing G than the NK RVD (Cong et al. 2012; Streubel et al. 2012). Interestingly, NH-containing TAL effectors exhibited significantly higher specificity for a G-based reporter than TAL effectors containing the canonical NN RVD (Cong et al. 2012). Computational analysis of the relative binding affinities between the TAL effector and its bound DNA revealed that substitution of a single NH for the NN RVD results in a gain in free energy in the DNA-bound state. Taken together, these data suggest that NH may be a more suitable RVD for targeting G nucleotides. Furthermore, the same study revealed that the NA RVD showed similar specificity for all four bases and may be useful for targeting degenerate DNA sequences.

It is now clear that individual RVDs have differing contributions to the overall interaction between TAL effectors and DNA. The concept of designating RVDs as "strong" or "weak" was supported when TAL effector function was restored after several NI RVDs, which flanked NK RVDs in a nonfunctional artificial TAL effector array, were replaced with HD (Streubel et al. 2012). It is likely that the strong RVDs stabilize the TAL effector/DNA interaction sufficiently such that the presence of weak RVDs does not decrease overall TAL effector activity. Quantitative assessment of the affinity of TAL effectors for DNA has suggested additional complexities in the previously reported simple TAL effector binding code (Meckler et al. 2013; Christian et al. 2012). In one study, the repeat type was the largest factor in affinity differences between TAL effectors (Meckler et al. 2013). Engineered TAL effectors with an even distribution of 10 NG, HD and NN RVDs bound their targets with high affinity compared to those containing NI or NK. Although there was not a clear linear correlation of affinity to activation of a reporter gene, the contribution of particular RVDs to affinity was: NG > HD = NN ≫ NI > NK.

Additional residues within the backbone of TAL effectors might contribute to DNA binding or specificity (Streubel et al. 2012). In addition to the RVDs at positions 12 and 13 in the repeat, there exists variability among natural TAL effectors at positions 4 and 32 (Boch and Bonas 2010). There is little evidence that these residues affect RVD specificity, but rather play a role in positioning the α-helices (Streubel et al. 2012). The differences in amino acid composition in the regions outside of the repeats also appear to have little effect on TAL effector binding, despite previous conjecture (Bogdanove et al. 2010; Streubel et al. 2012). Overall length of a TAL effector repeat array influences efficiency. Early experiments showed that the efficiency of TAL effector-mediated transcription increases until greater than nine to ten repeats are used (Boch et al. 2009). This is in agreement with experiments performed with a short TAL effector array of only 10.5 that contained either the RVD NN or NH at three fixed positions: only the NN-containing TAL effector was able to activate transcription of the reporter. When additional repeats were added to create a 17.5 repeat array, both the NN and NH-containing TAL effectors were active (Streubel et al. 2012).

The directionality of TAL effector protein binding begs the question of whether RVDs at different positions in the array contribute equally to affinity and specificity. Several studies suggested that RVD-nucleotide mismatches at the 5′ end of both TAL effector and TALEN targets have greater effects than those made at the 3′ end (Meckler et al. 2013; Christian et al. 2012; Blount et al. 2012). Specifically, substitution of nucleotides in the first 3 or 6 bp of a target site reduced TAL effector binding 15-fold and 370-fold, respectively. The same number of substitutions at the 3′ end had no significant effect (Meckler et al. 2013).

A major drawback for nuclease-mediated genome engineering is the sensitivity of DNA-binding proteins to cytosine methylation and other epigenetic marks, which are widespread in different cell types across species. It has been reported that TAL effector DNA-binding domains are affected by the presence of 5-methyl cytosine (5 mC) at their endogenous targets (Bultmann et al. 2012). This epigenetic modification is found in approximately 70 % of CpG dinucleotides in mammalian and

plant somatic and pluripotent cells, as well in CpA, CpT, and CpC dinucleotides (Jones 2012). Moreover, regions of genes most often targeted for gene editing (i.e., promoters and proximal exons) contain extensive 5 mC, particularly in CpG islands. The standard RVD HD is most frequently used to specify cytosine. However, in cases where cytosine is methylated, the structural data predict a clash between this 5-methyl group and the side chain of aspartic acid (D of HD) (Deng et al. 2012; Mak et al. 2012). Hence, methylation should work against this pairing.

Two groups independently reported that both NG and N* (asterisk indicates a deletion at residue 13 in the repeat unit) RVDs can accommodate 5 mC in vitro and in vivo (Deng et al. 2012; Valton et al. 2012). Based on solved crystal structures containing TAL effectors with NG RVDs bound to DNA targets, it was hypothesized that the lack of a side chain in Gly (of NG RVD) provides sufficient space to accommodate the 5-methyl group of thymine (T) nucleotides (Deng et al. 2012). This observation suggested that Gly13 of RVDs might also accommodate 5 mC due to its structural similarity to T. TAL effector arrays were tested for their binding affinity for 5 mC using electrophoretic mobility shift assays (EMSA), and only those containing NG RVDs in place of the standard HD to specify C were able to bind their cognate targets. The RVD N* was also used accommodate 5 mC as the first example of TALEN-mediated modification of a methylated locus in human cells (Valton et al. 2012). In the consensus of all repeats, the variable residues (aa 12 and 13) are followed immediately by two conserved glycine residues; thus N* is highly similar to NG except for a shortened RVD loop. The DNA-binding capacity and activity of N*-containing TAL effectors remained unaffected by 5 mC, whether present on one or both strands of the DNA (Valton et al. 2012). The RVD HG has similar binding specificity to NG and might also be a suitable alternative to specify 5 mC. Thus although binding to DNA by some TAL effectors is thwarted by DNA methylation, TAL effector arrays can be engineered to overcome this impediment.

6 Conclusions

The opportunity provided by engineered TAL effector proteins for the targeted manipulation of plant genomes has yet to be fully realized. Advances, however, are being made at a remarkable pace. In the span of just a few short years, we have come to understand the molecular basis by which TAL effectors recognize DNA. Importantly, the high degree of modularity of the TAL effector repeat achieves DNA targeting with unprecedented efficiency and fidelity. To date, most custom TAL effector arrays have been fused to nucleases to create chromosome breaks that enable precise DNA sequence alterations to plant genomes. Future work will focus on more effective means of delivering these reagents to plant cells and the implementation of strategies to better achieve desired DNA sequence modifications. Considerable potential also exists to control plant gene expression and the epigenetic state of target loci by fusing TAL effector arrays to a variety of DNA

and chromatin modifying enzymes. Without a doubt, effective DNA targeting through the use of reagents like custom TAL effector arrays has ushered in a new era in plant biology, providing newfound control over the plant genome to achieve a better understanding of plant gene function and to harness plant genes to create novel traits of value.

Acknowledgements We thank Kit Leffler for help in preparing the figures. This work was supported by a grant to DFV from the National Science Foundation (DBI 0923827).

References

Ansai S et al (2013) Efficient targeted mutagenesis in medaka using custom-designed transcription activator-like effector nucleases. Genetics 193:739–749

Beurdeley M et al (2013) Compact designer TALENs for efficient genome engineering. Nat Commun 4:1762

Blount BA, Weenink T, Vasylechko S, Ellis T (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. PLoS One 7:e33279

Boch J, Bonas U (2010) *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol 48:419–436

Boch J et al (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512

Bogdanove AJ, Schornack S, Lahaye T (2010) TAL effectors: finding plant genes for disease and defense. Curr Opin Plant Biol 13:394–401

Bonas U, Stall RE, Staskawicz B (1989) Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. vesicatoria. Mol Gen Genet 218:127–136

Bradley P (2012) Structural modeling of TAL effector-DNA interactions. Protein Sci 21:471–474

Briggs AW et al (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. Nucleic Acids Res 40:e117

Bultmann S et al (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. Nucleic Acids Res 40:5368–5377

Carlson DF et al (2012) Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci U S A 109:17382–17387

Cermak T et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39:e82

Chen LQ et al (2010) Sugar transporters for intercellular exchange and nutrition of pathogens. Nature 468:527–532

Chen S et al (2013) A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. Nucleic Acids Res 41:2769–2778

Christian M et al (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186:757–761

Christian ML et al (2012) Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. PLoS One 7(9):e45383

Cong L, Zhou R, Kuo YC, Cunniff M, Zhang F (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun 3:968

Deng D et al (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. Science 335:720–723

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3:e3647

Engler C, Gruetzner R, Kandzia R, Marillonnet S (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 4:e5553
Fonfara I, Curth U, Pingoud A, Wende W (2012) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. Nucleic Acids Res 40:847–860
Gao H, Wu X, Chai J, Han Z (2012) Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. Cell Res 22:1716–1720
Garg A, Lohmueller JJ, Silver PA, Armel TZ (2012) Engineering synthetic TAL effectors with orthogonal target sites. Nucleic Acids Res 40:7584–7595
Geissler R et al (2011) Transcriptional activators of human genes with programmable DNA-specificity. PLoS One 6:e19509
Gurlebeck D, Szurek B, Bonas U (2005) Dimerization of the bacterial effector protein AvrBs3 in the plant cell cytoplasm prior to nuclear import. Plant J 42:175–187
Gurlebeck D, Thieme F, Bonas U (2006) Type III effector proteins from the plant pathogen Xanthomonas and their role in the interaction with the host plant. J Plant Physiol 163: 233–255
Huang P et al (2011) Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol 29:699–700
Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13:484–492
Kay S, Boch J, Bonas U (2005) Characterization of AvrBs3-like effectors from a Brassicaceae pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. Mol Plant-Microbe Interact 18:838–848
Kay S, Hahn S, Marois E, Hause G, Bonas U (2007) A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science 318:648–651
Kay S, Hahn S, Marois E, Wieduwild R, Bonas U (2009) Detailed analysis of the DNA recognition motifs of the Xanthomonas type III effectors AvrBs3 and AvrBs3Deltarep16. Plant J 59: 859–871
Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A 93:1156–1160
Kim Y et al (2013) A library of TAL effector nucleases spanning the human genome. Nat Biotechnol 31:251–258
Kleinstiver BP et al (2012) Monomeric site-specific nucleases for genome editing. Proc Natl Acad Sci U S A 109:8061–8066
Li T et al (2010) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res 39:359–372
Li T, Liu B, Spalding MH, Weeks DP, Yang B (2012a) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol 30:390–392
Li L et al (2012b) Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. Plant Mol Biol 78:407–416
Ma S et al (2012) Highly efficient and specific genome editing in silkworm using custom TALENs. PLoS One 7:e45035
Maeder ML et al (2013) Robust, synergistic regulation of human gene expression using TALE activators. Nat Methods 10:243–245
Mahfouz MM et al (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A 108:2623–2628
Mahfouz MM et al (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. Plant Mol Biol 78:311–321
Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. Science 335:716–719
Mansfield J et al (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. Mol Plant Pathol 13:614–629

Meckler JF et al (2013) Quantitative analysis of TALE-DNA interactions suggests polarity effects. Nucleic Acids Res 41:4118–4128
Mercer AC, Gaj T, Fuller RP, Barbas CF 3rd (2012) Chimeric TALE recombinases with programmable DNA sequence specificity. Nucleic Acids Res 40(21):11163–11172
Miller JC et al (2011) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29:143–148
Morbitzer R, Romer P, Boch J, Lahaye T (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci U S A 1078:21617–21622
Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501
Mussolino C et al (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res 39:9283–9293
Pennisi E (2012) The tale of the TALEs. Science 338:1408–1411
Perez-Pinera P et al (2013) Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat Methods 10:239–242
Reyon D et al (2012) FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30:460–465
Romer P et al (2007) Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. Science 318:645–648
Romer P, Recht S, Lahaye T (2009a) A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. Proc Natl Acad Sci U S A 106:20526–20531
Romer P et al (2009b) Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper Bs3 alleles. Plant Physiol 150:1697–1712
Schmid-Burgk JL, Schmidt T, Kaiser V, Honing K, Hornung V (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. Nat Biotechnol 31:76–81
Scholze H, Boch J (2010) TAL effector-DNA specificity. Virulence 1:428–432
Shan Q et al (2013) Rapid and efficient gene modification in rice and Brachypodium using TALENs. Mol Plant 6:1365–1368
Streubel J, Blucher C, Landgraf A, Boch J (2012) TAL effector RVD specificities and efficiencies. Nat Biotechnol 30:593–595
Sun N, Liang J, Abil Z, Zhao H (2012) Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. Mol Biosyst 8:1255–1263
Szurek B, Rossier O, Hause G, Bonas U (2002) Type III-dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell. Mol Microbiol 46:13–23
Tremblay JP, Chapdelaine P, Coulombe Z, Rousseau J (2012) Transcription activator-like effector proteins induce the expression of the frataxin gene. Hum Gene Ther 23:883–890
Valton J et al (2012) Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J Biol Chem 287:38427–38432
Wendt T et al (2013) TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. Plant Mol Biol 83:279–285
Zhang F et al (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29:149–153
Zhang Y et al (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol 161(1):20–27

Chapter 5
Oligo-Mediated Targeted Gene Editing

Greg F.W. Gocal, Christian Schöpke, and Peter R. Beetham

Abstract Oligonucleotide-mediated targeted gene editing encompasses technology described by many names. These include Oligo-Directed Mutagenesis (ODM) and the commercial Rapid Trait Development System (***RTDS***™) from Cibus. ODM is a non-transgenic (non-GMO) base pair-specific oligonucleotide-directed gene editing platform that has been advanced at Cibus over the past decade and has achieved novel and commercially valuable traits in crops. This technology harnesses the cell's normal DNA repair system to correct and change specific targeted bases within the genome of a cell. The Gene Repair OligoNucleotide (GRON), a chemically synthesized oligonucleotide, is designed to create mismatched base pairs compared to the target sequence within the host organism's genome. The GRON hybridizes at the target region and the mismatched base pairs work to direct the cell's repair system at those sites to correct (replace, insert, or delete) the designated base(s). Once the correction process is complete the GRON is degraded and the now-modified or repaired gene retains its normal pattern of expression and stability within the genome. This technique has been successfully deployed in bacterial, fungal, mammalian, and plant systems. Our work in achieving herbicide tolerance traits in acetohydroxyacid synthase (AHAS) genes in oil seed rape (OSR) as well as work converting a blue fluorescent protein (BFP) transgene to green fluorescent protein (GFP) in an *Arabidopsis* model system will be discussed.

1 Introduction

Chemically synthesized and non-modified oligonucleotides were first used for gene editing in vivo in yeast (Moerschell et al. 1988). Almost 8 years later, oligonucleotides were used in mammalian cells initially to correct an episome (Yoon et al. 1996) and then to correct the point mutation within the nuclear genome in the human β-globin gene that causes sickle cell anemia (Cole-Strauss et al. 1996).

G.F.W. Gocal, M.Sc., Ph.D. (✉) • C. Schöpke, M.Sc., Ph.D. • P.R. Beetham, Ph.D.
Cibus, 6455 Nancy Ridge Dr., San Diego, CA 92121, USA
e-mail: ggocal@cibusllc.com; cschopke@cibusllc.com; pbeetham@cibusllc.com

F. Zhang et al. (eds.), *Advances in New Technology for Targeted Modification of Plant Genomes*, DOI 10.1007/978-1-4939-2556-8_5

At that time, since the oligonucleotide chemistry used was a RNA/DNA chimeric molecule, the first generation of the ODM technology was termed chimeraplasty.

This exciting result spawned a commercial foray into applying this technology for gene editing in a variety of organisms and systems demonstrating its applicability across plant, animal, human, fungal, and bacterial cells, and showing that this technique does not appear to be restricted by cell cycle, gene expression, transcription, or other known cellular or genetic activities (Cole-Strauss et al. 1996; Kren et al. 1997; Beetham et al. 1999; Zhu et al. 1999; Alexeev and Yoon 1998; Kochevenko and Willmitzer 2003; Okuzaki and Toriyama 2004; Dong et al. 2006). The ODM process produces non-transgenic, site-specific mutations. Traits produced using ODM are similar to the more than 2,500 products of classical mutagenesis in 180 crops including Clearfield canola, wheat, rice, and sunflower, which are already moving freely in world trade with no labeling or other restrictions applied (http://www.fao.org). In plants the initial proof of concept experiments were performed in tobacco and corn (Beetham et al. 1999; Zhu et al. 1999, 2000). In both cases point mutations in acetohydroxyacid synthase (AHAS) genes conferring tolerance to AHAS-inhibiting herbicides were targeted. AHAS-inhibiting herbicides are a modern class of herbicides that are used widely in broad acre agriculture. Similar to many modern herbicides the interaction between the plant's AHAS and the herbicide is well understood (Tan et al. 2005) and specific mutations in the gene(s) encoding AHAS are directly related to herbicide tolerance. Sulfonylurea herbicides are one of the five classes of AHAS-inhibiting herbicides that block the biosynthesis of the branched-chain amino acids valine, leucine, and isoleucine (Tan et al. 2005). Tobacco is an allotetraploid species with two AHAS-encoding loci, SuRA and SuRB. In the experiments of Beetham et al. (1999), one of two AHAS genes/loci (SuRA) was targeted in a tobacco cell line known as Nt-1. At a cellular level, this work targeted an amino acid substitution at P196 to confer resistance to the herbicide chlorsulfuron (Glean, DuPont). This paper demonstrated specificity for SuRA and reported evidence that ***RTDS*** could also reactivate a mutant transgene encoding a green fluorescent protein (GFP) marker gene, introduced into either the tobacco cells or whole plants.

In complementary studies by scientists at Pioneer Hi-Bred International, Inc., Zhu et al. (1999, 2000) converted the AHAS homologs in maize, again to making the cells herbicide tolerant. Additionally, they converted and therefore re-activated a mutant GFP transgene in maize. Cells of both the AHAS and GFP conversions were then cultured on various media; plants were regenerated and allowed to mature. Progeny of these plants confirmed that the gene conversions were heritable and stable as is expected for Mendelian inheritance.

It is well documented that the molecular basis of many traits is due to small genetic differences, or single nucleotide polymorphisms (SNPs), within critical genes. For more than a decade, Cibus, its predecessors and others have been developing and applying this molecular "spell checking" technology to *direct conversions* (nucleotide changes; SNPs) to the desired location in the specified gene in nuclear-encoded genes to develop traits in crop plants. These changes can be site-specific nucleotide substitutions, as well as insertions and deletions in individual

genes of multigene families. Therefore the conversion process can be used to repair mutant genes, alter genes, or interrupt normal gene function. This technology is not limited to manipulating genes that code for proteins, but has demonstrated that any nucleotide sequence (regulatory, coding, and noncoding) can be converted to enhance and/or reduce the function and activity of a gene product. In all cases, the SNP changes are emulating what often occurs in nature as part of the continuing natural genetic diversity. We see this in exponentially increasing detail as we generate more genome sequence data from plant populations.

From a regulatory stand point, the USDA has classified this technology as a mutagenesis technique like chemical mutagenesis, only with a directed outcome. An assessment of the ***RTDS*** technology, in Europe known as oligonucleotide-directed mutagenesis (ODM), was published several years ago by a Belgian group (Breyer et al. 2009). Their conclusions are in line with the USDA analysis where the outcomes of the technology are similar to traditional mutagenesis. It is clear that ***RTDS***-derived products are produced from a more precise and targeted technology as compared to traditional mutagenesis. In theory these products are isogenic to the parental genetics except for the targeted SNP. Regulators are favorable to technologies that are more defined and measurable (Kuzma and Kokotovich 2011). One difficulty will be that SNPs will occur in nature and if by chance these are the same as ***RTDS-***derived SNPs, they will be indistinguishable from each other. Interestingly these SNPs work well with more advanced breeding techniques now deployed in agriculture. Many quantitative trait loci (QTLs) are followed using SNPs as molecular markers in advanced breeding programs are well accepted protocols.

1.1 Gene Repair Oligonucleotide Structure

At Cibus and in the public domain, many oligonucleotide designs have demonstrated conversion activity (Beetham et al. 1999; Metz et al. 2002; Dong et al. 2006; Wagner et al. 2010). Since there are cellular enzymatic activities which can destroy open-ended RNA and DNA mono and duplex molecules, the ends of these molecules can be protected by various chemical modifications including capping with "hairpins." Such hairpins were used in the original DNA and RNA chimeric oligonucleotides. Self-complementary chimeric oligonucleotides contain a DNA "mutator" region of 5-bp nucleotides complementary to the target site flanked by 2′-*O*-methyl RNA segments. The "mutator" region is synthesized with a mutation designed to be introduced into the endogenous target gene (Beetham et al. 1999; Zhu et al. 1999; Gamper et al. 2000; Oh and May 2001; Kochevenko and Willmitzer 2003; Ruiter et al. 2003; Okuzaki and Toriyama 2004). This design attempted to mimic the formation of a complement-stabilized D-loop to direct the targeted conversion.

Second generation GRON designs are linear molecules approximately 40 nucleotides in length with blocked ends. These molecules are designed to be complementary with either the coding or the noncoding strand of the target gene, with targeting GRONs generally having the targeted nucleotide change (mismatch) towards their

center. Non-targeting GRONs, which are used as controls, are completely homologous with the target sequence. Within the peer-reviewed literature, the most common design generally blocks the oligo's termini with three phosphorothioate linkages (Pierce et al. 2003). Another design blocks the 5′ terminus with the fluorescent label Cy3 and 3′ terminus with a reverse cytosine base (Dong et al. 2006; Wagner et al. 2010).

1.2 Delivery

In plant systems, various methods for GRON deliver have been used successfully to convert endogenous gene targets. For protoplast systems, these include electroporation and PEG-mediated delivery, whereas for monocot crops including corn, rice, and wheat, particle bombardment has been used as the preferred method (Zhu et al. 1999; Kipp et al. 2000; Okuzaki and Toriyama 2004; Dong et al. 2006). Many of the early delivery studies were focused on single cells or embryogenic cell suspensions using biolistics (Kipp et al. 2000). They focused on whether plant cells had the gene repair machinery to correct a nonfunctional plasmid carrying the GFP gene. These studies confirmed that both delivery of the plasmid and GRONs and subsequent successful repair was possible in many cells. These cells also continued to grow. Studies also showed that cell-free extracts from various tissues and plants where also competent for plasmid repair using GRONs (Kmiec et al. 2001). These studies have clearly shown that once the parameters for delivery are optimized for GRONs, plant cells have the competence do complete ODM.

Importantly, the next endeavors for delivery were to complete chromosomal conversion. The two gene targets where this work has been described are AHAS and GFP (Beetham et al. 1999; Zhu et al. 1999, 2000; Kochevenko and Willmitzer 2003; Okuzaki and Toriyama 2004; Dong et al. 2006). Here we also describe work done more recently using a Blue Fluorescent Protein (BFP) chromosomal target, where by changing one nucleotide the BFP becomes GFP.

1.3 RTDS Mechanism/Process

By (a) mechanism(s) that are becoming more clearly understood, the mismatched nucleotide directs a specific mutation in the target gene (see Fig. 5.1). We loosely designate this process as "*gene conversion*." Once conversion is complete, the GRON is degraded and the modified gene remains expressed under the cell's normal control mechanisms. In vitro assays using "cell-free" extracts and various pure enzymes demonstrate that the GRON is stable for sufficient time to direct gene conversion and then is quickly degraded.

Requirements for gene conversion vary considerably between bacterial, yeast, and mammalian systems. In *E. coli* both MutS and RecA activities are required (Metz et al. 2002). However, in yeast overexpression of a RecA homolog RAD51 or

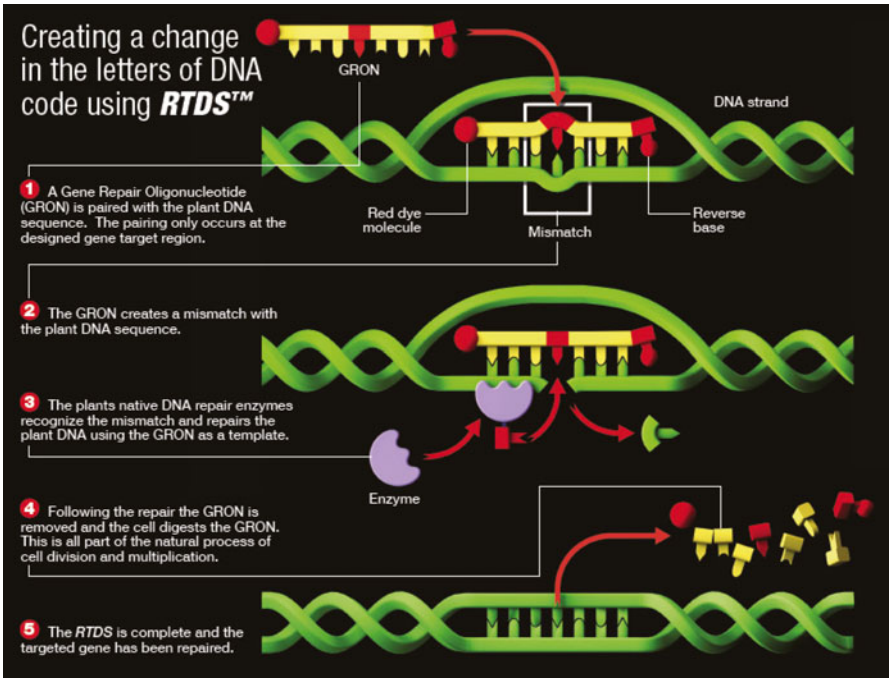

Fig. 5.1 The ***RTDS*** process. Included with permission from Cibus

the helicase RAD54 significantly improved conversion (Liu et al. 2002). In contrast, complete loss of expression of MSH2, a component of Mutα (MSH2 and MSH6), a eukaryotic homolog of MutS did not affect conversion efficiency (Rice et al. 2001) or might even reduce efficiency (Maguire and Rice 2007). Also in yeast, loss of function mutations in either of the two Mutβ components (MSH3 and MSH6) somewhat improved conversion efficiency (Rice et al. 2001). By comparison, in mouse embryonic stem cells, overexpression of RAD51 or RAD54 significantly improved conversion, whereas a loss of function mutant of msh2(−/−), a component of Mutα (MSH2 and MSH6) and a eukaryotic homolog of MutS, substantially improved conversion efficiency (Aarts et al. 2006; Morozov and Wawrousek 2008). Furthermore Morozov and Wawrousek (2008) significantly improved conversion efficiency by inhibiting Ku70/86, components of the nonhomologous end-joining pathway.

1.4 Application of RTDS

1.4.1 Targeting Acetohydroxyacid Synthase

As detailed above for tobacco and maize as well as in rice (Beetham et al. 1999; Zhu et al. 1999, 2000; Kochevenko and Willmitzer 2003; Okuzaki and Toriyama 2004), a common target used for conversion experiments of targets within plant nuclear

genomes is their acetohydroxyacid synthase (AHAS) gene. In OSR, the BnAHAS I and BnAHAS III genes both code for nearly identical (99 % homology) protein monomers. The BnAHAS I gene is located in the C genome, whereas BnAHAS III is located in the A genome of *Brassica napus*. The proteins encoded by these genes combine in all three permutations and combinations to form the active AHAS enzyme. Through chemical mutagenesis, a mutation was found in the AHAS I gene (Tan et al. 2005). This mutation is known as PM-1 (a mutation at an equivalent position known as 653 based in the AHAS amino acid sequence of *Arabidopsis*, a serine to asparagine amino acid change, respectively, encoded as AGT to AAT). Another mutation was found in the higher expressed gene AHAS III, known as PM-2 (a mutation at an equivalent position known as 574, a tryptophan to leucine amino acid change, respectively, encoded as TGG to TTG). These two mutations, PM-1 and PM-2, are combined in a commercial variety of Canola known as Clearfield Canola (Tan et al. 2005). Although the level of conferred tolerance to each herbicide may vary between the PM1 and PM2 mutations, each mutation provides tolerance to the plant by disrupting the herbicide's ability to interact with the AHAS. The introduction of a single base change in the nucleotide sequence at a specific location in each of the AHAS I and III genes produces enzymes, each with a change in a single amino acid that reduces the ability of the herbicide to block the active binding site of that enzyme, allowing branched chain amino acid synthesis to occur even in the presence of the herbicide.

1.4.2 Targeting Green Fluorescent Protein Transgenics

We utilized the well-known green fluorescent protein (GFP) gene for monitoring conversion resulting from ODM. In the peer reviewed literature, the single amino acid substitution at amino acid 66 of GFP from Y66 (Tyrosine encoded by the codon TAC) to H66 (Histidine encoded by the codon CAC) causes the protein to appear blue rather than green (Sommer et al. 2006). By transforming plants with this modified GFP at this position, we engineered a system where with one SNP we could switch its fluorescence from blue to green.

In this chapter we will discuss conversion experiments of two targets (1) the AHAS loci in oil seed rape (OSR; aka canola) and (2) blue fluorescent protein (BFP) in transgenic *Arabidopsis* lines.

2 Materials and Methods

2.1 Oil Seed Rape AHAS

In its simplest form, the ODM process involves introducing into cells a GRON designed to target a specific mutation within a gene target, the GRON pairing by homology with its target and causing the intended mutation followed by

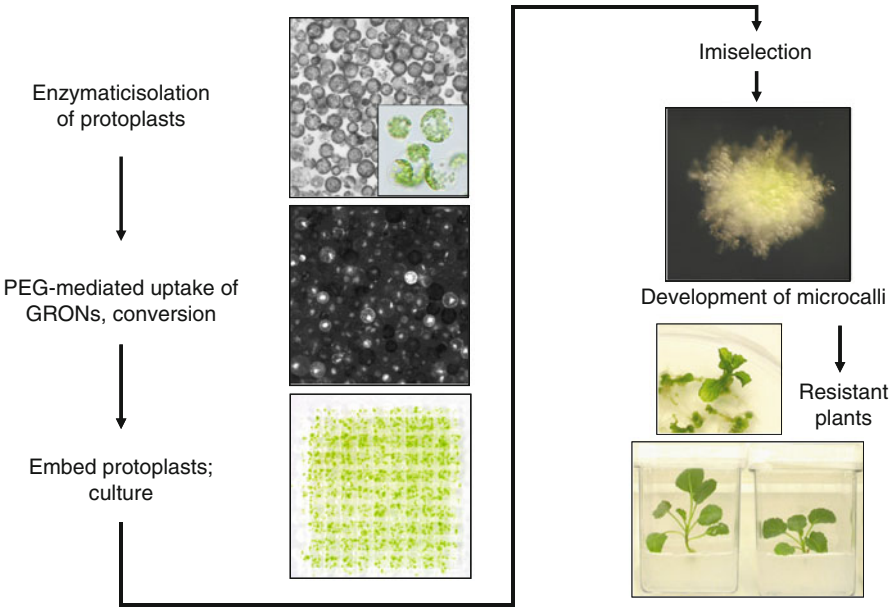

Fig. 5.2 The operating system for ODM in OSR. Leaf-derived protoplasts are isolated, GRONs introduced and the protoplasts are embedded in calcium alginate grids where they are cultured to form microcalli. After continuous selection, imazethapyr-tolerant calli develop. Tolerant calli are transferred to shoot induction media and the developing shoots rooted in vitro

regeneration of the converted cells into normal fertile plants. The cell biology process illustrating these steps is presented in Fig. 5.2. In brief, shoots derived both from seeds and from microspore-derived embryos were propagated under sterile conditions in vitro. Cuttings were subcultured every 2–4 weeks and cultured in Petri dishes (25×90 mm) in a volume of 40–45 mL RS medium (Dovzhenko 2001) and young leaves used for protoplast isolation.

As detailed in patent application WO 2009 046334, leaves of 2–3-week-old in vitro shoots were cut into small strips with a scalpel in a Petri dish with 5 mL medium B (Pelletier et al. 1983), and the pH adjusted to 5.8. After approximately 1 h, medium B was replaced with enzyme solution, consisting of medium B in which 0.5 % (w/v) Cellulase YC and 0.75 % (w/v) Macerozyme R10 (both from Karlan Research Products, Cottonwood, Arizona), 1 g/L bovine serum albumin, and 1 g/L 2-morpholinoethanesulfonic acid were dissolved. The enzyme solution was vacuum-infiltrated into the leaf tissue, and the dish with leaf pieces in enzyme solution was incubated for at 25 °C in darkness. Protoplast purification was performed using an iodixanol density gradient (adapted from Optiprep Application Sheet C18; Purification of Intact Plant Protoplasts; Axis-Shield USA, 10 Commerce Way, Norton, MA 02776). After the density gradient centrifugation, the band with purified

protoplasts was removed together with about 5 mL W5 medium (Negrutiu et al. 1987). The protoplast yield was determined with a hemocytometer, and the protoplasts were stored for 2 h at 4 °C.

The protoplast suspension was mixed with an equal volume of W5 medium, transferred to a 50 mL centrifuge tube, and centrifuged for 5 min at the lowest setting of a clinical centrifuge (about 50 × *g*). The supernatant was removed and replaced with TM medium (Klaus 2003), adjusting the protoplast density to 5×10^6/mL. Aliquots of 100 μL containing 5×10^6 protoplasts each were distributed into 12 mL round bottom centrifuge tubes. In each experiment a variety of control treatments were performed including no treatment, PEG treatment alone, and non-targeting GRONs (whose sequences were identical to the target sequences). GRONs targeting a mutation in one or both AHAS genes were then introduced into the protoplasts using a PEG treatment. To introduce the GRONs into the protoplasts, 12.5 μg GRON dissolved in 25 μL purified water and 125 μL of a polyethylene glycol solution (5 g PEG MW 1500, 638 mg mannitol, 207 mg $CaNO_3 \times 4H_2O$, and 8.75 mL purified water; pH adjusted to about 9.0) was added. After a 30 min incubation on ice, the protoplast-PEG suspension was washed with W5 medium and resuspended in medium B. The suspension was kept overnight in a refrigerator at about 4 °C.

One day after the GRON introduction, protoplasts were embedded in calcium alginate. The embedding of protoplasts in gel substrates (e.g., agarose, alginate) has been shown to enhance protoplast survival and to increase division frequencies of protoplast-derived cells. The method applied was based on that described in Dovzhenko (2001).

The selection of imazethapyr-resistant calli was carried out using sequential subcultures of the alginates in media according to Pelletier et al. (1983). Selection was started 1 week after the PEG/GRON treatment at a concentration of 0.5 μM imazethapyr. Cells and colonies were released by treating the alginate for 30–45 min with culture medium containing 50 mM sodium citrate. Two to three weeks after the transfer to solidified selection medium (occasionally earlier), actively growing calli appeared among a background of brownish cells and microcalli.

Over time, various SNP screening methods have been used to identify converted callus lines with targeted mutations in either BnAHAS I or BnAHAS III, and sequence confirmation was obtained for these mutations in the relevant locus. Plants were regenerated from protoplast-derived, herbicide-tolerant calli with a confirmed mutation in an AHAS gene. Imazethapyr-tolerant calli that had developed on solidified selection medium and whose DNA upon analysis had shown the presence of a mutation were transferred to herbicide-free medium E (Pelletier et al. 1983) to accelerate development. Individual callus lines varied in their growth rates and morphologies. In general, the development towards shoot regeneration followed these steps: Undifferentiated, green callus → callus with dark green areas → development of roots → development of shoot initials → development of stunted shoots with hyperhydric (vitrified) leaves.

Once shoots with three to four leaves had formed on medium E, they were transferred to RS medium (Dovzhenko 2001). On this medium, over time shoot and leaf tissue developed that was morphologically "normal" (i.e., non-hyperhydric). After in vitro plantlets had produced roots, standard protocols were used for the adaptation to greenhouse conditions.

2.2 BFP to GFP Conversion in Arabidopsis

A. thaliana ecotype Col-0 lines containing single or multiple BFP copies were produced through *Agrobacterium*-mediated transformation (Clough and Bent 1998). As described in Sect. 1.4, the BFP gene used for transformation was based on an eGFP gene into which a mutation was introduced at amino acid position 66 (Tyr66 encoded by codon TAC to His66 encoded by codon CAC). For expression in plant cells, the BFP coding sequence was driven by the mannopine synthase promoter (Guevara-García et al. 1999) and was used in combination with the pea rbcS-E9 terminator (Genbank Accession number M21375). Protoplast isolation, PEG-mediated DNA delivery, and protoplast culture were carried out essentially as described in Mathur et al. (1995).

The sequences of the targeting coding (C) and noncoding (NC) GRONs are as per Wagner et al. (2010). The non-targeting versions of these GRONs are identical to the BFP target in sequence (His66 encoded by codon CAC). The phosphorothioate version of these GRONs lack the Cy3 label and reverse base at their 5′ and 3′ termini, respectively, and have three phosphorothioate linkages at both ends.

For the determination of conversion efficiency, the percentage of green-fluorescing protoplasts was measured one day after GRON delivery using flow cytometry (Guava EasyCyte; Millipore). In other experiments protoplast were cultured in liquid medium in 24-well dishes, and the development of green-fluorescing (=converted) cells was followed over time. Images were acquired with an ImageXpress Micro System (Molecular Devices) equipped with a black-and-white CCD camera. Black-and-white images of green-fluorescing cells were colorized green using the imaging program MetaXpress (Molecular Devices).

3 Results

3.1 Chromosomal Conversion

3.1.1 Oil Seed Rape AHAS

Over the years, Cibus has successfully targeted three independent mutations in two AHAS loci in canola (BnAHAS I and BnAHAS III). The dataset shown in Fig. 5.3 consists of experiments targeting the mutations (SNPs) S653N (A**G**T to A**A**T) and

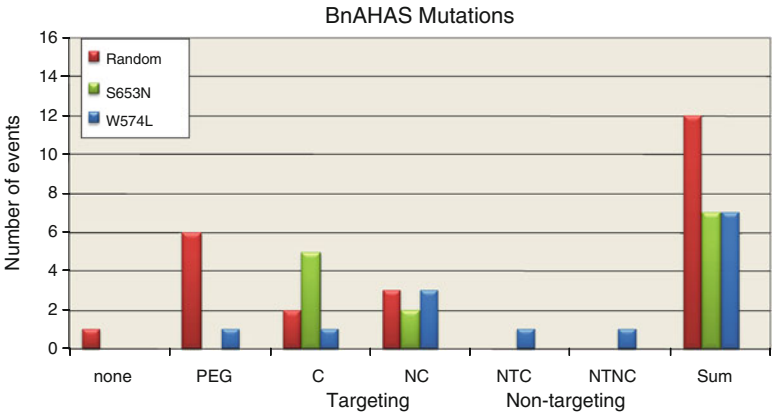

Fig. 5.3 Frequency of events with mutations in either BnAHAS I or BnAHAS III obtained in ten experiments with 8 targeting S653N and 2 targeting W574L. Callus forming efficiency is approximately tenfold higher in the control treatments without GRON (none and PEG) than in its presence (targeting or non-targeting GRON). *Green bars* indicate the presence of S653N (AAT), *blue bars* indicate the presence of W574L (TTG), and *red bars* indicate the presence of other mutations (not S653N or W574L) in either BnAHAS I or BnAHAS III

W574L (T**G**G to T**T**G) in both BnAHAS I and BnAHAS III. Of 10 experiments, 8 targeted S653N and 2 targeted W574L using the targeting GRONs detailed in Table 5.1. The target regions were identical to those in the reference sequences for BnAHAS I (Genbank accession# Z11524) and BnAHAS III (Genbank accession# Z11526; Rutledge et al. 1991). In each experiment, similar numbers of protoplasts were used in control treatments (no treatment, PEG alone, and/or non-targeting GRON—also detailed in Table 5.1). Under optimal conditions, up to 2 % of the PEG treated protoplasts develop into microcalli (see Fig. 5.2). In the absence of herbicide selection, from GRON introductions using up to 60 million protoplasts, as many as 1.2 million calli can form. As expected, somatic mutations were obtained in control treatments for nearly every experiment, the targeted S653N mutation in either BnAHAS I or III was *only* obtained in treatments employing either a coding or a noncoding targeting GRON. In fact the S653N mutation in BnAHAS III in the OSR lines we were able to obtain in our experiments was not obtained in chemical mutagenesis experiments performed over several decades in OSR (Tan et al. 2005).

The Cy3 label enables visualization of delivery efficiency with >80 % of the treated protoplasts receiving GRON in these experiments.

By contrast, in experiments targeting the W574L mutation, a similar number of targeted mutations (one line obtained with the coding GRON and 2 obtained using the noncoding GRON) were obtained in treatments with targeting GRONs compared with events obtained in control treatments. The dataset for the experiments targeting W574L is, however, significantly smaller than that for the S653N experiments. Regenerated plants and seed was obtained from more than 80 % of OSR calli with confirmed AHAS mutations with the level of Imi tolerance conferred by each targeted mutation being indistinguishable across multiple lines with each mutation.

Table 5.1 GRON sequences

	GRON	Sequence
Targeting	BnALS1653/C/41/5′Cy3/3′idC	VTGTGTTACCGATGATCCCAA**A**TGGTGGCACTTTCAAAGATGH
	BnALS1653/NC/41/5′Cy3/3′idC	VCATCTTTGAAAGTGCCACCA**T**TTGGGATCATCGGTAACACAH
	BnALS1574/C/41/5′Cy3/3′idC	VCTTGGGATGGTCATGCAATTGGAAGATCGGTTCTACAAAGCH
	BnALS1574/NC/41/5′Cy3/3′idC	VGCTTTGTAGAACCGATCTTCCAATTGCATGACCATCCCAAGH
Non-targeting	BnALS0653/C/41/5′Cy3/3′idC	VTGTGTTACCGATGATCCCAAGTGGTGGCACTTTCAAAGATGH
	BnALS0653/NC/41/5′Cy3/3′idC	VCATCTTTGAAAGTGCCACCACTTGGGATCATCGGTAACACAH
	BnALS0574/C/41/5′Cy3/3′idC	VCTTGGGATGGTCATGCAATGGGAAGATCGGTTCTACAAAGCH
	BnALS0574/NC/41/5′Cy3/3′idC	VGCTTTGTAGAACCGATCTTCCCATTGCATGACCATCCCAAGH

The converting base is shown in bold
V = Cy3; *H* = 3′DMT dC CPG

3.1.2 BFP to GFP Conversion in *Arabidopsis*

The first BFP-transgenic *Arabidopsis* lines obtained had multiple copies of the BFP gene. Experiments with these lines aimed at comparing the effect of coding and noncoding GRONs and showed that the coding GRON resulted in conversion barely above background, while the noncoding GRON achieved conversion rates in the range of 0.05–0.1 % (Fig. 5.4a). Conversion showed dependence on the concentration of the noncoding GRON. Increasing the GRON concentration from 40 to 80 μM led to a threefold increase, and at 160 μM the percentage of converted protoplasts was six times higher compared with the 40 μM treatment (Fig. 5.4b). Subsequent experiments with a line containing a single copy of the BFP transgene resulted in similar conversion frequencies.

Protoplast derived from BFP/GFP conversion experiments were cultured in 24-well dishes in liquid medium MSAR1 (Mathur et al. 1995). The first division of protoplast-derived cells was observed after about 1 week. Interestingly, the majority of dividing converted cells followed an unusual pattern: the division resulted in two daughter cells that were separate from each other (Fig. 5.5). This interpretation is based on the fact that in a given well there were only a few fluorescing units, most of them single cells, and some consisting in clusters of cells that were in close proximity, but not physically connected. In Fig. 5.5a the development of such a cluster composed of three cells, one of which is dividing, is shown over the course of 3 days.

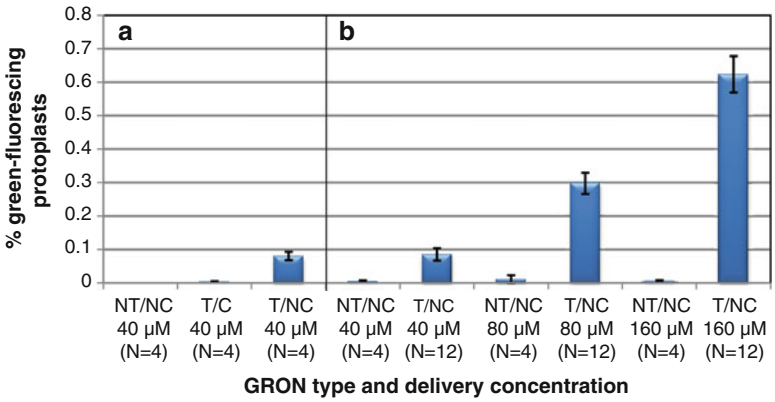

Fig. 5.4 Effect of GRON type and concentration on BFP to GFP conversion in *Arabidopsis* protoplasts containing eight copies of the BFP gene. The conversion frequency is the number of green-fluorescing protoplasts expressed as percentage of the protoplasts that were analyzed by flow cytometry 24 h after GRON introduction. (**a**) Targeting coding (T/C) and targeting noncoding (T/NC) were delivered into protoplasts. The non-targeting, noncoding GRON NC/NT was used as negative control. (**b**) Protoplasts were transfected with different concentrations of noncoding (NT) GRONs, either with the non-targeting (NT) or the targeting (T) version. The GRON concentration given in μM refers to the protoplast suspension after adding DNA and PEG

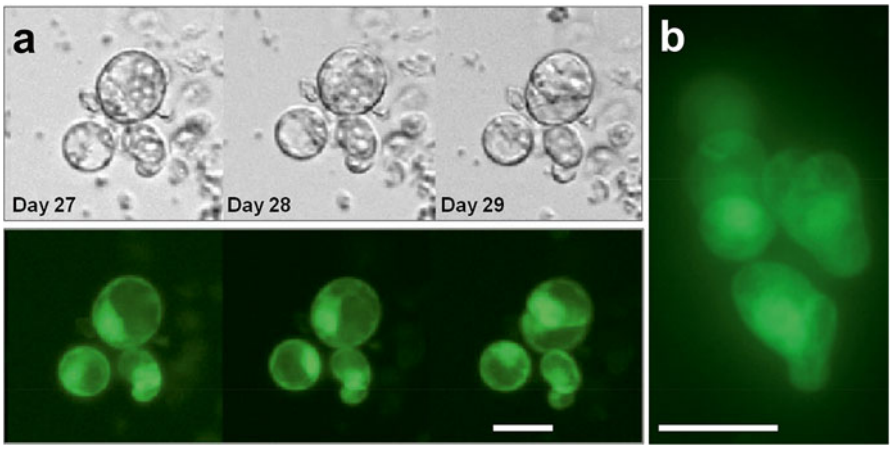

Fig. 5.5 Development of converted *Arabidopsis* cells. Protoplasts from a transgenic line containing eight copies of the BFP gene were treated with a GRON targeted at converting the BFP gene into a GFP gene. (**a**) Time lapse images of cells derived from a converted protoplast showing the division of a converted cell on day 29 after GRON delivery. *Upper panel*: images acquired in brightfield; *lower panel*: images acquired with a GFP filterset. (**b**) Colony of about six converted cells, 4 weeks after GRON delivery. Scale bars 10 μm

4 Discussion

Just like site-directed mutagenesis of plasmid DNA in vitro changed the course of molecular biology, in vivo gene targeting has been a goal that will forever change the course of biological research and product development. Concerted efforts at Cibus for more than a decade have advanced this technique in *Arabidopsis* as a model system and in OSR to develop the first commercial product using this technique. To our knowledge this is the first report of gene editing in OSR. Required elements to achieve this success were an understanding of the endogenous loci being targeted, conversion of these loci at a cellular level and regeneration of the converted cells into normal fertile plants.

Many peer reviewed publications have demonstrated conversion at a cellular level including some with show transient GRON-mediated conversion of plasmid templates in vitro and in plant cells (Rice et al. 2000; Kmiec et al. 2001; Dong et al. 2006).

A distinct advantage of traits developed using ***RTDS***, a directed mutagenesis technique, over those developed as transgenics is that the performance the trait is indistinguishable across lines (events) since the level and pattern of gene expression for the endogenous gene is retained in the convertant lines. The level of herbicide resistance of plants regenerated from calli with a specific targeted mutations detailed

in Fig. 5.3 was similar between lines with that mutation. As would be expected, the targeted mutations in these lines show stable Mendelian inheritance with that for our earliest product, to date demonstrating stability for more than 12 generations. To our knowledge, some of the specific mutations we obtained in the BnAHAS I and III targets in OSR were not previously obtained in chemical mutagenesis campaigns conducted by others over several decades. Together, we believe, the targeted and somatic mutations obtained in these experiments represent the largest collection of AHAS mutations in a single crop and parental background.

In tobacco a single clone and in rice two clones with the W574 (tryptophan) equivalent position successfully mutated to leucine were obtained (Kochevenko and Willmitzer 2003). Although the S653 (serine) homologous position (S627) was targeted in rice by Okuzaki and Toriyama (2004), no clones were obtained. By comparison, Zhu et al. (1999) obtained 13 independent events targeting the S653 homologous position (S621) in corn calli. The large number of control (no treatment, PEG alone, and/or non-targeting GRON) treatments as well as the depth of the 8 experiments targeting S653N (serine to asparagine) demonstrate the efficacy of ***RTDS***. This is further endorsed by our significant progress in improving conversion efficiency in our *Arabidopsis* model system, which has experienced improvements in excess of two orders of magnitude.

In the mammalian literature, transgenic cell lines containing one or multiple copies of mutated versions of the GFP gene are employed for the analysis of gene editing experiments because the phenotype of converted cells can be quantified by flow cytometry. The BFP gene in our transgenic *Arabidopsis* lines can be seen as a mutated GFP gene that has the advantage of potentially providing a phenotype for non-converted cells (Sommer et al. 2006). However, the background fluorescence in transgenic *Arabidopsis* protoplasts was too strong to permit the visualization of BFP fluorescence.

It is often difficult to compare BFP or mutated GFP conversion frequencies obtained by different authors. The efficiency of introduction of targeted changes through oligonucleotides measured within the first 2 days after transfection can depend on cell type, oligonucleotide concentration and chemistry, method of delivery, cell cycle status, copy number of the targeted gene, and many other factors. When flow cytometry is used for the quantification of converted cells, the type of flow cytometer (sensitivity) and the gating to distinguish green fluorescing cells from the background fluorescence can have an influence on the detectable number of converted cells.

There are few publications on the use of a BFP transgene to assess oligo-mediated gene targeting. Andrieu-Soler et al. (2005) compared the effect of different oligo chemistries and lengths on conversion frequency. Depending on the oligo used for transfection, chromosomal conversion rates varied from 0.007×10^{-3} to 2×10^{-3}. Sommer et al. (2006) describe the conversion of a chromosomal BFP gene to GFP. They evaluated conversion qualitatively by microscopy, but do not give information about the conversion efficiency per number of oligo-treated cells.

Wagner et al. (2010) used GRONs provided by Cibus (described in Sect. 2.2). The same GRON was used in our conversion experiments targeting BFP in transgenic *Arabidopsis* protoplasts as to convert their single copy BFP gene in a transgenic HEK cell line. In those experiments, Wagner et al. (2010) reported that neither the noncoding nor the coding version of the GRON resulted in conversion rates above background. However, in those experiments, no attempt was made to optimize the GRON concentration.

In the cited publications and through what is presented in this chapter, the successful application of the technology has been demonstrated in *Arabidopsis*, oil seed rape, corn, rice, tobacco, and wheat. To our knowledge, this is the first published report of gene targeting being performed in OSR and leading to a commercial outcome.

References

Aarts M, Dekker M, de Vries S, van der Wal A, te Riele H (2006) Generation of a mouse mutant by oligonucleotide-mediated gene modification in ES cells. Nucleic Acids Res 34:e147

Alexeev V, Yoon K (1998) Stable and inheritable changes in genotype and phenotype of albino melanocytes induced by an RNA-DNA oligonucleotide. Nat Biotechnol 16:1343–1346

Andrieu-Soler C, Casas M, Faussat AM, Gandolphe C, Doat M, Tempe D, Giovannangeli C, Behar-Cohen F, Concordet JP (2005) Stable transmission of targeted gene modification using single-stranded oligonucleotides with flanking LNAs. Nucleic Acids Res 33:3733–3742

Beetham PR, Kipp PB, Sawycky XL, Arntzen CJ, May GD (1999) A tool for functional plant genomics; chimeric RNA/DNA oligonucleotides cause *in vivo* gene-specific mutations. Proc Natl Acad Sci U S A 96:8774–8778

Breyer D, Herman P, Brandenburger A, Gheysen G, Remaut E, Soumillion P, Van Doorsselaere J, Custers R, Pauwels K, Sneyers M, Reheul D (2009) Genetic modification through oligonucleotide-mediated mutagenesis. A GMO regulatory challenge? Environ Biosafety Res 8:57–64

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16:735–743

Cole-Strauss A, Yoon K, Xiang Y, Byrne BC, Rice MC, Gryn J, Holloman WK, Kmiec EB (1996) Correction of the mutations responsible for sickle cell anemia by an RNA-DNA oligonucleotide. Science 273:1386–1389

Dong C, Beetham P, Vincent K, Sharp P (2006) Oligonucleotide-directed gene repair in wheat using a transient plasmid gene repair assay system. Plant Cell Rep 25:457–465

Dovzhenko A (2001) Towards plastid transformation in rapeseed (*Brassica napus* L.) and sugarbeet (*Beta vulgaris* L.). Ph.D. dissertation, LMU Munich, Faculty of Biology

Gamper HB, Parekh H, Rice MC, Youkey H, Kmiec EB (2000) The DNA strand of chimeric RNA/DNA oligonucleotides can direct gene repair/conversion activity in mammalian and plant cell-free extracts. Nucleic Acids Res 28:4332–4339

Guevara-García A, López-Bucio J, Herrera-Estrella L (1999) The mannopine synthase promoter contains vectorial cis-regulatory elements that act as enhancers and silencers. Mol Gen Genet 262:608–617

Kipp PB, Van Eck J, Beetham PR, May GD (2000) Gene targeting in plants via site-directed mutagenesis. In: Kmiec E (ed) Methods of molecular biology—gene targeting protocols. Humana, Totowa

Klaus S (2003) Markerfreie transplastome Tabakpflanzen (Marker-free transplastomic tobacco plants). Ph.D. dissertation, LMU Munich, Faculty of Biology

Kmiec EB, Johnson C, May GD (2001) Chloroplast lysates support directed mutagenesis via modified DNA and chimeric RNA/DNA oligonucleotides. Plant J 27:267–274

Kochevenko A, Willmitzer L (2003) Chimeric RNA/DNA oligonucleotide-based site-specific modification of the tobacco acetolactate synthase gene. Plant Physiol 132:174–184

Kren BT, Cole-Strauss A, Kmiec EB, Steer CJ (1997) Targeted nucleotide exchange in the alkaline phosphatase gene of HuH-7 cells mediated by chimeric RNA/DNA oligonucleotide. Hepatology 25:1462–1468

Kuzma J, Kokotovich A (2011) Renegotiating GM crop regulation. Targeted gene-modification technology raises new issues for the oversight of genetically modified crops. EMBO Rep 1:883–888

Liu L, Cheng S, van Brabant AJ, Kmiec EB (2002) Rad51p and Rad54p, but not Rad52p, elevate gene repair in *Saccharomyces cerevisiae* directed by modified single-stranded oligonucleotide vectors. Nucleic Acids Res 30:2742–2750

Maguire K, Kmiec EB (2007) Multiple roles for MSH2 in the repair of a deletion mutation directed by modified single-stranded oligonucleotides. Gene 15:107–114

Mathur J, Szabados L, Koncz C (1995) A simple method for isolation, liquid culture, transformation and regeneration of *Arabidopsis thaliana* protoplasts. Plant Cell Rep 14:221–226

Metz R, DiCola M, Kurihara T, Bailey A, Frank B, Roecklein B, Blaese M (2002) Mode of action of RNA/DNA oligonucleotides: progress in the development of gene repair as a therapy for alpha(1)-antitrypsin deficiency. Chest 121:91S–97S

Moerschell RP, Tsunasawa S, Sherman F (1988) Transformation of yeast with synthetic oligonucleotides. Proc Natl Acad Sci U S A 85:524–528

Morozov V, Wawrousek EF (2008) Single-strand DNA-mediated targeted mutagenesis of genomic DNA in early mouse embryos is stimulated by Rad51/54 and by Ku70/86 inhibition. Gene Ther 15:468–472

Negrutiu I, Shillito R, Potrykus I, Biasini G, Sala F (1987) Hybrid genes in the analysis of transformation conditions. I. Setting up a simple method for direct gene transfer in plant protoplasts. Plant Mol Biol 8:363–373

Oh TJ, May GD (2001) Oligonucleotide-directed plant gene targeting. Curr Opin Biotechnol 12:169–172

Okuzaki A, Toriyama K (2004) Chimeric RNA/DNA oligonucleotide-directed gene targeting in rice. Plant Cell Rep 22:509–512

Pelletier G, Primard C, Vedel F, Chetrit P, Remy R, Rouselle P, Renard M (1983) Intergeneric cytoplasm hybridization in *Cruciferae* by protoplast fusion. Mol Gen Genet 191:244–250

Pierce EA, Liu Q, Igoucheva O, Omarrudin R, Ma H, Diamond SL, Yoon K (2003) Oligonucleotide-directed single-base DNA alterations in mouse embryonic stem cells. Gene Ther 10:24–33

Rice MC, May GD, Kipp PB, Parekh H, Kmiec EB (2000) Genetic repair of mutations in plant cell-free extracts directed by specific chimeric oligonucleotides. Plant Physiol 123:427–438

Rice MC, Bruner M, Czymmek K, Kmiec EB (2001) *In vitro* and *in vivo* nucleotide exchange directed by chimeric RNA/DNA oligonucleotides in *Saccharomyces cerevisiae*. Mol Microbiol 40:857–868

Ruiter R, van den Brande I, Stals E, Delauré S, Cornelissen M, D'Halluin K (2003) Spontaneous mutation frequency in plants obscures the effect of chimeraplasty. Plant Mol Biol 53:675–689

Rutledge RG, Quellet T, Hattori J, Miki BL (1991) Molecular characterization and genetic origin of the *Brassica napus* acetohydroxyacid synthase multigene family. Mol Gen Genet 229: 31–40

Sommer JR, Alderson J, Laible G, Petters RM (2006) Reporter system for the detection of in vivo gene conversion. Mol Biotechnol 33:115–121

Tan S, Evans RR, Dahmer ML, Singh BK, Shaner DL (2005) Imidazolinone-tolerant crops: history, current status and future. Pest Manag Sci 61:246–257

Wagner S, McCracken J, Cole S, Laible G (2010) DNA oligonucleotides and plasmids perform equally as donors for targeted gene conversion. Biochem Genet 48:897–908

Yoon K, Cole-Strauss A, Kmiec EB (1996) Targeted gene correction of episomal DNA in mammalian cells mediated by a chimeric RNA.DNA oligonucleotide. Proc Natl Acad Sci U S A 93:2071–2076

Zhu T, Peterson DJ, Tagliani L, St. Clair G, Baszczynski CL, Bowen B (1999) Targeted manipulation of maize *in vivo* using chimeric RNA/DNA oligonucleotides. Proc Natl Acad Sci U S A 96:8768–8773

Zhu T, Mettenburg K, Peterson DJ, Tagliani L, Baszczynski CL (2000) Engineering herbicide-resistant maize using chimeric RNA/DNA oligonucleotides. Nat Biotechnol 18:555–558

Chapter 6
Gene Targeting in Crop Species with Effective Selection Systems

Takaki Yamauchi and Shigeru Iida

Abstract Gene targeting refers to the alteration of a specific DNA sequence in an endogenous gene at its original locus in the genome by homologous recombination and provides a powerful tool for both the functional analysis of the gene of interest and the molecular breeding of crop plants. In this chapter, we focus on gene targeting in crop plants with effective selection systems, namely, gene-specific selection and positive–negative selection. So far japonica rice is the only crop species in which the reproducible gene targeting of the endogenous genes in the nuclear genomes has been reported: two genes by gene-specific selection and at least nine genes by positive–negative selection. Gene-specific selection and positive–negative selection are applicable to only a limited number of genes and, in principle, any genes, respectively. We discuss some characteristic features and possible future developments associated with gene targeting of an endogenous gene with positive–negative selection in crop species.

Abbreviations

BARI	Border-associated random integration
BIRI	Border-independent random integration
DSB	Double-strand break
EGT	Ectopic gene targeting
GT	Gene targeting
HR	Homologous recombination
NHEJ	Nonhomologous end joining
OSI	One-sided invasion

T. Yamauchi, Ph.D. (✉)
Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa, Nagoya 464-8601, Japan
e-mail: atkyama@gmail.com

S. Iida, Ph.D.
Graduate School of Nutritional and Environmental Sciences, University of Shizuoka, Yada 52-1, Suruga-ku 422-8526, Japan
e-mail: shigiid@u-shizuoka-ken.ac.jp

F. Zhang et al. (eds.), *Advances in New Technology for Targeted Modification of Plant Genomes*, DOI 10.1007/978-1-4939-2556-8_6

TGT True gene targeting
Trp Tryptophan
ZFN Zinc-finger nuclease

1 Introduction

More than a quarter of a century has passed since the first report of successful gene targeting (GT) of an artificially truncated drug-resistance gene integrated in the tobacco genome (Paszkowski et al. 1988). GT refers to the alteration of a specific

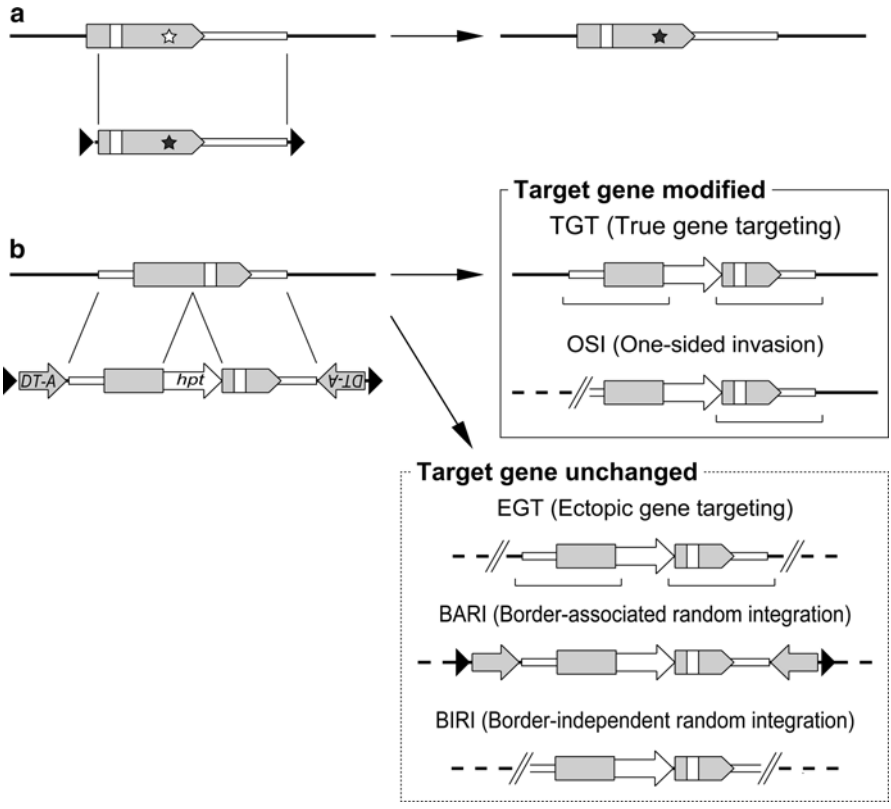

Fig. 6.1 Gene targeting with effective selection systems. (**a**) Gene targeting with gene-specific selection. The introduced point mutation confers a selectable phenotype. The wild-type and mutated nucleotides are indicated by *open* and *filled pentacles*, respectively. (**b**) Gene targeting with positive–negative selection and ectopic transgene integrations. *Open rectangles* and *filled arrowheads* indicate homologous regions carried by the introduced T-DNA segments and their border sequences, respectively. *Shaded* and *open areas* within the target gene represent exons and introns, respectively. The *hpt* and *DT-A* genes are used as positive and negative markers, respectively. The anticipated TGT is thought to occur via double crossovers at flanking homologous regions, and brackets under the maps indicate junction fragments generated by crossovers and used for screening HR events by PCR. The most efficient mode of transgene integration in *Agrobacterium*-mediated transformation is BARI, but the resulting transformants are to be killed by the expressed DT-A proteins

DNA sequence in an endogenous gene at its original locus in the genome by homologous recombination (HR) (Iida and Terada 2005; Da Ines and White 2013). GT provides a powerful tool for both the functional analysis of the gene of interest and the molecular breeding of crop plants, because it enables us in principle to convert an endogenous genomic sequence to desired sequences exactly as expected. In this chapter, we first briefly outline general backgrounds on GT and then focus on GT in crop plants with effective selection systems, namely, gene-specific selection and positive–negative selection (Fig. 6.1). Therefore, GT in model plants such as Arabidopsis and tobacco is briefly discussed only in the context of GT in crop plants. While gene-specific selection was successfully applied to the targeting of the endogenous 16S rRNA (*rrn 16*) gene in the plastid genome (Maliga 2004; Day and Goldschmidt-Clermont 2011), this review deals with the reproducible GT of endogenous natural genes in the nuclear genome. We also discuss possible future developments associated with GT of an endogenous gene with positive–negative selection in crop species.

2 General Background on Gene Targeting

The integration of an introduced transgene by somatic HR to yield true gene targeting (TGT) with the expected sequence alterations occurs in the order of 10^{-3} to 10^{-6} compared with random integration, and the overwhelming occurrence of random integration of transgenes by nonhomologous end joining (NHEJ) relative to targeted HR is the main obstacle in the development of an efficient GT system for flowering plants (Vergunst and Hooykaas 1999; Iida and Terada 2005; Da Ines and White 2013). In addition to random integration, aberrant recombination events associated with GT, called one-sided invasion (OSI) and ectopic gene targeting (EGT), have occasionally been detected (Fig. 6.1; Iida and Terada 2005). OSI results from one crossover by HR and another crossover by NHEJ at the target locus, whereas EGT is thought to be generated by ectopic integration (integration elsewhere in the genome without altering the target gene) of a recombinant molecule produced by HR between the introduced transgene and a copy of the target sequence. Therefore, an ectopically targeted locus can also be flanked by the expected sequences to a certain length and can be transmitted into progeny in a Mendelian fashion.

To circumvent such obstacles, various approaches have been developed to enhance GT efficiency and they include (1) induction of double-strand breaks (DSBs) at the target locus in the genome by meganucleases and sequence-specific engineered endonucleases, namely, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas)-based RNA-guided DNA endonucleases (Puchta et al. 1996; Belhaj et al. 2013; Marton et al. 2013; Mussolino and Cathomen 2013; Voytas 2013; Puchta and Fauser 2014; see also Chaps. 2, 3, 4, and 9); (2) engineering of the host recombination and/or repair

machinery (Lieberman-Lazarovich and Levy 2011; Waterworth et al. 2011; Da Ines and White 2013), or expressing of foreign genes involving HR (e.g., Shaked et al. 2005), for enhancing HR processes or for suppressing NHEJ processes; and (3) efficient enrichment by effective selection systems, namely, gene-specific selection and positive–negative selection (Table 6.1; Endo et al. 2007; Terada et al. 2002).

DSBs, which have extremely detrimental effects on plant growth, can be repaired by HR or NHEJ leading to GT by replacing the native genomic sequences with incoming DNA sequences or targeted mutagenesis by rejoining imprecisely at the break sites, respectively (Waterworth et al. 2011; Puchta and Fauser 2014). Integration of transgenes is also mediated predominantly by NHEJ; therefore, their integration sites are generally random in the genome. Thus, induction of DSBs can stimulate the HR-associated repair machinery. While the sequence-specific engineered endonucleases-promoted genome editing, GT and targeted mutagenesis, is a rapidly growing field, most published studies report on targeted mutagenesis (Belhaj et al. 2013; Marton et al. 2013) because NHEJ is the dominant DNA repair pathway in plants and because it is not necessary to appropriately provide a homologous segment to be recombined. To date, there appear to be only a few reports on the engineered endonuclease-promoted GT of an endogenous gene in crop species; for homologous donor molecules to be recombined, plasmids containing a selectable marker were used to obtain targeted transgenic maize plants (Shukla et al. 2009), whereas a single-stranded oligo containing restriction sites was used to isolate rice recombinants (Shan et al. 2013).

Although the engineered endonucleases-promoted genome editing is, in principle, applicable to diverse plant species including crop plants, one of the impediments to successful genome editing is off-target effects caused by these engineered endonucleases, which promote DSBs at secondary recognition sites in the genome, thereby leading to unwanted sequence alterations including DNA rearrangements and cytotoxicity (Voytas 2013; Puchta and Fauser 2014). Since genomic sequences are likely to contain such secondary recognition sites (e.g., Thyagarajan et al. 2000; see also Iida and Hiestand-Nauer 1987) and the probability of the presence of the secondary recognition sites depends upon genome size, genome editing induced by DSBs is intrinsically associated with off-target effects in crop plants bearing large genome sizes. Even in homodimeric meganucleases, which seem to be less toxic than these engineered endonucleases, they can still be harmful at very high doses (Gouble et al. 2006). Not only the induction of DSBs but also modulation of the recombination and/or repair machinery might hamper the functional characterization of the endogenous gene of interest as well as the molecular breeding of crop plants, because alteration of the recombination and/or repair machinery may result in unexpected or pleiotropic phenotypes. For example, suppression of Ku70 involved in NHEJ processes in rice led to increased HR frequency, and its null mutants exhibited severely retarded growth phenotypes (Nishizawa-Yokoi et al. 2012). In this respect, the approaches to enrich gene-targeted recombinants by effective selection systems appear to be less problematic and more straightforward than induction of DSBs or modulation of the recombination and/or

Table 6.1 Fertile transgenic rice plants having endogenous genes targeted by homologous recombination

Strategy	Gene	Gene delivery system	Selection and screening	Homologous sequences	Regenerants	Frequency (TGT)	Fresh weight calli (g) used	Transmission	References
Gene-specific selection									
	ALS (W548L/S627I)	*Agrobacterium*	GSS-PCR	8.3 kb	66 Independent TGT plants	66 TGT/72 BS^r calli (9.2×10^{-1})	~30	Mendelian inheritance	Endo et al. (2007)
	OASA2 (S126F/L530D)	*Agrobacterium*	GSS-CAPS	7.0 kb	2 Independent TGT plants	2 TGT/2 $5MT^r$ calli (1.0×10^{-0})	20.4	Mendelian inheritance	Saika et al. (2011)
	OASA2 (Y367A/L530D)	*Agrobacterium*	GSS-CAPS	7.0 kb	14 Independent TGT plants	14 TGT/14 $5MT^r$ calli (1.0×10^{-0})	55.6	Mendelian inheritance	Saika et al. (2011)
Positive–negative selection									
Knockout targeting									
	Waxy	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–6.3 kb, 3′–6.8 kb	6 Independent TGT plants	6 TGT/638 Hm^r calli (9.4×10^{-3})	194	Mendelian inheritance	Terada et al. (2002)
	Adh2	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–6.2 kb, 3′–6.0 kb	9 Independent TGT plants	9 TGT/468 Hm^r calli (1.9×10^{-2})	204	Mendelian inheritance	Terada et al. (2007)
	Waxy (liquid medium)	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–6.8 kb, 3′–5.7 kb	3 Independent TGT plants	3 TGT/202 Hm^r calli (1.5×10^{-2})	12	Mendelian inheritance	Ozawa et al. (2012)
	Xyl (liquid medium)	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–5.5 kb, 3′–5.6 kb	5 Independent TGT plants	5 TGT/276 Hm^r (1.8×10^{-2})	16	Mendelian inheritance	Ozawa et al. (2012)

(continued)

Table 6.1 (continued)

Strategy	Gene	Gene delivery system	Selection and screening	Homologous sequences	Regenerants	Frequency (TGT)	Fresh weight calli (g) used	Transmission	References
Knockin targeting									
	MET1a	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–3.1 kb, 3′–3.1 kb	15 Independent TGT plants	15 TGT/284 Hm^r calli (5.3×10^{-2})	107	Mendelian inheritance	Yamauchi et al. (2009)
	DRM2	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–3.0 kb, 3′–3.1 kb	7 Independent TGT plants	7 TGT/366 Hm^r calli (1.9×10^{-2})	73	Transmittable	Moritoh et al. (2012)
	DRM2-Del	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–3.0 kb, 3′–3.0 kb (3 kb internal deletion)	2 Independent TGT plants	2 TGT/504 Hm^r calli (4.0×10^{-3})	57	No information	Moritoh et al. (2012)
	ROS1a	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–3.0 kb, 3′–3.0 kb	6 Independent TGT plants	6 TGT (1.1×10^{-2})	No information	Not transmittable	Ono et al. (2012)
	MET1b	*Agrobacterium*	PNS-PCR [*HPT*(Hmr)/DT-A]	5′–3.6 kb, 3′–1.8 kb	6 Independent TGT plants	6 TGT/680 Hm^r calli (8.8×10^{-3})	290	Mendelian inheritance	Yamauchi et al. (2014)

Subtle mutagenesis									
	IRE1 [K519A][a] (liquid medium)	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–7.0 kb, 3′–4.5 kb	4 Independent TGT plants	4 TGT (0.7×10^{-2})	No information	Homozygous mutant lethal	Wakasa et al. (2012)
	IRE1 [K833A][a] (liquid medium)	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–7.0 kb, 3′–4.5 kb	2 Independent TGT plants	2 TGT (2.2×10^{-2})	No information	Mendelian inheritance	Wakasa et al. (2012)
	Rac1[b]	*Agrobacterium*	PNS-PCR [*HPT*(Hm^r)/*DT-A*]	5′–3.0 kb, 3′–3.0 kb	5 Independent TGT plants	5 TGT/94 Hm^r calli (5.3×10^{-2})	No information	Mendelian inheritance	Thi Dang et al. (2013)

Abbreviations: *GSS-PCR* gene-specific selection followed by PCR screening; *GSS-CAPS* gene-specific selection followed by screening with cleaved polymorphic sequence (CAPS) markers; *PNS-PCR* positive–negative selection followed by PCR screening; *TGT* true gene targeting; *BS^r* bispyribac resistance; *5MT^r* 5-methyl-Trp analogue resistance; *Hm^r* hygromycin resistance

Reportes published as of December 2013

[a]The positive selection marker *hpt* gene is present in the genome

[b]The positive selection marker *hpt* gene was removed by the site-specific Cre recombinase

repair machinery, because no materials other than homologous segments to be recombined with target sequences are necessary to be introduced into the plant cells, although an efficient and large-scale transformation procedure needs to be developed (see below).

3 Gene Targeting with Effective Selection Systems

Gene-specific selection is based on the facts that certain point mutations in a limited number of endogenous genes can confer resistance to certain chemicals (e.g., herbicides), whereas positive–negative selection is, in principle, applicable to any endogenous gene (Fig. 6.1). It is thus absolutely necessary to know the phenotypes beforehand in the case of predesigned mutations obtained by gene-specific selection, whereas no such condition is required for GT with positive–negative selection, although genes with predictable mutant phenotypes, such as *Waxy*, *Xyl*, *IRE1*, and *Rac1*, were often chosen (Table 6.1). Indeed, the first endogenous gene successfully targeted by gene-specific selection in crop species was the rice *ALS* gene encoding acetolactate synthase, in which introduction of appropriate point mutations conferred an herbicide-tolerant phenotype (Endo et al. 2007), whereas the first endogenous gene successfully targeted by positive–negative selection in crop species was the rice *Waxy* gene for granule-bound starch synthase, controlling an agronomically important trait, and its null mutants can be easily detected by iodine staining (Terada et al. 2002). It appears that japonica rice is so far the only crop species in which the reproducible GT of the endogenous genes in the nuclear genomes has been succeeded by either gene-specific selection or positive–negative selection; several independently isolated fertile transgenic plants carry the exact anticipated sequence alterations in the targeted gene (Table 6.1).

Rice, with its small genome of ~389 Mb, is an important staple food for more than half the world's population and its functional genomics is being vigorously pursued (Upadhyaya 2007; Wang et al. 2013). The success of GT in rice stems partly from the development of efficient *Agrobacterium*-mediated large-scale transformation procedures (Terada et al. 2004, 2007; Toki et al. 2006; Ozawa et al. 2012). Calli derived from mature rice seeds were used for transformation, and such in vitro culture may concomitantly induce genetic and epigenetic changes termed somaclonal variation that might hamper the functional characterization of the gene to be targeted (Kaeppler et al. 2000; Jain 2001). To minimize somaclonal variation, one of the procedures is to shorten the duration of the tissue culture (Toki et al. 2006). Another possible approach to explore efficient GT in cereal crops such as japonica and indica rice as well as maize is to infect immature embryos directly with *Agrobacterium* for efficient transformation, which may shorten the transformation time (Hiei and Komari 2008). Since the *Agrobacterium*-mediated floral-dip transformation that does not require tissue culture is a routine procedure in Arabidopsis (Clough and Bent 1998), *in planta* transformation of cereals has the potential to be developed to suppress somaclonal variation (e.g., Zale et al. 2009).

4 Gene Targeting with Gene-Specific Selection

Since only a limited number of endogenous genes can confer acquired resistance to certain chemicals by introducing appropriate point mutations (Fig. 6.1a), GT by gene-specific selection has been applied to only a few endogenous genes (Iida and Terada 2005; Da Ines and White 2013). The first reproducible GT of an endogenous gene by gene-specific selection was the *PPO* gene encoding protoporphyrinogen oxidase in Arabidopsis, and appropriate base substitutions in *PPO* conferred herbicide resistance (Hanin et al. 2001). For crop species, fertile targeted rice plants carrying anticipated base substitutions in two genes, *ALS* and *OASA2*, encoding acetolactate synthase and α-subunit of anthranilate synthase were successfully obtained by gene-specific selection with a bispyribac herbicide and a tryptophan (Trp) analogue 5-methyl-Trp, respectively (Table 6.1; Endo et al. 2007; Saika et al. 2011). The strategy for gene-specific direct selection is efficient and straightforward, and the targeting frequencies of *ALS* and *OASA2* were reported to be more than 90 % TGT events per selected calli. Because anthranilate synthase (AS) plays a key role in the biosynthesis of Trp, and because it is susceptible to feedback inhibition by Trp or its analogue, plants carrying designed sequence alterations produced seeds with higher Trp content due to insensitivity of the feedback inhibition and higher AS activity (Saika et al. 2011), thereby potentially contributing to rice breeding.

5 Gene Targeting with Positive–Negative Selection

The vectors to be introduced for all successful GT events with positive–negative selection in rice have been essentially the same gene constellation as originally developed for targeting of the *Waxy* gene; in the T-DNA region, a positive-selection marker is placed within the sequence homologous to the target locus and each of two copies of the same negative-selection marker is located outside of the homologous region in inverted orientation in order to avoid the excision of the homologous segment from the T-DNA segment (Fig. 6.1b; Table 6.1; Terada et al. 2002). The positive marker used is the *hpt* gene for hygromycin resistance, because it is more effective than the *nptII* gene for kanamycin resistance in rice, although *nptII* can be effectively used in rice for paromomycin resistance (Komari et al. 1999). A similar gene constellation of the vectors with positive–negative selection was originally designed for targeting Lotus genes without succeeding in identifying targeted events; *nptII* is placed within the sequence homologous to the target locus and the entire homologous region is flanked by two *codA* genes in direct orientation (Thykjær et al. 1997). The *codA* gene for the *E. coli* cytosine deaminase, which catalyzes the conversion of nontoxic 5-fluorocytosine to toxic 5-fluorouracil, has been developed as a substrate-dependent negative-selection marker, but it often does not appear to be stringent enough for negative selection (Terada et al. 2004),

probably because 5-fluorocytosine does not seem to efficiently penetrate the core region of calli and consequently a significant number of escapees appear. Thus, the success of positive–negative selection is partly attributable to utilization of the *DT-A* gene encoding the diphtheria toxin A fragment, which inhibits eukaryotic protein biosynthesis without containing its B fragment, which is required for the recognition of membrane receptors and for the import of the toxin A fragment into the cell. Therefore, the DT-A protein cannot be taken up by adjacent cells, and this cell-autonomous and unconditional mode of killing is suitable not only for genetic ablation (Day and Irish 1997) and but also for negative selection (Terada et al. 2002; Iida and Terada 2004, 2005).

The majority of the randomly integrated single-copy T-DNA molecules in *Agrobacterium*-mediated transformation are known to contain the entire T-DNA segment with a well-conserved right border and a left-border sequence that is either conserved or truncated, which is collectively termed border-associated random integration (BARI), and two inversely oriented *DT-A* genes placed at both ends of the T-DNA segment adjacent to its border sequences are intended to efficiently eliminate the overwhelming BARI events (Fig. 6.1b; Terada et al. 2007). As expected, the majority of the surviving calli with positive–negative selection have the truncated segments containing the active *hpt* gene integrated into the genome by NHEJ processes independent of BARI, which is termed border-independent random integration (BIRI). The estimated frequency of BIRI compared with BARI in the targeting of the *Waxy* gene was 6.9×10^{-2} (Terada et al. 2002), indicating that positive–negative selection resulted in a nearly 15-fold enrichment of gene-targeted calli. Subsequently, the surviving calli were subjected to PCR analysis to detect two junction fragments generated by HR between the homologous segments in the introduced vector and the endogenous target gene. End sequencing of candidate junction fragments, which appeared to be the anticipated size, was then performed to unambiguously exclude false-positive junction fragments that resulted from OSI and other ectopic recombination evens. We noticed that it was difficult to amplify the long junction fragment in the *Waxy* promoter region (Terada et al. 2002), and there may be other regions in the genome that are not conducive to PCR amplification. Except for such cases, PCR analysis for detecting both junction fragments combined with end sequencing can effectively eliminate the OSI events, even if they have occurred (Fig. 6.1). Subsequent Southern-blot analysis revealed whether the transformed calli or their regenerated T0 plants were heterozygotes with TGT alleles without containing an ectopically inserted additional transgene. The transgenic plants carrying the TGT allele in the heterozygous condition should contain one copy each of the truly targeted and untargeted wild-type alleles. If EGT occurred in the heterozygous condition, however, the copy numbers of the ectopically targeted and wild-type alleles should be one and two, respectively. In case the integrated locus of an ectopically targeted allele is unlinked to the targeted locus, T1 segregants homozygous for the EGT allele in the selfed progeny clearly show that two copies of the gene at the target locus remain intact. Such a feature of EGT is quite different from that of TGT, in which T1 segregants homozygous for the TGT

allele bear two copies of a properly modified gene at the target locus (e.g., Yamauchi et al. 2009). It should be emphasized that such extensive characterization of the targeted gene structure in the genome could be performed independent of the function of the target gene. In other words, the confirmation of TGT can be accomplished without knowing the function of the target gene in advance.

Although we repeatedly performed several transformation experiments in our first attempt to target the *Waxy* gene (Terada et al. 2002), a single transformation is generally sufficient to obtain several independent calli bearing the TGT alleles by employing an improved routine transformation procedure (Terada et al. 2007; Yamauchi et al. 2009). For the PCR analysis, it is feasible for one operator to handle 192 calli in two 96-well plates in a single day. It has been reported recently that efficient *Agrobacterium*-mediated GT using rice suspension-cultured calli resulted in a more than fivefold increase in the number of surviving calli per weight of the starting material after positive–negative selection, as compared with the conventional transformation using calli on a solid medium (Ozawa et al. 2012). Not only the gene constellation but also the positive- and negative-selection markers used in their vectors (Ozawa et al. 2012; Wakasa et al. 2012) were essentially the same as those in our vectors (Terada et al. 2002, 2007; Yamauchi et al. 2009; Ono et al. 2012), and the observed GT frequencies in their system, as determined by TGT events per surviving callus with positive–negative selection, were comparable to those in our GT system (Table 6.1). Therefore, it remains to be seen whether the observations that the surviving calli grown in a suspension culture are more frequent than those grown on a solid medium are due to the increased frequency of BIRI in particular or due to that of NHEJ in general. Based on our observations of rice GT of *Waxy*, *Adh2*, *MET1a*, *MET1b*, and *ROS1a* with positive–negative selection (Table 6.1), the following general conclusions can be drawn: (1) targeted calli or their regenerated T0 plants are heterozygotes with TGT alleles without containing an ectopically inserted additional copy of transgene; (2) the observed GT frequencies are roughly 1–5 %; (3) no ectopic events including OSI and EGT have been observed; (4) a targeted null allele that is not transmittable to progeny is repeatedly obtained; and (5) single base changes within the homologous segments in the vector are efficiently transferred into the corresponding genomic sequences at the targeted locus. Not only *DRM2* (Moritoh et al. 2012) but also two rice *DDM1* genes, *DDM1a* and *DDM1b*, encoding SWI2/SNF chromatin remodeling proteins in GT were found to meet criteria (1) to (3) (Terada et al. 2007; Johzuka-Hisatomi et al. 2008; Ono et al. 2012).

Apart from flowering plants including crop species, the moss *Physcomitrella patens* is well known to exhibit high frequencies of HR-dependent GT (Schaefer 2002; Strotbek et al. 2013) although both the transformation methods and the terminology of GT in moss appear to be slightly different from that in rice and other flowering plants (e.g., see Terada et al. 2007). In this respect, it may be worth noting here that HR-mediated GT with positive–negative selection using an efficient *Agrobacterium*-mediated transformation system has been successfully demonstrated in the liverwort *Marchantia polymorpha* L. with a frequency of ~2 % of the surviving thalli (Ishizaki et al. 2013).

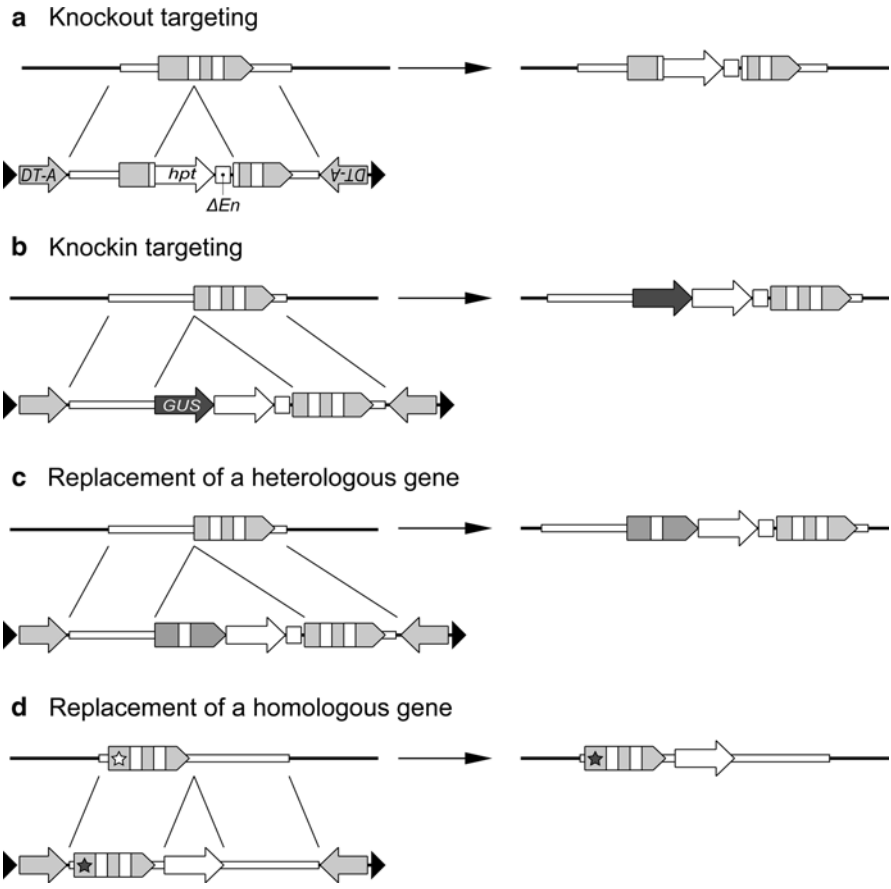

Fig. 6.2 Various strategies for gene targeting. (**a**) Knockout targeting. The target gene is disrupted by the *hpt* gene. (**b**) Knockin targeting. The promoterless *GUS* reporter gene is fused with the promoter of the target gene. (**c**) Replacement of a heterologous gene. The target gene is replaced by a heterologous gene. (**d**) Replacement of a homologous gene. The target gene is replaced with its base changed allele. The strategy was used to introduce point mutations in the *IRE1* gene (Table 6.1). The *ΔEn* element is used for efficient transcriptional termination, but it may not be a prerequisite depending on circumstances (**d**). The elimination of the *hpt* and its regulatory elements from the targeted locus leads to the gene activation in knockout targeting or the desirable modifications in other targeting (see Fig. 6.3). Other *symbols* are as in Fig. 6.1

5.1 Knockout and Knockin Targeting

Knockout targeting is the ablation of the function of a target gene by introducing a foreign DNA segment, usually a selection marker, thereby generating a disrupted null allele (Fig. 6.2a). On the other hand, knockin targeting is the placing of a foreign gene of interest or a reporter gene under the control of the *cis*-acting regulatory elements, including the promoter of a target gene, thereby also generating another

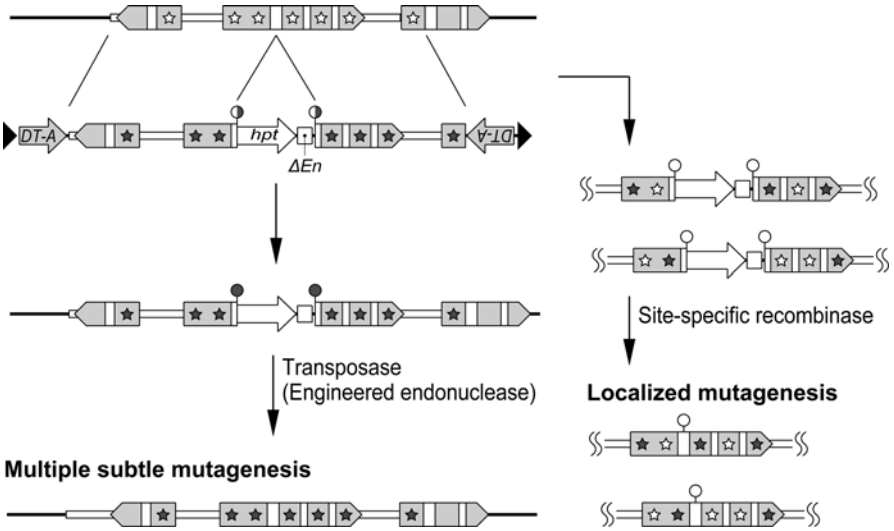

Fig. 6.3 Subtle and localized mutageneses with the elimination of a positive-selection marker. Multiple base changes scattered throughout the homologous segments on a vector can be transferred efficiently to the rice genome and often may be intermingled with the corresponding plant genome sequences, probably due to mismatch correction of heteroduplex intermediates formed. The elimination of a positive-selection marker from the targeted locus results in subtle mutagenesis and localized mutagenesis. The *hpt* gene represents a positive-selection marker. Site-specific recombinase and transposase indicate the Cre recombinase and the *piggyBac* transposase, respectively. *Open* and *filled lollipops* indicate the Cre recognition sequence *loxP* and the *piggyBac* terminal inverted repeat, respectively. Alternatively, an engineered endonuclease can be used for eliminating the positive-selection marker flanked by the directly repeated endogenous sequences, instead of the *piggyBac* terminal inverted repeats, although sequence alterations may occur at the excision site. Other *symbols* are as in Fig. 6.1

disrupted null allele (Fig. 6.2b). By combining HR-mediated GT with a site-specific recombinase system (e.g., Cre-*loxP*), strategies for conditional activation (gain of function) or inactivation (loss of function) of gene expression in vivo have been developed to control spatiotemporal or tissue-specific expression of target genes in mice (Fig. 6.3; Lewandoski 2001; Sorrell and Kolb 2005). Bearing a possible adaptation of such strategies to plants in mind, *hpt* flanked by two *loxP* sequences was designed to insert into an intron of the *Waxy* gene or the 5′-untranslated region of the *Adh2* gene in the knockout targeting of endogenous genes in rice (Terada et al. 2002, 2007). To ensure efficient termination of the transcriptional read-through, the *hpt* gene used contains the 3′ terminal part of the maize *En* element, *ΔEn*, and its effectiveness was subsequently confirmed (Yamauchi et al. 2009).

While the *Waxy* gene is unique on chromosome 6 in the rice genome, the *Adh3*, *Adh2*, and *Adh1* genes all comprise 10 exons with identical intron positions and are clustered in the same orientation on chromosome 11, and highly repetitive *Copia-like* and *Gypsy*-like elements are present 1.0 kb downstream of *Adh2* and less than 1.0 kb upstream of *Adh1*, respectively (Terada et al. 2007). Even though the *Adh*

genes, located immediately adjacent to highly repetitive elements, comprise a small multigene family and the introduced targeting vector contains the entire *Adh2* and part of the *Copia* sequence, no apparent preclusion by the presence of homologous and repetitive sequences was observed. GT resulted in the *waxy*::*hpt* and *adh2*::*hpt* (*waxy* and *adh2* disrupted by the *hpt* integration) null alleles conferring the anticipated *waxy* phenotype and no apparent phenotypic alterations, respectively (Terada et al. 2002, 2007), and Cre-promoted excision of *hpt* led to reactivation of the *Waxy* gene (Terada et al. 2010). Subsequently, knockout targeting of the single-copy genes, *Waxy* and *Xyl* (encoding β1,2-xylosyltransferase involved in plant-specific *N*-glycosylation), using rice suspension-cultured calli has recently been reported (Ozawa et al. 2012). In these cases, the disruption of the target genes was confirmed by the functional deficiencies of these genes: *waxy* by iodine staining and *xyl* by analyzing *N*-glycans with mass spectrometry.

So far only the promoterless *GUS* reporter gene encoding β-glucuronidase has been used for knockin targeting, in which it has been fused with the endogenous promoters of the following target genes: *MET1a* and *MET1b* for maintenance DNA methyltransferases, *DRM2* for de novo DNA methyltransferase, and *ROS1a* for 5-methyl cytosine DNA demethylase in rice (Yamauchi et al. 2009, 2014; Moritoh et al. 2012; Ono et al. 2012). It should be emphasized that the null mutant phenotypes of these genes were all unknown in advance. The transgenic rice plants having the disrupted null mutations in *MET1a*, *MET1b*, *DRM2*, and *ROS1a* were found to exhibit no overt phenotypes, abnormal seed phenotypes associated with viviparous germination or early embryonic lethality, pleiotropic developmental phenotypes, and no transmission of the disrupted null allele to progeny, respectively. Although the *ros1a* null mutation, being not transmittable to progeny, is impossible to isolate by conventional mutagenesis techniques that are used to identify and isolate mutants in the progeny population (Upadhyaya 2007; Wang et al. 2013), several independent mutants carrying an identical structure of the predesigned targeted knockin allele in the heterozygous condition were reproducibly obtained (Ono et al. 2012). Moreover, independently isolated mutants bearing the knockin-targeted *met1a-GUS* and *met1b-GUS* alleles showed the reproducible, dosage-dependent, and spatiotemporal *GUS* expression patterns (Yamauchi et al. 2009, 2014), and independently isolated *ros1a-GUS* mutants in the heterozygous condition also displayed the reproducible and spatiotemporal *GUS* expression patterns (Ono et al. 2012). The level of *ROS1a* transcripts that accumulated in the independently obtained heterozygous lines was approximately half that in the wild-type plants, suggesting that *ROS1a* is also expressed in a dosage-dependent manner. The spatiotemporal expression patterns of *GUS* generally coincided with their expression patterns indicated by RT-PCR analysis and with the in situ hybridization patterns of the corresponding wild-type transcripts in wild-type rice. These characteristic features of the *GUS* expression patterns in the knockin-targeted plants are in sharp contrast to the interindividual variation of transgene expression observed among independently obtained plants bearing the randomly integrated transgenes (Butaye et al. 2005; Yamauchi et al. 2009).

Although various genes may be expressed in a dosage-dependent manner, no apparent incomplete dominance has been detected in the heterozygous conditions, according to Mendel's first law (e.g., see Watson et al. 2008). This apparent discrepancy can be explained by assuming that these ordinary genes are generally expressed more than twice as much as the necessary levels in the wild type. Consistent to this notion, dosage-dependent expression of a certain gene is the primary cause of the observed incomplete dominance (Johzuka-Hisatomi et al. 2011). Because the knockin-targeting strategy allows detection of spatiotemporal expression of the target gene, utilization of other more sensitive markers such as green fluorescent protein, in combination with other available techniques including fluorescence-activated cell sorting, would facilitate further detailed characterization of the target gene function through in vivo analysis and/or isolation of viable tissues expressing the target gene (Ono et al. 2012). It is worth noting that the introduction of a 3-kb internal deletion caused a reduced targeting frequency of *DRM2* (Table 6.1). Similar reduced targeting frequencies associated with introducing internal deletions were observed in the liverwort *Marchantia polymorpha* L. (Ishizaki et al. 2013). A strategy for the replacement of a heterologous gene or sequence (e.g., substitution of the endogenous petunia *DFR* gene with the maize *DFR* gene for flower pigmentation; see Meyer et al. 1987) may closely resemble that for knockin targeting with an internal deletion(s) adjacent to the positive-selection marker (Fig. 6.2c), followed by elimination of the positive-selection marker (see Fig. 6.3).

5.2 *Toward the Development of Desired Subtle and Localized Mutageneses*

Not only the knockout- or knockin-targeted null mutations but also subtle mutations such as point mutations or small deletions and insertions provide valuable information about the function of the target gene, e.g., particular amino acid residues within a functional domain of a protein or specific sites within *cis*-regulatory elements (Sorrell and Kolb 2005; Menke 2013). The first such attempt using GT with positive–negative selection was to introduce two different predesigned missense mutations separately into either a kinase domain or a ribonuclease domain of the *IRE1* gene involved in the unfolded protein response (Table 6.1; Wakasa et al. 2012). Instead of *hpt* being placed within the target gene in the knockout or knockin targeting (Fig. 6.2a, b), it was placed adjacent to the target gene (Fig. 6.2d). Subsequently, a single base substitution was introduced in the *Rac1* gene encoding a Rac/Rop GTPase by gene targeting with positive–negative selection (Table 6.1; Thi Dang et al. 2013), followed by elimination of *hpt* from a *Rac1* intron by site-specific Cre recombinase (Fig. 6.3; see below).

An important observation was made for developing an efficient generation of subtle mutations during the characterization of two crossover sites between the flanking ~6-kb homologous regions of the *hpt* gene on the T-DNA-based vectors

and their corresponding genomic regions: efficient transfer of multiple base changes (point mutations) scattered throughout the two homologous segments of ~12 kb in total occurred from the vector to the rice genome, implying that major crossover events tended to occur near the outer ends of each homologous segment (Fig. 6.3; Johzuka-Hisatomi et al. 2008). Not only the point mutations within the *Adh2* gene but also those on the adjacent *Copia*-like element were simultaneously transferred into genome, indicating that multiple point mutations on two or more contiguous genes could be concurrently introduced. Since the transferred point mutations from the vector were often intermingled with the corresponding plant genome sequences (Fig. 6.3), it would also be feasible to conduct localized mutagenesis to create new combinations of point mutations derived from the vector and the corresponding genome sequences. The observed intermingled point mutations on the entire homologous segments are most likely to be a result of the mismatch correction of heteroduplex molecules formed between the genomic and introduced T-DNA sequences (Johzuka-Hisatomi et al. 2008). Although double-stranded DNA intermediates are thought to serve mainly in the normal integration processes of T-DNA into the plant genome by the NHEJ machinery (Tzfira et al. 2004; Weinthal et al. 2013), single-stranded T-DNA molecules are likely intermediates in the HR-mediated TGT processes with positive–negative selection because the transient expression of the double-stranded *DT-A* gene in the intermediates formed in the plant nucleus is thought to kill host plant cells and because the efficient transfer of base changes along with the long heteroduplex formation and mismatch correction appears to be easily explainable by a single-stranded T-DNA intermediate model among other alternative models (Iida and Terada 2004, 2005; Johzuka-Hisatomi et al. 2008).

For generation of subtle mutations, it is prerequisite to eliminating positive-selection markers from the targeted loci after introduction of point mutations into the target gene (Fig. 6.3). Various strategies have been developed to remove positive-selection markers from transgenic plants, and they include the utilization of site-specific recombinases, transposases, and sequence-specific engineered endonucleases (Hohn et al. 2001; Yau and Stewart 2013; Li et al. 2013; Weinthal et al. 2013). Of these, site-specific recombinase Cre-promoted excision of *hpt* from the targeted *Waxy* and *Rac1* loci was shown to result in reactivation of these genes (Terada et al. 2010; Thi Dang et al. 2013). However, excision of selection markers by site-specific recombinases leaves one copy of the recognition sequence (e.g., *loxP* site for Cre recombinase) at the excised site. Likewise, excision of selection markers by the *Ac* transposase originated from maize nearly always leaves a footprint sequence (small sequence alterations including insertions and deletions) at the excised site. The presence of the foreign *loxP* sequence may affect the expression of the nearest gene under certain circumstances. Since the 34-bp *loxP* site is originated from bacteriophage P1, plants containing a *loxP* sequence are regarded to be transgenic plants (Nielsen 2003). In this respect, the *piggyBac* transposase system is superior to the Cre-*loxP* system: the *piggyBac* transposon, originally isolated from the cabbage looper moth *Trichoplusia ni* genome, inserts at target sites with the sequence TTAA and usually excises itself precisely without leaving any footprint sequences (Li et al. 2013). Moreover, it has been demonstrated

recently that *piggyBac* transposase-promoted precise excision of selection markers without leaving any sequence alterations at the excised site in rice (Nishizawa-Yokoi et al. 2014). Thus, it would be feasible to generate subtle mutant rice plants by GT with positive–negative selection, in which a positive marker is designed to place into an appropriate TTAA sequence, followed by *piggyBac* transposase-mediated precise excision of the selection marker. A viral gene-expression system for the delivery of ZFNs into target cells in order to introduce genomic DSBs has also been recently reported, and a characteristic feature of this system is that the targeted cells can be classified as nontransgenic because viral RNA genomes do not integrate into the plant genome (Marton et al. 2010, 2013). It is conceivable that the *piggyBac* transposase, instead of ZFNs, can also be delivered into the target cells in order to eliminate positive-selection markers without leaving any sequence alterations at the excised site.

6 Future Prospects

Even though the sequence-specific engineered endonuclease-promoted genome editing has become quite popular in applying the functional analysis of the gene of interest and the molecular breeding of crop plants, there has been slow but steady progress in HR-mediated GT with positive–negative selection in rice, an important staple food for more than half the world's population (Table 6.1). One of the characteristic features of rice GT with positive–negative selection is that multiple base changes on vectors can be concurrently transferred into the rice genome, and the transferred base change is localized at least in a 6-kb range of segment from the *hpt* gene (Fig. 6.3; Johzuka-Hisatomi et al. 2008). Another is that a null allele that is not transmittable to progeny can be repeatedly obtained, because reproducibly obtained targeted calli or their regenerated T0 plants are heterozygotes with the TGT alleles (Ono et al. 2012). Although the nontransmittable null alleles must provide invaluable information about the function of the gene of interest, it is impossible to isolate by conventional mutagenesis techniques (Upadhyaya 2007; Wang et al. 2013) or by an *in planta* gene-targeting procedure proposed recently (Fauser et al. 2012), because the desired mutants are isolated in the progeny population.

For molecular breeding in rice, it would be feasible to develop a new vector system in order to minimize various potential risk factors. Instead of the *hpt* gene, the mutated rice *ALS* gene for herbicide tolerance can be used as a positive-selection marker (Endo et al. 2007) for increasing public acceptance of the technology and for fulfilling criteria for intragenic crop plants, in which the inserted DNA segment is derived from the transformed species itself or from closely related species capable of sexual hybridization (Nielsen 2003; Rommens et al. 2007; Holme et al. 2013). P-DNA, a plant-derived transfer DNA containing border-like sequences (Rommens et al. 2007), can be employed even though P-DNA sequences are not retained in the targeted genome. Although the DT-A protein is not regarded as a toxin because of

the cell-autonomous killing activity and its gene is not carried by the targeted genome, its designation itself may give a discomforting impression when applied to a breeding program in crop species. In this respect, other possible negative-selection markers suitable for breeding programs in the future may be genes encoding RNases, Barnase and RNase T1, from *Bacillus amyloliquefaciens* and *Aspergillus oryzae*, respectively, because they are suitable for genetic ablation (Day and Irish 1997). In particular, *A. oryzae* is a fungus widely used in traditional Japanese fermentation industries including soy sauce, sake, and vinegar production. To utilize the gene encoding RNase T1 as a negative-selection marker, it must be modified, e.g., by using an appropriate rice intron (intragenic sequence), in order to express efficiently in rice and simultaneously to prevent from killing the prokaryotic host cells, namely *E. coli* and *Agrobacteria*, through its leaky expression. To remove the mutated rice *ALS* gene, an appropriate viral gene-expression system can be utilized for the delivery of *piggyBac* transposase into target cells (Fig. 6.3). For desired subtle mutagenesis, the resulting intragenic plants should bear only a predesigned subtle mutation, only the sequence of which differs from the parental sequence. Such a project will be a challenge even with regard to rice, but if it could be realized, it would serve as a prototype for the future development of optimal GT in other agronomically important crops.

Acknowledgement We thank Mikio Nakazono and Hirokazu Kobayashi for their encouragement.

References

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods 9:39

Butaye KMJ, Cammue BPA, Delauré SL, De Bolle MFC (2005) Approaches to minimize variation of transgene expression in plants. Mol Breed 16:79–91

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16:735–743

Da Ines O, White CI (2013) Gene site-specific insertion in plants. In: Renault S, Duchateau P (eds) Site-directed insertion of transgenes. Springer, Dordrecht, pp 287–315

Day A, Goldschmidt-Clermont M (2011) The chloroplast transformation toolbox: selectable markers and marker removal. Plant Biolechnol J 9:540–553

Day CD, Irish VF (1997) Cell ablation and the analysis of plant development. Trends Plant Sci 2:106–111

Endo M, Osakabe K, Ono K, Handa H, Shimizu T, Toki S (2007) Molecular breeding of a novel herbicide-tolerant rice by gene targeting. Plant J 52:157–166

Fauser F, Roth N, Pacher M et al (2012) *In planta* gene targeting. Proc Natl Acad Sci U S A 109:7535–7540

Gouble A, Smith J, Bruneau S et al (2006) Efficient *in toto* targeted recombination in mouse liver by meganuclease-induced double-strand break. J Gene Med 8:616–622

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

Hiei Y, Komari T (2008) *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. Nat Protoc 3:824–834

Hohn B, Levy AA, Puchta H (2001) Elimination of selection markers from transgenic plants. Curr Opin Biotechnol 12:139–143

Holme IB, Wendt T, Holm PB (2013) Intragenesis and cisgenesis as alternatives to transgenic crop development. Plant Biotechnol J 11:395–407

Iida S, Hiestand-Nauer R (1987) Role of the central dinucleotide at the crossover sites for the selection of quasi sites in DNA inversion mediated by the site-specific Cin recombinase of phage P1. Mol Gen Genet 208:464–468

Iida S, 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

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

Ishizaki K, Johzuka-Hisatomi Y, Ishida S, Iida S, Kohchi T (2013) Homologous recombination-mediated gene targeting in the liverwort *Marchantia polymorpha* L. Sci Rep 3:1532

Jain SM (2001) Tissue culture-derived variation in crop improvement. Euphytica 118:153–166

Johzuka-Hisatomi Y, Terada R, Iida S (2008) Efficient transfer of base changes from a vector to the rice genome by homologous recombination: involvement of heteroduplex formation and mismatch correction. Nucleic Acids Res 36:4727–4735

Johzuka-Hisatomi Y, Noguchi H, Iida S (2011) The molecular basis of incomplete dominance at the *A* locus of *CHS-D* in the common morning glory, Ipomoea purpurea. J Plant Res 124:299–304

Kaeppler SM, Kaeppler HF, Rhee Y (2000) Epigenetic aspects of somaclonal variation in plants. Plant Mol Biol 43:179–188

Komari T, Hiei Y, Kasaoka K. Method for selecting transformed cells. *World Patent Application*, WO/1999/005296; 1999

Lewandoski M (2001) Conditional control of gene expression in the mouse. Nat Rev Genet 2:743–755

Li X, Burnight ER, Cooney AL, et al. *piggyBac* transposase tools for genome engineering. *Proc Natl Acad Sci USA*. 2013; 110: E2279–2287

Lieberman-Lazarovich M, Levy AA (2011) Homologous recombination in plants: an antireview. Methods Mol Biol 701:51–65

Maliga P (2004) Plastid transformation in higher plants. Annu Rev Plant Biol 55:289–313

Marton I, Zuker A, Shklarman E et al (2010) Nontransgenic genome modification in plant cells. Plant Physiol 154:1079–1087

Marton I, Honig A, Omid A et al (2013) From *Agrobacterium* to viral vectors: genome modification of plant cells by rare cutting restriction enzymes. Int J Dev Biol 57:639–650

Menke DB (2013) Engineering subtle targeted mutations into the mouse genome. Genesis 51: 605–618

Meyer P, Heidmann I, Forkmann G, Saedler H (1987) A new petunia flower colour generated by transformation of a mutant with a maize gene. Nature 330:677–678

Moritoh S, Eun C-H, Ono A et al (2012) Targeted disruption of an orthologue of *DOMAINS REARRANGED METHYLASE 2*, *OsDRM2*, impairs the growth of rice plants by abnormal DNA methylation. Plant J 71:85–98

Mussolino C, Cathomen T (2013) RNA guides genome engineering. Nat Biotechnol 31:208–209

Nielsen KM (2003) Transgenic organisms—time for conceptual diversification? Nat Biotechnol 21:227–228

Nishizawa-Yokoi A, Nonaka S, Saika H, Kwon Y-I, Osakabe K, Toki S (2012) Suppression of Ku70/80 or Lig4 leads to decreased stable transformation and enhanced homologous recombination in rice. New Phytol 196:1048–1059

Nishizawa-Yokoi A, Endo M, Osakabe K, Saika H, Toki S (2014) Precise marker excision system using an animal-derived *piggyBac* transposon in plants. Plant J 77:454–463

Ono A, Yamaguchi K, Fukada-Tanaka S, Terada R, Mitsui T, Iida S (2012) A null mutation of *ROS1a* for DNA demethylation in rice is not transmittable to progeny. Plant J 71:564–574

Ozawa K, Wakasa Y, Ogo Y, Matsuo K, Kawahigashi H, Takaiwa F (2012) Development of an efficient *Agrobacterium*-mediated gene targeting system for rice and analysis of rice knockouts lacking granule-bound starch synthase (*Waxy*) and β1,2-Xylosyltransferase. Plant Cell Physiol 53:755–761

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

Puchta H, Fauser F (2014) Synthetic nucleases for genome engineering in plants: prospects for a bright future. Plant J 78:727–741

Puchta H, Dujon B, 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 U S A 93:5055–5060

Rommens CM, Haring MA, Swords K et al (2007) The intragenic approach as a new extension to traditional plant breeding. Trends Plant Sci 12:397–403

Saika H, Oikawa A, Matsuda F, Onodera H, Saito K, Toki S (2011) Application of gene targeting to designed mutation breeding of high-tryptophan rice. Plant Physiol 156:1269–1277

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

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

Shan Q, Wang Y, Li J et al (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31:686–688

Shukla VK, Doyon Y, Miller JC et al (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. Nature 459:437–441

Sorrell DA, Kolb AF (2005) Targeted modification of mammalian genomes. Biotechnol Adv 23: 431–469

Strotbek C, Krinninger S, Frank W (2013) The moss *Physcomitrella patens*: methods and tools from cultivation to targeted analysis of gene function. Int J Dev Biol 57:553–564

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

Terada R, Asao H, Iida S (2004) A large-scale *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, Johzuka-Hisatomi Y, Saitoh M, Asao H, Iida S (2007) Gene targeting by homologous recombination as a biotechnological tool for rice functional genomics. Plant Physiol 144: 846–856

Terada R, Nagahara M, Furukawa K, Shimamoto M, Yamaguchi K, Iida S (2010) Cre-*loxP* mediated marker elimination and gene reactivation at the *waxy* locus created in rice genome based on strong positive–negative selection. Plant Biotechnol 27:29–37

Thi Dang T, Shimatani Z, Kawano Y, Terada R, Shimamoto K (2013) Gene editing a constitutively active OsRac1 by homologous recombination-based gene targeting induces immune responses in rice. Plant Cell Physiol 54:2058–2070

Thyagarajan B, Guimarães MJ, Groth AC, Calos MP (2000) Mammalian genomes contain active recombinase recognition sites. Gene 244:47–54

Thykjær T, Finnemann J, Schauser L, Christensen L, Poulsen C, Stougaard J (1997) Gene targeting approaches using positive–negative selection and large flanking regions. Plant Mol Biol 35:523–530

Toki S, Hara N, Ono K et al (2006) Early infection of scutellum tissue with *Agrobacterium* allows high-speed transformation of rice. Plant J 47:969–976

Tzfira T, Li J, Lacroix B, Citovsky V (2004) *Agrobacterium* T-DNA integration: molecules and models. Trends Genet 20:375–383

Upadhyaya N (ed) (2007) Rice functional genomics—challenges, progress and prospects. Springer, New York

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

Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. Annu Rev Plant Biol 64:327–350

Wakasa Y, Hayashi S, Ozawa K, Takaiwa F (2012) Multiple roles of the ER stress sensor IRE1 demonstrated by gene targeting in rice. Sci Rep 2:944

Wang N, Long T, Yao W, Xiong L, Zhang Q, Wu C (2013) Mutant resources for the functional analysis of the rice genome. Mol Plant 6:596–604

Waterworth WM, Drury GE, Bray CM, West CE (2011) Repairing breaks in the plant genome: the importance of keeping it together. New Phytol 192:805–822

Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R (2008) Molecular biology of the gene, 6th edn. Cold Spring Harbor, Cold Spring Harbor Laboratory Press

Weinthal DM, Taylor RA, Tzfira T (2013) Nonhomologous end joining-mediated gene replacement in plant cells. Plant Physiol 162:390–400

Yamauchi T, Johzuka-Hisatomi Y, Fukada-Tanaka S, Terada R, Nakamura I, Iida S (2009) Homologous recombination-mediated knock-in targeting of the *MET1a* gene for a maintenance DNA methyltransferase reproducibly reveals dosage-dependent spatiotemporal gene expression in rice. Plant J 60:386–396

Yamauchi T, Johzuka-Hisatomi Y, Terada R, Nakamura I, Iida S (2014) Disruption of the rice *MET1b* gene encoding a maintenance DNA methyltransferase resulted in viviparous germination or early embryonic lethality. Plant Mol Biol 85:219–232

Yau Y-Y, Stewart CN Jr (2013) Less is more: strategies to remove marker genes from transgenic plants. BMC Biotechnol 13:36

Zale JM, Agarwal S, Loar S, Steber CM (2009) Evidence for stable transformation of wheat by floral dip in *Agrobacterium tumefaciens*. Plant Cell Rep 28:903–913

Chapter 7
Recombinase Technology for Precise Genome Engineering

James G. Thomson and Ann Blechl

Abstract Here we review multiple uses of site-specific recombination and their applications in plant genomic engineering. We also outline various strategies for the combined use of multiple site-specific recombinase systems to precisely target transgenes into a predetermined locus and remove unwanted selectable marker genes.

For three decades, genetic engineering via introduction of DNA via *Agrobacterium* or biolistic technologies has been used to make plants with new traits. These methods can be employed to introduce new genes and/or alter the expression of native genes. These modifications of plant genomes have allowed scientists to establish relationships between genes and their effects on plant development, metabolism, and composition. The use of a variety of promoters has been established to drive gene expression in a constitutive, developmental, or tissue-specific fashion. Inducible promoters (e.g., heat shock, wounding) and promoter systems (glucocorticoid and estrogen based) are now available that allow the use of external signals to control gene expression. As we detail below, such control, when combined with site-specific recombination, allows one to remove unwanted selectable marker genes, turn genetic switches on or off, resolve complex insertions, and perform targeted integration.

With the release of genetically engineered (GE) versions of major crops such as cotton, soybean, canola, and maize with commercially useful traits, control of genetic modification for commercial production (Bioscience 58:391–401, 2008) has become an issue of global importance. The enhancements in crop quality through GE technology have had significant and positive impacts on farm income due to a combination of increased productivity and reduced inputs. In 2007, the direct global farm income benefit from biotech crops was $10.1 billion. Since 1996, farm incomes have increased by $44.1 billion (GM Crops Food 3:265–272, 2012) and thus the economic incentives for GE crops continue to grow. The isolation of stable transgenic lines with predictable and stable levels of transgene expression is

Author Contributions Conceived the paper: JT. Wrote the paper: AB and JT. Edited the paper: AB and JT.

J.G. Thomson, Ph.D. (✉) • A. Blechl, Ph.D.
Crop Improvement and Genetics Unit, USDA Agricultural Research Service,
Western Regional Research Center, 800 Buchanan St., Albany, CA 94710, USA
e-mail: James.Thomson@ars.usda.gov; ann.blechl@ars.usda.gov

F. Zhang et al. (eds.), *Advances in New Technology for Targeted Modification of Plant Genomes*, DOI 10.1007/978-1-4939-2556-8_7

highly desired but using standard approaches (e.g., *Agrobacterium* or biolistic technologies) is labor intensive and costly. It is often necessary to screen hundreds of independently transformed plants to identify those with suitable transgene structure and expression. Therefore, research endeavors are focused on goals to eliminate random DNA integration and/or reduce the frequency of multi-copy transgene insertions, and to reduce or eliminate events that exhibit unreliable transgene expression (In Vitro Cell Dev Biol Plant 41:213–219, 2005). Site-specific recombination is a promising tool that can be used to address these challenges for crop genome engineering. In this review, we examine previous studies and discuss recent advances in the applications of site-specific recombinase technology.

1 Introduction

Here we review multiple uses of site-specific recombination and their applications in plant genomic engineering. We also outline various strategies for the combined use of multiple site-specific recombinase systems to precisely target transgenes into a predetermined locus and remove unwanted selectable marker genes.

For three decades, genetic engineering via introduction of DNA via *Agrobacterium* or biolistic technologies has been used to make plants with new traits. These methods can be employed to introduce new genes and/or alter the expression of native genes. These modifications of plant genomes have allowed scientists to establish relationships between genes and their effects on plant development, metabolism, and composition. The use of a variety of promoters has been established to drive gene expression in a constitutive, developmental, or tissue-specific fashion. Inducible promoters (e.g., heat shock, wounding) and promoter systems (glucocorticoid and estrogen based) are now available that allow the use of external signals to control gene expression. As we detail below, such control, when combined with site-specific recombination, allows one to remove unwanted selectable marker genes, turn genetic switches on or off, resolve complex insertions, and perform targeted integration.

With the release of genetically engineered (GE) versions of major crops such as cotton, soybean, canola, and maize with commercially useful traits, control of genetic modification for commercial production (Wang and Moeller 2008) has become an issue of global importance. The enhancements in crop quality through GE technology have had significant and positive impacts on farm income due to a combination of increased productivity and reduced inputs. In 2007, the direct global farm income benefit from biotech crops was $10.1 billion. Since 1996, farm incomes have increased by $44.1 billion (Brookes and Barfoot 2012) and thus the economic incentives for GE crops continue to grow. The isolation of stable transgenic lines with predictable and stable levels of transgene expression is highly desired but using standard approaches (e.g., *Agrobacterium* or biolistic technologies) is labor intensive and costly. It is often necessary to screen hundreds of independently transformed

plants to identify those with suitable transgene structure and expression. Therefore, research endeavors are focused on goals to eliminate random DNA integration and/or reduce the frequency of multi-copy transgene insertions, and to reduce or eliminate events that exhibit unreliable transgene expression (Ow 2005). Site-specific recombination is a promising tool that can be used to address these challenges for crop genome engineering. In this review, we examine previous studies and discuss recent advances in the applications of site-specific recombinase technology.

2 Recombinase Types and Catalysis

Site-specific recombinase systems were discovered in bacteria and yeast, where they facilitate a number of biological processes, including the phase variation of certain bacterial virulence factors and the integration of bacteriophages into host genomes. Site-specific recombination occurs at a specific DNA sequence or recognition site and involves DNA cleavage and reunion, resulting in integration, deletion, or inversion of a DNA fragment without the gain or loss of nucleotides (Fig. 7.1). The orientation of the recognition sites determines whether the intervening DNA is integrated, deleted, or inverted (Grindley et al. 2006).

The recombinase superfamily can be split into two fundamental groups, the tyrosine and serine recombinases. This division is based on the active amino acid (Tyr or Ser) within the catalytic domains of the enzymes in each family. Each family can be further subdivided into subfamilies based on either size or mode of recombinase action. The first and best-characterized group are bidirectional tyrosine recombinases and includes the Cre-*lox* (Sauer and Henderson 1990), FLP-*FRT* (Golic and Lindquist 1989), and R-*RS* (Onouchi et al. 1991) systems, where Cre, FLP, and R are the recombinases and *lox*, *FRT*, and *RS* are their respective DNA recognition sites (i.e., sequences the enzymes recognize to perform recombination). Both recognition sites, while unique to each of the enzymes in this subfamily, are identical and configured as inverted repeats surrounding a spacer sequence. The spacer sequence is also known as the "region of crossover" and is the DNA that is exchanged during recombination. This sequence is further required for directionality or orientation of the recognition sites and determines whether the outcome is an excision, integration, or inversion event. The inverted repeats comprise the binding domain of the recombinase enzyme. Because the recognition sites of this subfamily are identical (Fig. 7.2a) and DNA exchange occurs without loss of nucleotides, the recombination reaction is fully reversible. However, intramolecular recombination (excision) is kinetically favored over intermolecular reaction (integration).

During recombination, a Holliday junction is formed between the DNA recognition sites and the tyrosine recombinase protomers (Grindley et al. 2006). DNA strand swapping is a two-step process with a mandatory isomerization to reconfigure the synaptonemal complex and allow the reaction to go to completion (Grindley et al. 2006). Cleavage of the DNA occurs at the spacer sequence (region of crossover) where an 8 bp staggered cut is formed between adjacent strands (Fig. 7.2a).

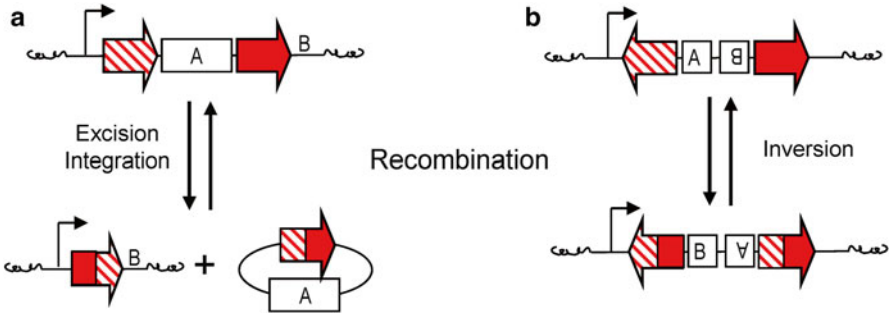

Fig. 7.1 Schematic representations of reactions catalyzed by site-specific recombinases. The recombinases (not shown) for these scenarios can be provided in *cis* or *trans*. (**a**) Excision and integration events. The "A" DNA is flanked by directly oriented recombinase recognition sites (*red* and *striped arrows*). A promoter (*bent arrow*) is outside the recognition sites. The excision event removes the "A" DNA between the associated recognition sites, leaving the external DNA "B" intact and adjacent to a single hybrid recognition site. The non-replicating circular DNA fragment is most often lost or rarely reintegrated if the hybrid site is functional. The system can also be used in reverse to integrate the "A" DNA into the genome located recognition site. A specific application of the excision reaction is in Molecular Switches (see text). In that case, The "A" DNA is a stuffer, a selectable marker coding region with transcription terminators that prevent read-through to the "B" DNA, which is a GOI. The marker gene in "A" is driven by a promoter (*bent arrow*) outside of the recognition sites. The excision event that removes the stuffer sequence will delete the transcription terminators and bring the previously distal GOI "B" in proximity for transcription. A specific application of the integration reaction is when the "B" gene is a negative/positive selectable marker and integration of the "A" DNA prevents "B's" expression. Site-specific integrants can be detected by applying the negative selection for "B" expression. This strategy facilitates gene targeting and the reuse of genomic locations that support stable reproducible levels of transgene expression. (**b**) Inversion event. The "A" and "B" DNA's are flanked by inversely oriented recombinase recognition sites (*red* and *striped arrows*). A promoter (*bent arrow*) is outside the recognition sites. Site-specific recombination results in an inversion of the DNA between the original sites and production of hybrid sites. The reaction is reversible if the hybrid sites are functional. A specific application of inversion is in Molecular Switches (see text). In that case, expression of "A" gene can be inactivated and expression of the "B" gene can be activated an inversion reaction that brings "B" DNA into correct proximity of the promoter

Homology between the two "regions of crossover" is required for completion of the recombination reaction. Modifications to the 8 bp "region of crossover" have become a crucial element in the engineering of unique heterospecific recognitions sites for more sophisticated recombinase strategies (see below).

Members of the other tyrosine recombinase subfamily have nonidentical recognition sites typically known as *attB* (attachment site bacteria) and *attP* (attachment site phage) (Fig. 7.2b). These enzymes perform irreversible recombination unless helper proteins are also present. The unidirectional tyrosine recombinases that have been shown to be useful for genome manipulation include HK022 (Gottfried et al. 2005; Kolot et al. 1999) and a modified form of λ (Christ and Dröge 2002). While these are the least utilized division of the recombinase superfamily for DNA manipulation, it should be pointed out that the modified version of λ is widely used in the commercially available "Gateway®" cloning system (http://www.lifetechnologies.com/us/en/home/life-science/cloning/gateway-cloning.html).

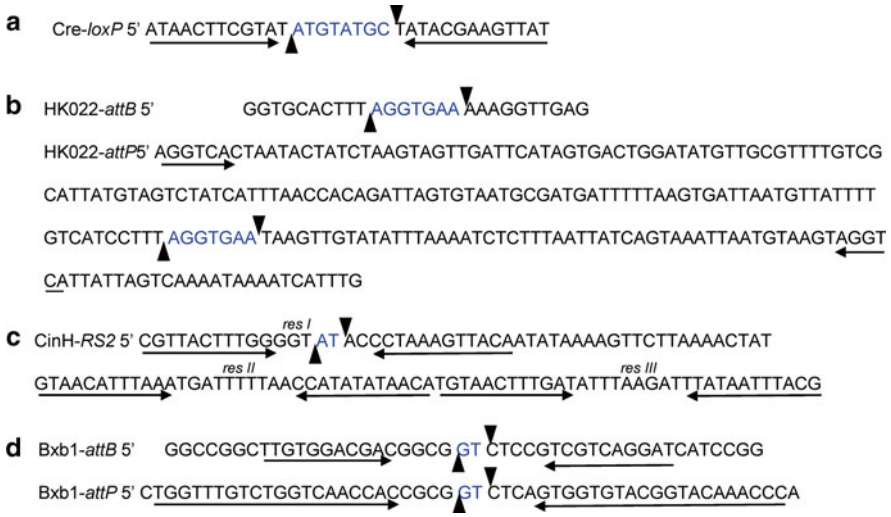

Fig. 7.2 Examples of site-specific recombinase recognition sites of the four recombinase families. The *horizontal arrows* delineate inverted repeats within the recognition sites. The *black triangles* show the sites of cleavage during recombination. The *blue* nucleotides are the region of crossover between the recognition sites. In order for recombination to occur, these *blue* nucleotides need to be identical in the two interacting sites. (**a**) The Cre/loxP system is a member of the tyrosine family with identical recognition sites. The reactions catalyzed by these enzymes are reversible. (**b**) The HK022/att system is a member of the tyrosine recombinase family with nonidentical recognition sites. The reactions catalyzed by these enzymes are nonreversible unless accessory proteins are also present. (**c**) The CinH/RS2 system is a member of the small serine family with identical recognition sites. However, the reactions catalyzed by these enzymes are nonreversible due to formation of synaptonemal complexes with considerable strain that act as directional filters. (**d**) The Bxb1/att system is a member of the large serine recombinase family with nonidentical recognition sites. The reactions catalyzed by these enzymes are nonreversible

The serine recombinase family also has two distinct subfamilies, the division being based on the size of the enzyme and the type of recognition sites used. Catalysis by these enzymes involves the covalent bonding of a serine residue and the formation of a synaptonemal complex, which leads to a concerted four-strand DNA cleavage and 180° twist event to complete the recombination process (Grindley et al. 2006). While stand cleavage for all the serine recombinases results in a 2 bp overhang (region of crossover), the composition of the recognition sites can be drastically different. Nevertheless, homology between the recognition sites' 2 bp "region of crossover" is required for reaction completion.

The small serine subfamily (aka. resolvase/invertase) includes β-*six* (Diaz et al. 2001), γδ-*res* (Schwikardi and Dröge 2000), CinH-*RS2* (Kholodii 2001; Thomson and Ow 2006), and ParA-*MRS* (Gerlitz et al. 1990; Thomson and Ow 2006) where β, γδ, CinH, and ParA are small serine recombinases, and *six*, *res*, *RS2*, and *MRS* are their respective DNA recognition sites. For the archetype Tn3, the recognition site contains three 12 bp inverted repeats (aka. binding sites) with space between regions of differing lengths designated *res I*, *II*, and *III* (Fig. 7.2c). *Res I* is the site

of cleavage, crossover, and exchange and is separated from *res II* and *III* by 22 bp in an asymmetric arrangement. The subsites *res II* and *III* participate in the twisted antiparallel complex configuration that forms the structure (synaptonemal complex) required for recombination. A total of six small serine resolvase protomers are required per recognition site, thus 12 per synaptonemal complex. While recombination mediated by these small serine recombinases utilizes identical recognition sites, only intramolecular excision events result. This appears to be due to tight requirements for synaptonemal formation, where recognition sites must be located in the correct orientation on the same stand of negatively supercoiled DNA in order for the accessory sites (*res II* and *III*) to interact prior to synapsis (Ghosh et al. 2003). Studies have determined that due to conformational strain of the synaptonemal complex formation, small serine recombinases cannot effectively facilitate intermolecular integration (Mouw et al. 2008). This is also known as a topological filter (Grindley et al. 2006). Therefore, an excision event mediated by a small serine recombinase family member is considered irreversible.

The large serine subfamily is represented by phiC31 (Rubtsova et al. 2008; Thomason et al. 2001), TP901-1 (Stoll et al. 2002), R4 (Olivares et al. 2001), Bxb1 (Keravala et al. 2006; Kim et al. 2003; Thomson and Ow 2006), and U153 (Thomson and Ow 2006). These enzymes act on two small recognition sites approximately 50 bp in size that differ in sequence. The recognition sites are typically known as *attB* and *attP* and after recombination, yield hybrid product sites known as *attL* and *attR* (Fig. 7.2d). Excision, inversion, or integration reactions can occur, but because the recognition site sequences of *attB* and *attP* become the hybrid sites *attL* and *attR*, the reactions cannot be reversed, unless a second protein, the corresponding excisionase, is also present (Ghosh et al. 2006; Thorpe et al. 2000). Compared to the small serine recombinases, the large recombinases form relatively simple synaptonemal complexes mediated by only four protomers. Complex formation is not the limiting step for recombination and thus integration reactions are possible. Although the catalytic domains are similar among the serine recombinases, the large types have longer C-terminal domains that have been shown to participate in DNA binding (Smith et al. 2010). They suggest that these domains stabilize the synaptonemal complexes, allowing intermolecular reactions to occur (Smith et al. 2010). The direction of the recombination event appears to be influenced by at least two factors: a protein/effector called recombinase directionality factor (RDF) (Ghosh et al. 2003; Khaleel et al. 2011) and the homology between the 2 bp "region of crossover" formed after cleavage (Zhang et al. 2010).

3 Application: Excision

Site-specific recombination was among the first methods used to create transgenic plants without retention of the selectable marker transgenes used to identify them (Dale and Ow 1991; Russell et al. 1992). Widespread adoption of this technology would constitute a bioconfinement strategy that would eliminate the possibility of

selectable marker transgenes spreading to other plants (Sang et al. 2013). Removal of the selectable marker also allows reuse of the same selection regime for subsequent rounds of gene transfer. A number of recombinase-mediated marker deletion strategies have been reported in model plants (Gils et al. 2008; Hare and Chua 2002; Hohn et al. 2001; Thomson et al. 2009, 2010, 2012), as well as in food crop species (Ballester et al. 2007; Blechl et al. 2012; Cao et al. 2006; Chawla et al. 2006; Chong-Pérez et al. 2012; Cuellar et al. 2006; Djukanovic et al. 2008; Gilbertson et al. 2003; Hoa et al. 2002; Hu et al. 2008; Kapusi et al. 2012; Kempe et al. 2010; Kerbach et al. 2005; Lyznik et al. 1996; Radhakrishnan and Srivastava 2005; Rubtsova et al. 2008; Sreekala et al. 2005; Srivastava et al. 1999; Srivastava and Ow 2003; Wurdig et al. 2013; Zhang et al. 2003; Zou et al. 2013). Recent reviews on the topic of recombinase-mediated excision strategies in plants are Gidoni et al. (2008), Moon et al. (2011), and Wang et al. (2011).

Of particular note is the recent development of the large serine irreversible phiC31 and Bxb1 systems as potential plant genome engineering tools. PhiC31 was shown to be fully functional in germ line cells of *Arabidopsis*, barley and wheat, where it mediated precise excision of a DNA fragment from the genome (Kapusi et al. 2012; Kempe et al. 2010; Thomson et al. 2010). It has also been used for excision in tobacco plastid genomes (Kittiwongwattana et al. 2007). Bxb1 also has the ability to excise DNA from both the plant nuclear (Thomson et al. 2012) and plastid genomes (Shao et al. 2013). These studies demonstrate that these recombination systems are suitable for the generation of marker-free, recombinase-free progeny from transgenic plants.

Several researchers have designed sophisticated refinements to the excision strategy, including the use of developmentally specific or inducible promoters to control recombinase expression (reviewed in Gidoni et al. 2008) and one-step auto-excision strategies for DNA removal. In the latter strategies, a gene(s) of interest (GOI), the marker gene, and the recombinase are cloned into a single transformation construct with the recombinase gene under the control of an inducible promoter such as a heat shock promoter (e.g., HSP81-1) (Hoff et al. 2001; Liu et al. 2005; Takahashi et al. 1992). The selection gene is generally flanked by directly oriented recognition sites with the GOI inserted outside of the region flanked by the sites allowing it to remain in the genome after an excision event. After transformation, the transgenic plants are treated (e.g., by heat shock) to induce expression of the recombinase and removal of the selectable marker.

In some of these designs, the recombinase Cre was fused to various ligand-binding domains (Logie and Stewart 1995; Metzger et al. 1995) or to the estrogen hormone receptor so that the fusion protein could not reach the nucleus in the absence of the ligand (Brocard et al. 1998; Danielian et al. 1998; Feil et al. 1996; Hunter et al. 2005; Kellendonk et al. 1996; Wunderlich et al. 2001). In the presence of the hormone inducer, a recombinase-mediated excision event deleted the intervening region between directly oriented recognition sites (e.g., *lox*) removing both the recombinase and the selectable marker genes. Auto-excision strategies like this have been applied in *Arabidopsis* (Zuo et al. 2001), tomato (Zhang et al. 2006), maize (Zhang et al. 2003), aspen (Matsunaga et al. 2002), rice (Sreekala et al. 2005),

and tobacco (Liu et al. 2005; Wang et al. 2005). Because chemicals or heat shock treatments are required for recombinase synthesis or nuclear localization, such marker gene deletion strategies may not be applicable to all plant species. Another potential limitation of this approach is that ectopic or leaky expression of the recombinase could lead to premature excision that will decrease the number of transformants recovered by selection (Li et al. 2007). On the other hand, if induction is not complete and recombinase levels are low, transgene excision may not occur in all cells of the germ line and chimeric plants will result.

The auto-excision strategy can be used to excise the entire backbone of the *Agrobacterium* plasmid after its transfer into plant cells (Kondrak et al. 2006). In this application, the R recombinase was employed to eliminate the insertion of non-T-DNA bacterial sequences into the plant genome, an undesirable side effect of *Agrobacterium*-mediated transformation. Such extraneous DNA is thought to enhance gene silencing and destabilize transmission of the T-DNA to progeny. Auto-excision strategies could also be applied in "intragenic" approaches to transformation, where only "native" plant DNA is used in the vectors that integrate into plant genomes. For *Agrobacterium*-mediated intragenic transformation, native (aka. cryptic) functional recognition sites for the desired recombinase(s) would need to first be identified in and isolated from the plant's genome. Such sites have been found in mammalian genomes (Semprini et al. 2007; Thyagarajan et al. 2000, 2001).

Another permutation of the marker gene excision application is the use of developmentally regulated promoters to activate recombinase expression only within specific organs or tissues. This strategy may be more efficient than using inducible promoters because the induction of expression is built into the plant and no external agents are needed. Various germline-specific promoters have been employed for recombinase expression (Kopertekh et al. 2010; Li et al. 2007; Luo et al. 2007; Mlynárová et al. 2006; Verweire et al. 2007). In particular, activation of recombinase-mediated excision during male germ line development has provided a containment system to prevent transgene movement via pollen (Luo et al. 2007; Mlynárová et al. 2006; Moon et al. 2011). This technique could potentially reduce the risk of transgene flow to nearby plants in the environment, eliminating the adventitious presence of transgenes in non-GE crops or related wild species. An alternative version of this strategy was the use of an embryo-specific promoter to drive the temporal expression of the recombinase Cre to eliminate selectable marker transgenes in soybean seeds (Li et al. 2007).

A different way to control recombinase expression is by introducing an expression cassette into the target cells in such a way that it is not expected to be stably integrated (Albert et al. 1995; Araki et al. 1995; Srivastava and Ow 2002; Vergunst et al. 1998). In these cases, the targeted genome is only transiently exposed to the recombinase. However, Srivastava and Ow (2002) unexpectedly found that genomic integration of the recombinase gene had occurred in 40 % of host cells that underwent recombinase-mediated excision. This led to the design of vectors specifically for transient recombinase expression: one utilizes *Agrobacterium tumefaciens* transformation proteins to transport the recombinase (Kopertekh and Schiemann 2005;

Vergunst et al. 2000) while the other is a virally derived vector expressing the Cre recombinase (Jia et al. 2006; Kopertekh et al. 2004). More options for limiting the duration of recombinase expression include direct transformation of recombinase mRNA (de Wit et al. 1998) and use of peptides to facilitate cellular uptake of the recombinase protein itself (Peitz et al. 2002). Use of transient expression strategies eliminates the need to segregate away or auto-excise the recombinase expression cassette.

4 Application: Molecular Switches

A special use of site-specific recombinase-mediated excision is in "Molecular Switches." The purpose of a molecular switch is to make expression of a GOI dependent on introduction or expression of the recombinase. In these applications, a piece of DNA (aka. stuffer) flanked by recognition sites is placed between a GOI's coding sequence and its promoter ("A" in Fig 7.1a). The stuffer sequence usually contains a transcription terminator, a series of terminators or in some extreme cases, an intron splice site followed by terminators; these elements force termination and/or mis-splicing of the transcript produced by the promoter. Thus, the GOI is off until the recombinase is introduced either by transformation, crossing, or induction of its promoter. The presence of the recombinase results in excision of the "stuffer," and the joining of the GOI to its promoter and thereby its expression in the plant (Fig. 7.1a). When the GOI is lethal and even more stringent control is required, the recombinase can be brought in via crossing to a separate line in a "Dual Switch" configuration (Sang et al. 2013). In a reverse strategy, a recombinase could be used to turn genes off by excision of the promoter or coding sequence from the host genome.

Inversion of DNA fragments offers another avenue for using a "Molecular Switch" to activate or inactivate a gene of interest (Fig. 7.1b). In this illustrated example, the recombinase-mediated inversion brings the "B" coding region into the correct orientation and in proximity to the promoter, allowing production of the "B" protein. The "flexing" technique uses inversion activation to identify which random transgene insertion sites would be suitable for transgene expression by screening transgenics for high levels of expression of a visual reporter (e.g., eGFP). Introduction of the recombinase swaps the positions of the reporter gene and GOI, thus activating the latter (Floss and Schnutgen 2008). Further use of the "flexing" technique is discussed below in the RMCE section. A very recent application of recombinase-mediated inversion has been for the production of synthetic circuits in *E. coli* cells using the basic computer programming techniques of AND, OR, and NOR (Bonnet et al. 2013; Siuti et al. 2013). A remarkable achievement! While such research so far is limited to prokaryotes, the ability to build memory into a living circuit is very impressive and raises possibilities for its employment in agricultural research.

5 Application: Transgene Insertion Site Resolution

Various factors affect transgene expression and stability in plants. The most prominent of these are the genomic location of transgene integration (aka. position effects) and the complexity of inserted DNA structures. In the case of position effects, transgene expression may be increased, decreased, or mis-regulated, depending on the surrounding genomic elements. Aspects of insertion structures that affect expression include transgene copy number, orientation, presence of transgene fragments, interspersion with native genomic DNA, and number of independent locations. For genomic engineering, the ability to insert a single-copy transgene into a location that allows faithful expression, both quantitatively and qualitatively, is most desirable. The ideal transgene is reliably expressed at expected levels in predictable tissues and at specific developmental times (Akbudak et al. 2010, 2013; Chawla et al. 2006; Day et al. 2000; Iyer et al. 2000) and is not subject to position effects (Albert et al. 1995; Hobbs et al. 1990; Kohli et al. 2003; Meyer 2000), sporadic silencing (mosaicism or variegation) (Burdon and Wall 1992; Marenkova et al. 2012), or complete silencing (Akbudak et al. 2013; Henikoff 1998; Martinez de Alba et al. 2013). Furthermore, the ideal transgene insertion does not result in genomic instability (Maqbool and Christou 1999) or endogenous gene inactivation (Wang et al. 2011). Research supports that single complete transgene copies inserted into the genome will produce additive gene expression (Akbudak et al. 2010) while partial or truncated copies can lead to gene silencing (Akbudak et al. 2013). These studies further indicate that a complete intact T-DNA insertion into the genome is critical for stable, predictable, and heritable gene expression (Srivastava et al. 2004). A study by Chawla et al. (2006) demonstrated that rice with multi-copy transgene inserts, initially silenced, recovered expression if the copies that were illegitimately integrated were removed by segregation in subsequent generations, leaving a single genomic copy in the targeted site. Thus an application to resolve transgene arrays removing deleterious copies could be used to decrease the total number of original transgenic plants required to find single-copy lines with faithful and heritable expression.

For all these reasons, considerable effort and time is spent in generating and characterizing transgenic plants to find those with simple insertion structures, reliable inheritance, and predictable expression. Site-specific recombination can be used in two ways to reduce these efforts. Site-specific integration can target a transgene to a genomic locus that allows high levels of faithful expression, as discussed below. Recombinase-mediated excision can be used to resolve complex multi-transgene loci into simpler structures that house one or two transgenes (Fig. 7.3). The success of this strategy depends on the design of the vector, which includes recognition sites in the same orientation flanking unwanted (e.g., marker gene) DNA, a GOI outside of those tandem sites, and a single recognition site in the opposite orientation (Fig. 7.3). Tandem transgene arrays will be excised due to the presence of the multiple recognition sites in direct orientation (Srivastava and Ow 2001). Recombinase-meditated excision will continue until a single recognition site is left (Fig. 7.3).

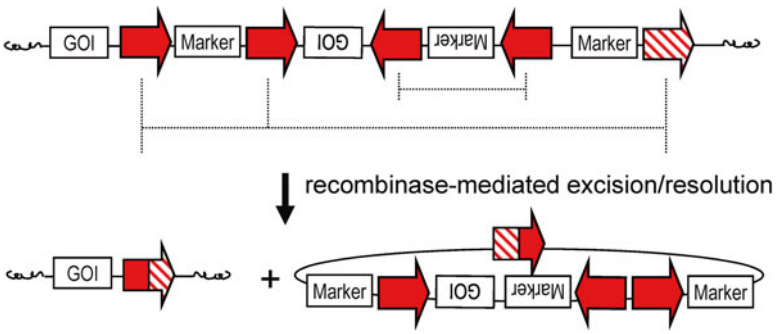

Fig. 7.3 Schematic representation of recombinase-mediated resolution of a complex transgene locus. This technique uses the recombinases' ability to excise DNA from between any two directly oriented recognition sites (*red* and *striped arrows*). The locus can consist of many different copies of the GOI and selectable marker genes in both orientations. However, excision will continue until a single site remains in the locus. The excised DNA exists as a non-replicating circle that can be further resolved and is lost. *Dotted lines* designate all possible excision events mediated by directly oriented recognition sites. DNA (including transgene DNA) outside the flanking recognition sites will not be removed. The recombinase (not shown) can be provided in *cis* or in *trans*. Not all possible excision intermediate products are shown

However, transgene fragments outside of the outermost recognition site will not be eliminated by this method. Use of recombinases to resolve complex insertions was originally demonstrated in wheat by (Srivastava et al. 1999): in four different transgenic lines generated by biolistics, multi-transgene insertions were reduced to single copies by Cre-mediated excision. Despite this success and the apparent importance of single-copy transgene insertion structures for faithful expression and commercial deregulation, this application of site-specific recombinase-mediated excision remains underutilized. This may be because it had been thought for many years that *Agrobacterium*-mediated transformation methods used to transform most plant species generally resulted in transgene loci with the desired properties. However, recent studies have shown that bacterial DNA outside of the T-DNA borders is more often than not integrated into the plant genome along with the T-DNA (Drs. Thilmony and Cook, unpublished observations). Thus, this application may garner more interest in the future.

6 Application: Integration

Recombinase-mediated integration can be used to insert a single copy of foreign DNA into predetermined locations within a genome, thus minimizing the likelihood of the position effects that plague random insertions (O'Gorman et al. 1991; Sauer and Henderson 1990). This application has been demonstrated in *Arabidopsis* (Louwerse et al. 2007; Vergunst and Hooykaas 1998; Vergunst et al. 1998), aspen (Fladung and Becker 2011), tobacco (Albert et al. 1995; Choi et al. 2000; Day et al. 2000;

Nanto and Ebinuma 2008; Nanto et al. 2005, 2009), maize (Kerbach et al. 2005), rice (Akbudak et al. 2010; Chawla et al. 2006; Nandy and Srivastava 2011; Srivastava and Ow 2002, 2004), soybean (Li et al. 2009), barley (Kapusi et al. 2012), and the plastid genome of tobacco (Lutz et al. 2004). Documented rates of integration range from ~33 % in tobacco (Albert et al. 1995; Day et al. 2000) to nearly 50 % in rice (Srivastava and Ow 2004). In other words, at least one plant in three contained single-copy insertions in predetermined genomic locations in these studies. Of the single-copy insertion events reported in rice, nearly all displayed heritable and predictable expression patterns, while in tobacco approximately half showed the expected expression pattern. The remaining half of the tobacco single-copy transgene insertions were affected by methylation-dependent DNA silencing (Day et al. 2000). These outcomes are much better than ~1–10 % single-copy faithfully expressed insertions that result from random integration transformation methods. In rice transformants generated by biolistics, Chawla et al. (2006) documented that site-specific integration into a preexisting genomic location yielded single-copy transgenes with stable gene expression over multiple generations in 11 of 18 events. Also noted in this study was how the initially silenced targeted transgenes recovered expression when segregation removed the extra transgene copies that were unlinked to the targeted location. This can most likely be attributed to reversal of homology-dependent and/or improper-termination gene silencing when repetitive or truncated, respectively, transgene DNA was removed (Luff et al. 1999; Luo et al. 2007).

Use of site-specific recombination for targeted integration requires two sequential transformations. First, a transgenic plant must be generated and shown to contain a single recombinase recognition site in a genomic location that allows faithful heritable expression of a marker gene or a GOI. Cells of this plant are then retransformed with one or two plasmids containing the complementary recognition site flanking the next GOI to be inserted and the recombinase expression cassette. Targeted integration in eukaryotes was first shown in yeast with the Cre recombinase (Sauer and Henderson 1990). However, the use of reversible recombinase systems such as Cre, Flp, and R for integration is inherently inefficient unless modifications are made that favor the bi-molecular integration reaction over the mono-molecular excision reaction. Improved targeted integration using Cre was achieved by providing the recombinase transiently, thereby trapping the DNA in its final position after the enzyme was gone (Albert et al. 1995; Baubonis and Sauer 1993; Srivastava and Ow 2001; Vergunst et al. 1998). This strategy was improved upon by placing the initial target *lox* recognition site in the genome between the promoter and open reading frame of the Cre recombinase. The recombinase protein is thus preloaded within the cell before introduction of the targeting plasmid, facilitating its integration. Once integration has occurred, the recombinase open reading frame is displaced from the promoter and effectively shut off, thereby trapping the DNA in the integrated state. This technique has been very effective for the reversible recombination systems of Cre and Flp. It has been used for integration in *Arabidopsis* (Vergunst et al. 1998), tobacco (Albert et al. 1995; Choi et al. 2000; Day et al. 2000), and rice (Chawla et al. 2006; Srivastava and Ow 2002, 2004). A further refinement was incorporation of two *lox* sites into the targeting DNA, such

that in the presence of Cre, unwanted "backbone/vector" DNA was removed to produce a circular fragment that was subsequently integrated into the genomic *lox* site (Kolb and Siddell 1997; Nandy and Srivastava 2011; Srivastava and Ow 2004; Vergunst and Hooykaas 1998; Vergunst et al. 1998). The result was a "clean" transgene lacking plasmid backbone sequences that could trigger transgene silencing (Iglesias et al. 1997).

Another improvement in using reversible recombinase systems for integration was first proposed by (Hoess and Abremski 1984), who suggested that Cre recombinase could tolerate mutations in one of the two 13 bp-binding domains of a *lox* recognition site because cooperative Cre–Cre interaction would allow *lox* site attachment to occur normally, but mutations in both 13 bp-binding domains would result in a large drop in overall binding efficiency, inhibiting further Cre/*lox* interactions. Thus, recombination between two *lox* sites with mutant-binding domains, but complementary to one another, would occur at relatively normal rates. The resulting double mutant *lox* recognition site would resist further recombination, trapping the integrated DNA. The Hoess and Abremski prediction was tested with several randomly generated mutant *lox* sites in vitro with a plasmid inversion template. Measurements of the forward and reverse reactions validated the prediction (Albert et al. 1995). Subsequent studies with other mutant-binding sites have shown that this strategy helps stabilize the integration events by inhibiting the reverse excision reaction (Albert et al. 1995; Araki et al. 2010; Ohtsuka et al. 2010; Thomson et al. 2003). Thus, the use of mutant *lox* sites has allowed targeted integration to be achieved at useful levels in plants such as tobacco (Albert et al. 1995; Day et al. 2000), rice (Chawla et al. 2006; Srivastava et al. 2004; Srivastava and Ow 2002), and maize (Srivastava and Ow 2001). However, use of double mutant *FRT* recognition sites for enhanced integration stability with the Flp system has not been successful, probably due to the intrinsic nature of the Flp/*FRT* interaction: Flp recombinase binding is not cooperative (Huang et al. 1991; Senecoff et al. 1988).

Unidirectional recombinases comprise a unique set of tools for genome engineering and can facilitate DNA integration without the possibility of reversal. Members of the large serine recombinase family, such as phiC31 (Thorpe and Smith 1998), R4 (Olivares et al. 2001), TP901-1 (Stoll et al. 2002; Thomson and Ow 2006), Bxb1 (Keravala et al. 2006; Russell et al. 2006; Thomson and Ow 2006), and phiBT1 (Zhang et al. 2010), offer this advantage over reversible recombinase systems because recombination occurs between nonidentical *attB* and *attP* sites. Direct comparison of integration into the same genomic location in *Schizomyces pombe* yielded completion rates of 85 %, 95 %, and 78 % for Bxb1, phiC31, and TP901-1, respectively (Thomson and Ow 2006). In plant applications, the phiC31 recombinase has been used for integration (Lutz et al. 2004) within the tobacco plastid genome. For lines expressing high levels of phiC31 integrase from a nuclear transgene, 17 independent transplastomic lines per bombarded leaf were obtained via site-specific integration into the plastid genome target. More recently, Yau et al. (2011) described the use of the Bxb1 large serine recombinase for site-specific integration into a predetermined locus of the tobacco nuclear genome. In this experiment, the Bxb1 *attP* site was inserted into the tobacco genome via random integration.

Single-copy transgenic plants were identified and used as targets for site-specific integration of a plasmid containing the Bxb1 *attB* site and a hygromycin resistance gene. Doubly transformed plants were identified by hygromycin selection and the integration sites were characterized by DNA blot analysis and by sequencing the PCR product junctions between the insert and target DNA's. Approximately 5 % of the secondarily transformed plants contained integration sites formed by site-specific recombination. More examples of the use of site-specific integration for crop plant improvement can be found in Srivastava and Gidoni (2010).

The use of endogenous recognition sites (or cryptic sites) in genomes provides an additional option for targeted chromosomal integration. This method has been demonstrated to work in various prokaryotes and eukaryotes (Groth et al. 2000; Olivares et al. 2001; Ou et al. 2009; Sauer 1996; Sauer and Henderson 1990; Thomson et al. 2003; Thyagarajan et al. 2000, 2001). Searches for potential cryptic recognition sites in plant genome sequences have only recently been conducted (Thomson et al. 2009, 2010, 2012). The utility of such sequences would have to be demonstrated by empirical studies before they could be deployed in GE crops.

7 Application: Recombinase-Mediated Cassette Exchange

A powerful application of recombinase technology combines the integration and excision reactions into a strategy to replace transgenes in situ. This technique is termed Recombination-Mediated Cassette Exchange (RMCE) and was originally used in cultured mammalian cells with the Flp/*FRT* recombinase system (Schlake and Bode 1994; Seibler and Bode 1997). The technique was effective in generating targeted modifications with a minimum of excess DNA remaining in the host genome (Bouhassira et al. 1997). The approach involves targeted integration of a GOI followed by a recombinase-mediated excision event removing selectable marker, plasmid backbone, and other unwanted DNA (Fig. 7.4). To implement RMCE in plants, a founder line containing recombinase recognition sites in the genome, a "TAG" plant, must first be generated. The "TAG" usually contains a positive selectable marker, such as an herbicide or antibiotic resistance gene, flanked by two inverted recognition sites (Fig. 7.4a). The inverted orientation of the recognition sites inhibits unwanted "TAG" auto-excision when identical sites are used (Feng et al. 1999; Nanto et al. 2005, 2009). The "EXCH" plasmid contains complimentary inverted recognition sites flanking the GOI to be inserted into the genome (Fig. 7.4a). Recombinase gene expression can be provided in *cis*, encoded on the EXCH plasmid but outside the recognition sites flanking the GOI cassette, or in *trans* from a separate incoming plasmid. Retransformation of the "TAG" or founder lines with the "EXCH" (and recombinase, if necessary) vector triggers two successive recombinase-mediated events, integration followed by excision, in which both sets of recognition sites will be used and the original "TAG" DNA can be exchanged with the incoming "EXCH" GOI (Fig. 7.4c). The predicted intermediate between integration and excision (Fig. 7.4b) consists of the entire construct integrated into

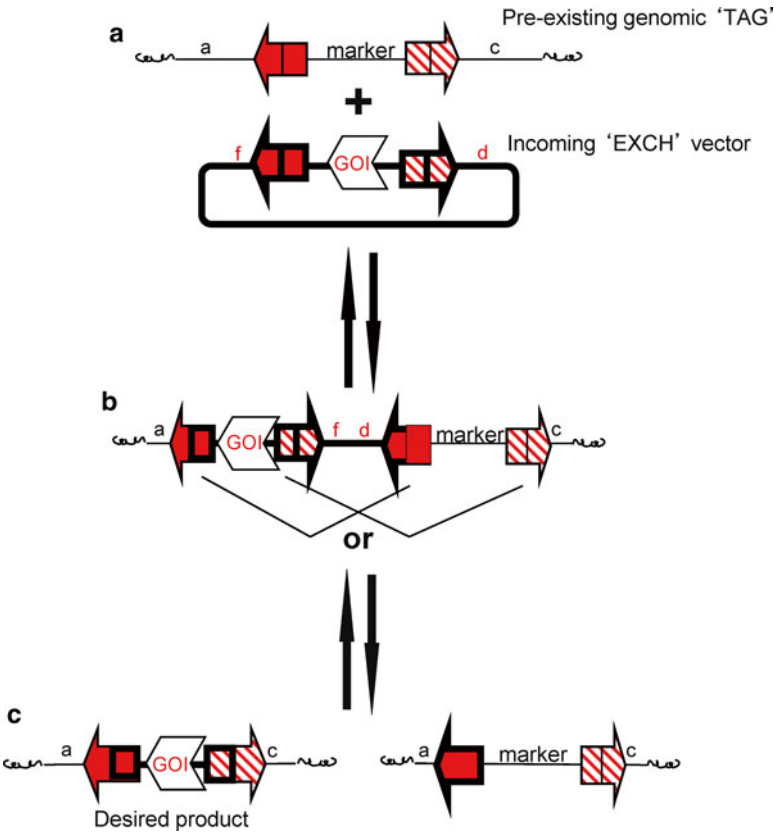

Fig. 7.4 RMCE schematic for a single bidirectional recombinase using heterologous sites. The *red* and *red-striped arrows* represent recognition sites and are designed to recombine exclusively with only their matching partner, but both set of sites recognize the same recombinase enzyme. This strategy combines both site-specific integration and excision mediated by a single recombinase and is an efficient way to produce transgenics with a minimum of extraneous DNA remaining in the host genome. Specifically, RMCE is a technique whereby DNA can be site-specifically integrated into a preexisting genomic target "TAG" with unwanted DNA removed at the conclusion. (**a**) A preexisting GE plant "TAG" contains a selectable marker gene (*marker*) flanked by inverted, and in this scenario heterologous, recognition sites (*thin lines*). The incoming plasmid DNA (*thick lines*) contains a GOI also flanked by the matching inverted heterologous recognition sites and is termed the exchange "EXCH" cassette. The recombinase (not shown) can be provided in *cis* or *trans* and will integrate the incoming "EXCH" DNA into the "TAG" genomic location utilizing one of the two flanking recognition sites. (**b**) Either set of directly oriented recognition sites can then be used for site-specific excision of the intervening DNA. (**c**) One of the two possible products (*left*) contains the GOI of the "EXCH" cassette in place of the marker of the "TAG" DNA, while the other possible product (*right*) retains only the marker gene. Reversible recombinase systems such as Cre and Flp can generate both forward and reverse reactions generating numerous intermediates (not shown)

the genome (Feng et al. 1999; Fiering et al. 1993; Nanto et al. 2005; Thomason et al. 2001). Recent evidence obtained by atomic force microscopy has confirmed that RMCE is a two-step process (Malchin et al. 2008).

The RMCE strategy has been used successfully with several recombinase systems, including Cre/*lox*, Flp/*FRT*, R/*RS*, phiC31/*att*, and Bxb1/*att*, and in many different organisms, including yeast, mammalian cells, mice, *Drosophila*, and plants (Belteki et al. 2003; Feng et al. 1999; Fladung and Becker 2011; Horn and Handler 2005; Li et al. 2009; Louwerse et al. 2007; Nanto and Ebinuma 2008; Nanto et al. 2005, 2009; Obayashi et al. 2012; Turan et al. 2013). In mammalian systems, targeting efficiencies into a preexisting chromosomal site can approach 100 % when a negative selection is used to identify cells in which exchange has occurred (Feng et al. 1999). Without selection, 1 % of cells contain exchanges (Feng et al. 1999). In plants, RCME has been demonstrated for both the direct DNA and *Agrobacterium-mediated* transformation methods (Dafhnis-Calas et al. 2005; Nanto and Ebinuma 2008).

Use of bidirectional tyrosine recombinases in RCME presents some difficulties because their reactions are reversible at each step of the process, decreasing its overall efficiency and complicating molecular analyses. If identical recognition sites are used, self-excision can occur, short circuiting the whole process. Use of heterologous binding sites, discussed above and illustrated in Fig. 7.4, reduces the likelihood of self-excision (Louwerse et al. 2007). Heterologous sites take advantage of the homology requirement between the 8 bp "regions of crossover" of the participating *FRT*, *lox*, or *RS* sites (Hoess et al. 1986; Lee and Saito 1998; Nanto et al. 2005). In particular, binding sites with 3–4 mutations rarely interact with wild type or with one another (Takata et al. 2011). Heterologous sites for both the Cre (Torres et al. 2011) and Flp (Tsanov et al. 2012) systems have been used in directing RMCE gene transfer into the genome in a predetermined orientation whereas the original design utilizing identical recognition sites allowed insertion in both directions (Seibler and Bode 1997). Scientists at DuPont Corporation used RMCE with heterologous recognition sites for the modification of the soybean genome to improve seed oil quality. Multiple genes were inserted at the same chromosomal site to simultaneously silence the *FAD2* and *FATB* genes, thereby improving the oxidative stability of the oil, and to over-express the *DGAT1* gene to increase overall oil content (Li et al. 2009, 2010).

A variation on RCME termed "Flex-ing" (Floss and Schnutgen 2008) used recombinase-mediated inversion to activate expression of a GOI. The coding sequence of the GOI was flanked by a pair of inverted heterologous FRT sites (F F′) (similar to Fig. 7.1b where "A" and inverted "B" are between inverted recognition sites). Transformants with randomly inserted DNA were screened for strong/specific promoter expression using the enhanced Green Fluorescent Protein (eGFP) as a reporter ("A" in Fig. 7.1b). Plants containing strongly expressed *eGFP* were retransformed with a plasmid that provided FLP, causing the *eGFP* cassette to invert and bringing the GOI adjacent to the promoter. The resultant genomic site had a single FRT (F) site on one side and three FRT sites on the other. An excision event on the triple stack (F′ F F′) of directly oriented FRT sites resolved it to a single

FRT (F′) site, eliminating further recombination due to the heterologous nature of the remaining FRT sites trapping the DNA in its flipped orientation. By design, the GOI coding sequence was now flanked by heterologous FRT sites (F-coding sequence-F′) and could serve as the target for future RMCE integration.

Another version of RMCE is dual RMCE (dRMCE), which employs two different recombinase expression systems. The first successful demonstration of this strategy combined the Cre/*lox* and Flp/*FRT* systems in conjunction with mutant recognition sites to stabilize the integration event (Lauth et al. 2002). Cells containing precise exchange events were identified without selection at a frequency of 3 %. Put into perspective, this is 3× better than the efficiency of obtaining cells by specific targeting with homologous recombination and identified using a selectable marker gene. Since then, many other groups have used variations of dRMCE. Dafhnis-Calas et al. (2005) used the Cre/*lox* and phiC31/*att* recombinase systems sequentially in mammalian cell cultures to stack genes into the same locus. They called their version "Iterative Site-Specific Integration." PhiC31 has also been paired with Flp for embryonic stem (ES) cell genome targeting (Monetti et al. 2011). This combination has become so successful that the North American Conditional Mouse Mutagenesis Project (NorCOMM) has created docking or "TAG" sites in more than 600 genes of the mouse genome for knockin, knockout, and/or gene replacement studies. Nanto and Ebinuma (2008) demonstrated the use of the R/*RS* and Cre/*lox* systems for dRMCE in tobacco. This well-executed study successfully generated several lines with marker-free single-copy transgenes at targeted integration sites. *Agrobacterium* was used as the vehicle for transformation, proving that RMCE can be executed from an incoming T-DNA molecule. More recently, the Cre/*lox* and phiC31/*att* recombinase systems were shown to be effective for dRMCE in *Arabidopsis* (De Paepe et al. 2013).

The use of unidirectional recombinase systems in dRMCE can make it even more effective for precise insertion of new transgenes into the genomes of previously characterized plant lines having stable expression patterns (aka transgene stacking). The large serine recombinases, namely Bxb1 and phiC31, are naturally unidirectional, and therefore trap the DNA within their target loci. Successive genome targeting has been demonstrated with phiC31 integration followed by Cre excision followed in the next generation by Bxb integration and Cre-mediated excision (Huang et al. 2009). In this design, integration by one unidirectional recombinase (in this case, phiC31) brings in the target recognition sites for a second recombinase (in this case, Bxb1) and vice versa. An unlimited number of successive dRMCE targeting events can be envisioned with this approach. The successive introduction of the various recombinases can be via genetic crosses or by transformation. Although it involves more steps and generations, the use of crosses and segregation provides more sequential control of the process. Recent atomic force microscopy experiments have confirmed that while RMCE is a two-step process, simultaneous expression of two integration competent recombinase enzymes (in this case, HK022 and FLP) results in competitive inhibition of the process (Malchin et al. 2008). Research indicates that only one synaptonemal complex can be formed at a time. These studies went on to demonstrate that sequential addition of the recombinases

was much more efficient than simultaneous addition. The competition hypothesis is further supported by research that demonstrated that the best efficiencies of RMCE were obtained in cells when Flp levels were higher than Cre levels (Anderson et al. 2012). This study demonstrated that when the levels of Cre were greater than or equal to Flp, RMCE efficiency dropped drastically. The optimum ratio for the two recombinase enzymes in dRMCE may depend on which recombinase systems are used.

Use of a third type of recombinase from the excision-only small serine category may increase dRCME efficiencies. Due to topological constraints (Mouw et al. 2008; Sarkis et al. 2001), these recombinases cannot form synaptonemal complexes without first having both recognition sites on the same strand of DNA and in the correct orientation. As such, these recombinases will not act until after the integration step has occurred, regardless of their concentrations relative to the integrase. Use of the excision-only small serine recombinase would also inhibit reintegration of the excised fragment.

An improvement to the overall RCME process is the employment of negative selectable markers (e.g., *codA*) to directly select cells in which gene replacements have occurred (Fig. 7.5) (Kondrak et al. 2006; Nanto and Ebinuma 2008). When the "EXCH" vector is site-specifically inserted into the chromosomal "TAG" cassette, the recognition sites required for excision will align, allowing removal of the negative selectable marker gene (Fig. 7.5). Cells in which the "EXCH" vector integrated at a random location, or where the excision event failed to remove the negative selection, would leave the marker gene in the genome and the cells will be killed by the negative selection agent.

Efficient RCME strategies are especially needed for high throughput gene integrations such as those used by biopharmaceutical companies to make complex proteins, for engineering complex metabolic pathways in plants, or for the study of genes with regard to biotechnology risk assessment. For plant GE, RCME offers the intriguing possibility of stacking multiple genes into a single genomic location. The TAG and EXCH vectors can be designed for multiple rounds of exchange. As RCME becomes routine, the plant researcher community would benefit from collections of crop germplasm with several "TAG" lines ready for RMCE targeting.

8 Application: Megabase Modifications/Chromosome Level Engineering

Site-specific recombination can occur between any two recognition sites in DNA that are present in the same cell (prokaryotes) or nucleus (eukaryotes). The probability of recombination is governed by physical parameters including the distance between recognition sites, their relative orientations, and their relationships (i.e., intra-chromosomal, homologous, nonhomologous) (Burgess and Kleckner 1999). Thus, site-specific recombination makes planned chromosomal rearrangements feasible for genetic engineering. To date the majority of the research that utilizes this approach has been done in mammalian cells and using the Cre/*lox* system.

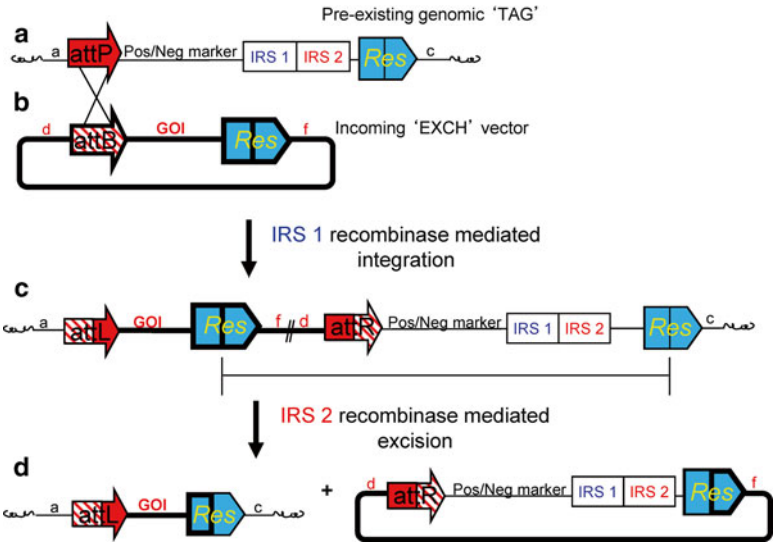

Fig. 7.5 Schematic representation of a sophisticated design for Dual Recombinase-Mediated Cassette Exchange. This strategy uses a positive/negative selection marker gene and two unidirectional recombinases, one for integration into the preexisting chromosomal "TAG," and one for excision of unwanted DNA. Cells that have undergone irreversible exchange of the GOI for the marker gene can be identified by negative selection. The preexisting "TAG" line (**a**) contains the target in a random location and is identified using the positive selection. The target contains a positive–negative selectable marker gene and expression cassettes for a large and a small serine recombinase (Irreversible Recombinase System—IRS1 and IRS2), all flanked by the recombinase recognition sites (*red attP* and *blue Res*, respectively). The EXCH cassette (**b**, *thick lines*) contains sites for each recombinase (*red striped attB* and *blue Res*) flanking the GOI and is introduced by transformation. Although both recombinases and their sites are present in the same transformed cells, the excisionase can only act after irreversible site-specific integration into the TAG site brings the Res sites into alignment on the same chromosome (**c**). Irreversible excision deletes the positive/negative selectable marker gene (**d**), allowing only the cells that contain the excision (*left* in **d**) to survive the negative selection agent. Cells in which the EXCH vector integrated at random or did not complete RMCE do not survive the selection. Alternative designs allow this same strategy to be used for repetitive gene stacking

Large scale chromosome modifications can be disruptive to the genome, and many will not be tolerated or stably maintained, particularly in diploid organisms such as mammals. Nevertheless, large deletions, duplications, or chromosomal translocations have facilitated the study of human genetic diseases such as Down's syndrome, Smith–Magenis syndrome, Cri-du-chat, and Charcot-Marie-Thooth type 1A (Chen et al. 1997; Korenberg et al. 1994; Street et al. 2002). Cre-induced site-specific translocations have been reported in mammalian ES cells at frequencies of 1:1,200–2,400 (nonrandom:random) (Van Deursen et al. 1995).

As previously explained, the orientation of the *lox* sites will determine the type of recombination event observed. With two *lox* sites located on a single chromosome in directly repeated orientation, a deletion event is expected, but duplication events have also been detected (Medberry et al. 1995; Ramirez-Solis et al. 1995;

Uemura et al. 2010; Van Deursen et al. 1995). The deletion/duplication events result from non-sister chromatid recombination that generates a set of balanced and unbalanced chromatids; cells with the latter will not survive. A more controlled and effective strategy was to employ the temporal-specific synaptonemal complex protein1 (SYCP1) gene promoter to control recombinase expression. In the spermatocytes of mice during meiosis, the homologous chromosomes are tightly paired in the synaptonemal complex. Using the SYCP1 promoter to drive Cre expression boosted the efficiency of generating deletion and duplication events between the paired chromosomes in the synaptonemal complex (Herault et al. 1998; Olson et al. 2005).

With the *lox* sites in the opposite orientation and located on the same chromosome, an inversion event is generated. This form of chromosomal rearrangement has been used in mice to study genetic abnormalities and establish "balanced" lethal systems to facilitate stock maintenance (Zheng et al. 1999). With the age of genomics upon us, the use of designed "balancer chromosomes" could be a critical element in understanding how genomic arrangements influence gene expression, metabolic pathway function, and pathway interactions on a global scale. However, inverted *lox* sites can also lead to unequal recombination between sister chromatids, generating dicentric and acentric chromosomes that are lost during subsequent cell divisions, precluding cell survival (Lewandoski and Martin 1997; Uemura et al. 2010).

Placement of the *lox* sites in the same orientation on nonhomologous chromosomes can lead to balanced and unbalanced chromosomal translocations. In human cells, these rearrangements have been used to study the effects of inappropriate regulation of spatial and temporal gene expression leading to various forms of cancer, developmental abnormalities, and genetic diseases (Van Deursen et al. 1995). In plants, the efficiency of using Cre to mediate translocations was studied. Transgenic tobacco plants were constructed in which one chromosome contained a promoterless hygromycin coding region adjacent to a 5′ *loxP* site and another chromosome contained the 35*S* promoter adjacent to a 3′ *loxP* site. After Cre protein was introduced, 2.5 % of the resulting plants were able to grow in the presence of hygromycin, indicating that recombination and presumably translocation had occurred between the two *loxP* sites (Qin et al. 1994). Translocations have also been engineered between chromosomes of two different plant species (Koshinsky et al. 2000). Protoplasts from *Arabidopsis* and tobacco plants were fused in culture. Induction of *Cre* expression resulted in a recombination event that joined a promoter region with the coding sequence of the hygromycin resistance gene. The DNA's from hygromycin-resistant calli were analyzed and found to contain the predicted junction between *Arabidopsis* and tobacco chromosomes. However, plants regenerated from the calli did not pass the hybrid chromosome to their progeny (Koshinsky et al. 2000).

A potential use of chromosomal recombination, as yet undemonstrated in crop plants, is to introgress useful transgene combinations from genotypes that are easy to transform (lab lines) into commercially adapted varieties (elite lines) that are difficult to transform. Such transfers are possible through traditional crossing, but several generations of backcrossing are required to recover most of the genome of the elite variety with only the transgene and a small amount of linked DNA from the

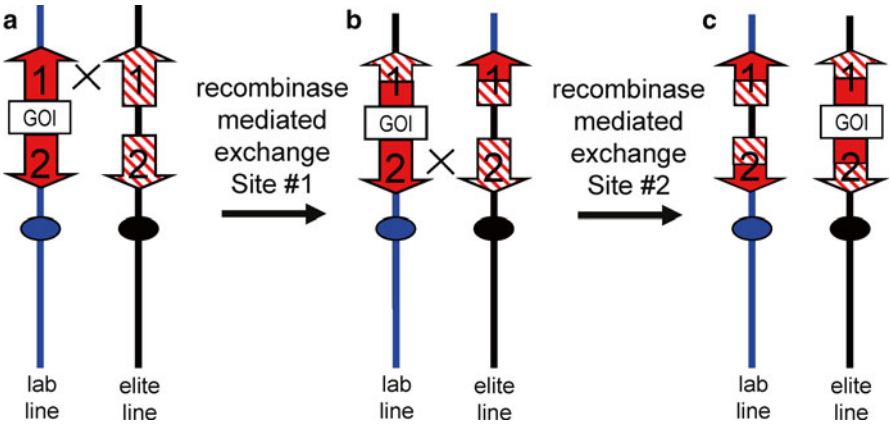

Fig. 7.6 Schematic representation of a recombinase-mediated introgression event to transfer a transgene from a genotype that is easy to transform (*lab line*) into an adapted commercially viable variety (*elite line*). The "lab" line contains a GOI flanked by divergently oriented recombinase recognition sites (*red arrows* in a, *left blue* chromosome) and the appropriated recombinase expression cassette. Inverted recognition sites prevent unwanted DNA excisions. The "elite" line has recognition sites for the same recombinase(s) (*striped red arrows* in a, *right black* chromosome) flanking a region of nonessential DNA. Two recombination events are needed can exchange the information located between the sites. (**b**) An intermediate transposition between the chromosomes produced by recombination of one of the two sites. (**c**) The final products after the second recombination event reverses the transposition. After the exchange is complete, progeny that have lost the lab line chromosome by segregation can be identified in a subsequent generation. While this strategy could theoretically be used with any recombinase, it would be much more efficient with the irreversible integrases

lab line parent. The association of the transgene and adjacent DNA is called linkage drag. If divergently oriented recognition sites for site-specific recombination could be introduced into one chromosome of the elite line, then a GOI flanked by the matching sites could be swapped into the elite chromosome without any extraneous DNA or disruption (Fig. 7.6). The lab line expressing recombinase would be crossed to the target elite line, exchange of chromosomal segments would occur, and progeny containing only elite chromosomes (including the exchanged one) would be recovered. Back-crossing would not be necessary, although one round would make it more likely to find the elite-only descendent. This strategy would speed the transfer of genomic modifications from laboratory to elite lines (Fig. 7.6) (Ow 2005). The most suitable recombinases for this approach are irreversible large serine type.

An interesting way to produce mutations in a genome on a megabase scale is through the use of "Targetrons" (Enyeart et al. 2013). Currently only proven to be effective in prokaryotes, the system uses a combination of mobile group II introns and the Cre/*lox* recombinase system. The introns embedded with a single *lox* sites are integrated into specific locations in the *E. coli* genome. Introduction of Cre expression results in large deletions or inversions that can be used to study how large scale genome rearrangements influence cellular processes.

An organized approach to chromosome reconstruction through utilization of site-specific technology would be extremely useful in the production of artificial chromosomes or "minichromosomes" (Gaeta et al. 2012; Houben et al. 2008). Artificial chromosomes offer an advantage to crop improvement by providing genomic loci that can be manipulated without regard for surrounding endogenous genes. They provide platforms to stack GOI and/or to study gene expression in individual chromatin domains. Such modules would be transmitted through meiosis and genetically stable. These platforms could also provide targets for recombinase-mediated integration, which would allow the addition of more genes even after establishment of the "minichromosome" in plant cells. Production of circular "minichromosomes" is easily done with site-specific recombinases in vitro or in single cells. However, in germ line cells of whole plants, such circles can form dicentric rings during meiosis leading to DNA breakage during segregation (McClintock 1938). To counter this, the presence of a "minichromosome"-specific resolvase such as Tn3 or CinH could in theory separate the rings and help stabilize inheritance. Currently, artificial chromosome inheritance rates are not up to the 100 % level required for commercialization (Carlson et al. 2007), but research continues to investigate possible uses of these amazing structures.

9 Conclusion: Benefits of Site-Specific Recombinase Technology and Future Directions

This review discusses the application of site-specific recombination-based technologies and ways they can be used to improve genetic engineering in both basic and applied research. The technology enables the removal of extraneous DNA such as selectable markers (e.g., antibiotic and herbicide resistance genes) from the genome. The use of site-specific recombination technology in plant genome manipulation has also been demonstrated to effectively resolve complex transgene insertions to single copies, to remove unwanted plasmid backbone DNA, and to precisely insert transgenes into genomic target sites with known expression attributes. In plants, this technology speeds the transition from laboratory manipulation to field production, an important benefit for both the industry and the general public. From an application point of view, the number and types of recombinases available along with the innovative strategies being developed offer a multitude of ways for genome manipulation including efficient single-copy targeted integration, sequential gene stacking, and complete transgene removal from the pollen and/or seed. These strategies are applicable to both commonly used transformation methods, biolistics and *Agrobacterium*, and thus can be used in any transformable plant. The final question to address is how these methods will change genomic engineering in the future. Up to now, the technology behind site-specific recombination has mainly been evolving by changes in the timing of recombinase action and in the placement of recognition sequences, as discussed above. More recently, with the demonstration that these enzyme

systems can function in nonnative cell environments, the emphasis has shifted toward modifications of the enzymes themselves to enhance their efficiencies in target organisms and improve the specificities of the recombinase for natural and mutated recognition sites. In 1998 Buchholz performed error-prone PCR on the *Flp* gene, inducing a series of mutations that increased the enzyme's thermo-stability and relative rates of activity. Second generation modifications included mutagenesis for increased activity (Buchholz et al. 1996), recognition site alteration (Abi-Ghanem et al. 2013; Keravala et al. 2008; Sclimenti et al. 2001), codon optimization for increased rates of translation and thus higher cellular concentration (Blechl et al. 2012; Moon et al. 2011; Raymond and Soriano 2007), and addition of nuclear localization signals (NLS) to increase effective concentrations of the recombinase in the nucleus (Andreas et al. 2002; Blechl et al. 2012). While these modifications have been effective in enhancing activity or altering binding site recognition, they did not fundamentally change the structure of the recombinase or allow it to recognize off-site targets. Enter the third generation of site-specific recombinases or "designer" recombinases. The capacity to excise or integrate DNA without the gain or loss of nucleotides has been paired with the both zinc fingers (Gaj et al. 2013; Prorocic et al. 2011; Proudfoot et al. 2011) and TALENs (Mercer et al. 2012) to yield adjustable binding site recognition. In theory, these third generation recombinases would have the capacity to target most any region of the genome in a site-specific manner. This means that DNA insertion into genomes would no longer be subject to the randomness and rearrangements of illegitimate recombination nor depend on the host cell's own inefficient homologous recombination machinery. Use of recombinase technology has become routine in genomic engineering and is becoming an integral part of crop biotechnology, yielding improved crops that rely less on pesticides and fertilizers and produce high quality, abundant foods that benefit the farm economy. Further refinements to this technology have the potential to improve public acceptance for genetically engineered crops in the future as plants produced with this expertise may raise fewer public and regulatory concerns about the presence of markers genes, novel proteins and potential unintended effects of extraneous DNA.

Acknowledgments Research was funded by USDA-ARS project 5325-21000-020-00D and by the Biotechnology Risk Assessment Program competitive grant 2010-33522-21773 from the USDA - National Institute of Food and Agriculture. USDA is an equal opportunity provider and employer.

References

Abi-Ghanem J, Chusainow J, Karimova M, Spiegel C, Hofmann-Sieber H, Hauber J, Buchholz F, Pisabarro MT (2013) Engineering of a target site-specific recombinase by a combined evolution- and structure-guided approach. Nucleic Acids Res 41:2394–2403

Akbudak MA, More AB, Nandy S, Srivastava V (2010) Dosage-dependent gene expression from direct repeat locus in rice developed by site-specific gene integration. Mol Biotechnol 45:15–23

Akbudak MA, Nicholson S, Srivastava V (2013) Suppression of Arabidopsis genes by terminator-less transgene constructs. Plant Biotechnol Rep 7:415–424

Albert H, Dale EC, Lee E, Ow DW (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J 7:649–659

Anderson RP, Voziyanova E, Voziyanov Y (2012) Flp and Cre expressed from Flp-2A-Cre and Flp-IRES-Cre transcription units mediate the highest level of dual recombinase-mediated cassette exchange. Nucleic Acids Res 40:e62

Andreas S, Schwenk F, Küter-Luks B, Faust N, Kühn R (2002) Enhanced efficiency through nuclear localization signal fusion on phage ϕC31-integrase: activity comparison with Cre and FLPe recombinase in mammalian cells. Nucleic Acids Res 30:2299–2306

Araki K, Araki M, Miyazaki J, Vassalli P (1995) Site-specific recombination of a transgene in fertilized eggs by transient expression of Cre recombinase. Proc Natl Acad Sci U S A 92: 160–164

Araki K, Okada Y, Araki M, Yamamura K (2010) Comparative analysis of right element mutant lox sites on recombination efficiency in embryonic stem cells. BMC Biotechnol 10:29

Ballester A, Cervera M, Pena L (2007) Efficient production of transgenic citrus plants using isopentenyl transferase positive selection and removal of the marker gene by site-specific recombination. Plant Cell Rep 26:39–45

Baubonis W, Sauer B (1993) Genomic targeting with purified Cre recombinase. Nucleic Acids Res 21:2025–2029

Belteki G, Gertsenstein M, Ow DW, Nagy A (2003) Site-specific cassette exchange and germline transmission with mouse ES cells expressing [phgr]C31 integrase. Nat Biotechnol 21:321–324

Blechl A, Lin J, Shao M, Thilmony R, Thomson J (2012) The Bxb1 recombinase mediates site-specific deletion in transgenic wheat. Plant Mol Biol Report 30:1357–1366

Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D (2013) Amplifying genetic logic gates. Science 340:599–603

Bouhassira E, Westerman K, Leboulch P (1997) Transcriptional behavior of LCR enhancer elements integrated at the same chromosomal locus by recombinase-mediated cassette exchange. Blood 90:3332–3344

Brocard J, Feil R, Chambon P, Metzger D (1998) A chimeric Cre recombinase inducible by synthetic, but not by natural ligands of the glucocorticoid receptor. Nucleic Acids Res 26: 4086–4090

Brookes G, Barfoot P (2012) The income and production effects of biotech crops globally 1996-2010. GM Crops Food 3:265–272

Buchholz F, Ringrose L, Angrand P-O, Rossi F, Stewart AF (1996) Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. Nucleic Acids Res 24:4256–4262

Burdon TG, Wall RJ (1992) Fate of microinjected genes in preimplantation mouse embryos. Mol Reprod Dev 33:436–442

Burgess SM, Kleckner N (1999) Collisions between yeast chromosomal loci in vivo are governed by three layers of organization. Genes Dev 13:1871–1883

Cao M-X, Huang J-Q, Yao Q-H, Liu S-J, Wang C-L, Wei Z-M (2006) Site-specific DNA excision in transgenic rice with a cell-permeable Cre recombinase. Mol Biotechnol 32:55–63

Carlson SR, Rudgers GW, Zieler H, Mach JM, Luo S, Grunden E, Krol C, Copenhaver GP, Preuss D (2007) Meiotic transmission of an in vitro assembled autonomous maize minichromosome. PLoS Genet 3:e179

Chawla R, Ariza-Nieto M, Wilson A, Moore S, Srivastava V (2006) Transgene expression produced by biolistic-mediated, site-specific gene integration is consistently inherited by the subsequent generations. Plant Biotechnol J 4:209–218

Chen K-S, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. Nat Genet 17:154–163

Choi S, Begum D, Koshinsky H, Ow D, Wing R (2000) A new approach for the identification and cloning of genes: the pBACwich system using Cre/lox site-specific recombination. Nucleic Acids Res 28:E19

Chong-Pérez B, Kosky RG, Reyes M, Rojas L, Ocaña B, Tejeda M, Pérez B, Angenon G (2012) Heat shock induced excision of selectable marker genes in transgenic banana by the Cre-lox site-specific recombination system. J Biotechnol 159:265–273

Christ N, Dröge P (2002) Genetic manipulation of mouse embryonic stem cells by mutant λ integrase. Genesis 32:203–208

Cuellar W, Gaudin A, Solorzano D, Casas A, Nopo L, Chudalayandi P, Medrano G, Kreuze J, Ghislain M (2006) Self-excision of the antibiotic resistance gene nptII using a heat inducible Cre-loxP system from transgenic potato. Plant Mol Biol 62:71–82

Dafhnis-Calas F, Xu Z, Haines S, Malla SK, Smith MCM, Brown WRA (2005) Iterative in vivo assembly of large and complex transgenes by combining the activities of φC31 integrase and Cre recombinase. Nucleic Acids Res 33:e189

Dale EC, Ow DW (1991) Gene transfer with subsequent removal of the selection gene from the host genome. Proc Natl Acad Sci U S A 88:10558–10562

Danielian PS, Muccino D, Rowitch DH, Michael SK, McMahon AP (1998) Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. Curr Biol 8:1323–S1322

Day CD, Lee E, Kobayashi J, Holappa LD, Albert H, Ow DW (2000) Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. Genes Dev 14:2869–2880

De Paepe A, De Buck S, Nolf J, Van Lerberge E, Depicker A (2013) Site-specific T–DNA integration in Arabidopsis thaliana mediated by the combined action of CRE recombinase and ϕC31 integrase. Plant J 75:172–184

de Wit T, Drabek D, Grosveld F (1998) Microinjection of Cre recombinase RNA induces site-specific recombination of a transgene in mouse oocytes. Nucleic Acids Res 26:676–678

Diaz V, Servert P, Prieto I, Gonzalez MA, Martinez-A C, Alonso JC, Bernad A (2001) New insights into host factor requirements for prokaryotic beta-recombinase-mediated reactions in mammalian cells. J Biol Chem 276:16257–16264

Djukanovic V, Lenderts B, Bidney D, Lyznik LA (2008) A Cre::FLP fusion protein recombines FRT or loxP sites in transgenic maize plants†. Plant Biotechnol J 6:770–781

Enyeart PJ, Chirieleison SM, Dao MN, Perutka J, Quandt EM, Yao J, Whitt JT, Keatinge-Clay AT, Lambowitz AM, Ellington AD (2013) Generalized bacterial genome editing using mobile group II introns and Cre-lox. Mol Syst Biol 9

Feil R, Brocard J, Mascrez B, LeMeur M, Metzger D, Chambon P (1996) Ligand-activated site-specific recombination in mice. Proc Natl Acad Sci U S A 93:10887–10890

Feng Y-Q, Seibler J, Alami R, Eisen A, Westerman KA, Leboulch P, Fiering S, Bouhassira EE (1999) Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. J Mol Biol 292:779–785

Fiering S, Kim CG, Epner EM, Groudine M (1993) An "in-out" strategy using gene targeting and FLP recombinase for the functional dissection of complex DNA regulatory elements: analysis of the beta-globin locus control region. Proc Natl Acad Sci U S A 90:8469–8473

Fladung M, Becker D (2011) Targeted integration and removal of transgenes in hybrid aspen (Populus tremula L. × P. tremuloides Michx.) using site-specific recombination systems. Plant Biol 13:223–223

Floss T, Schnutgen F (2008) Conditional gene trapping using the FLEx system. In: Davis G, Kayser K (eds) Chromosomal mutagenesis. Humana, Totowa, pp 127–138

Gaeta RT, Masonbrink RE, Krishnaswamy L, Zhao C, Birchler JA (2012) Synthetic chromosome platforms in plants. Annu Rev Plant Biol 63:307–330

Gaj T, Mercer AC, Sirk SJ, Smith HL, Barbas CF (2013) A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic Acids Res 41:3937–3946

Gerlitz M, Hrabak O, Schwab H (1990) Partitioning of broad-host-range plasmid RP4 is a complex system involving site-specific recombination. J Bacteriol 172:6194–6203

Ghosh P, Kim AI, Hatfull GF (2003) The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. Mol Cell 12:1101–1111

Ghosh P, Wasil LR, Hatfull GF (2006) Control of phage Bxb1 excision by a novel recombination directionality factor. PLoS Biol 4:e186
Gidoni D, Srivastava V, Carmi N (2008) Site-specific excisional recombination strategies for elimination of undesirable transgenes from crop plants. In Vitro Cell Dev Biol Plant 44:457–467
Gilbertson L, Dioh W, Addae P, Ekena J, Keithly G, Neuman M, Peschke V, Petersen M, Samuelson C, Subbarao S, Wei L, Zhang W, Barton K (2003) Cre/lox mediated marker gene excision in transgenic crop plants. In: Vasil I (ed) Plant biotechnology 2002 and beyond. Springer, Netherlands, pp 225–228
Gils M, Marillonnet S, Werner S, Grützner R, Giritch A, Engler C, Schachschneider R, Klimyuk V, Gleba Y (2008) A novel hybrid seed system for plants. Plant Biotechnol J 6:226–235
Golic KG, Lindquist S (1989) The FLP recombinase of yeast catalyzes site-specific recombination in the drosophila genome. Cell 59:499–509
Gottfried P, Lotan O, Kolot M, Maslenin L, Bendov R, Gorovits R, Yesodi V, Yagil E, Rosner A (2005) Site-specific recombination in Arabidopsis plants promoted by the Integrase protein of coliphage HK022. Plant Mol Biol 57:435–444
Grindley NDF, Whiteson KL, Rice PA (2006) Mechanisms of site-specific recombination. Annu Rev Biochem 75:567–605
Groth AC, Olivares EC, Thyagarajan B, Calos MP (2000) A phage integrase directs efficient site-specific integration in human cells. Proc Natl Acad Sci U S A 97:5995–6000
Hare PD, Chua N-H (2002) Excision of selectable marker genes from transgenic plants. Nat Biotechnol 20:575–580
Henikoff S (1998) Conspiracy of silence among repeated transgenes. Bioessays 20:532–535
Herault Y, Rassoulzadegan M, Cuzin F, Duboule D (1998) Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). Nat Genet 20:381–384
Hoa TTC, Bong BB, Huq E, Hodges TK (2002) Cre/lox site-specific recombination controls the excision of a transgene from the rice genome. Theor Appl Genet 104:518–525
Hobbs SA, Kpodar P, DeLong CO (1990) The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. Plant Mol Biol 15:851–864
Hoess RH, Abremski K (1984) Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. Proc Natl Acad Sci U S A 81:1026–1029
Hoess R, Wierzbicki A, Abremski K (1986) The role of the loxP spacer region in P1 site-specific recombination. Nucleic Acids Res 14:2287–2300
Hoff T, Schnorr K, Mundy J (2001) A recombinase-mediated transcriptional induction system in transgenic plants. Plant Mol Biol 45:41–49
Hohn B, Levy AA, Puchta H (2001) Elimination of selection markers from transgenic plants. Curr Opin Biotechnol 12:139–143
Horn C, Handler AM (2005) Site-specific genomic targeting in Drosophila. Proc Natl Acad Sci U S A 102:12483–12488
Houben A, Dawe RK, Jiang J, Schubert I (2008) Engineered plant minichromosomes: a bottom-up success? Plant Cell Online 20:8–10
Hu Q, Kononowicz-Hodges H, Nelson-Vasilchik K, Viola D, Zeng P, Liu H, Kausch AP, Chandlee JM, Hodges TK, Luo H (2008) FLP recombinase-mediated site-specific recombination in rice. Plant Biotechnol J 6:176–188
Huang L-C, Wood EA, MCox M (1991) A bacterial model system for chromosomal targeting. Nucleic Acids Res 19:443–448
Huang J, Zhou W, Dong W, Watson AM, Hong Y (2009) Directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. Proc Natl Acad Sci U S A 106: 8284–8289
Hunter NL, Awatramani RB, Farley FW, Dymecki SM (2005) Ligand-activated Flpe for temporally regulated gene modifications. Genesis 41:99–109
Iglesias VA, Moscone EA, Papp I, Neuhuber F, Michalowski S, Phelan T, Spiker S, Matzke M, Matzke AJ (1997) Molecular and cytogenetic analyses of stably and unstably expressed transgene loci in tobacco. Plant Cell Online 9:1251–1264

Iyer L, Kumpatla S, Chandrasekharan M, Hall T (2000) Transgene silencing in monocots. Plant Mol Biol 43:323–346

Jia H, Pang Y, Chen X, Fang R (2006) Removal of the selectable marker gene from transgenic tobacco plants by expression of Cre recombinase from a tobacco mosaic virus vector through agroinfection. Transgenic Res 15:375–384

Kapusi E, Kempe K, Rubtsova M, Kumlehn J, Gils M (2012) phiC31 integrase-mediated site-specific recombination in barley. PLoS One 7:e45353

Kellendonk C, Tronche F, Monaghan A-P, Angrand P-O, Stewart F, Schütz G (1996) Regulation of Cre recombinase activity by the synthetic steroid RU 486. Nucleic Acids Res 24:1404–1411

Kempe K, Rubtsova M, Berger C, Kumlehn J, Schollmeier C, Gils M (2010) Transgene excision from wheat chromosomes by phage phiC31 integrase. Plant Mol Biol 72:673–687

Keravala A, Groth A, Jarrahian S, Thyagarajan B, Hoyt J, Kirby P, Calos M (2006) A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. Mol Genet Genomics 276:135–146

Keravala A, Lee S, Thyagarajan B, Olivares EC, Gabrovsky VE, Woodard LE, Calos MP (2008) Mutational derivatives of PhiC31 integrase with increased efficiency and specificity. Mol Ther 17:112–120

Kerbach S, Lorz H, Becker D (2005) Site-specific recombination in Zea mays. Theor Appl Genet 111:1608–1616

Khaleel T, Younger E, McEwan AR, Varghese AS, Smith MCM (2011) A phage protein that binds φC31 integrase to switch its directionality. Mol Microbiol 80:1450–1463

Kholodii G (2001) The shuffling function of resolvases. Gene 269:121–130

Kim AI, Ghosh P, Aaron MA, Bibb LA, Jain S, Hatfull GF (2003) Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. Mol Microbiol 50:463–473

Kittiwongwattana C, Lutz K, Clark M, Maliga P (2007) Plastid marker gene excision by the phiC31 phage site-specific recombinase. Plant Mol Biol 64:137–143

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

Kolb AF, Siddell SG (1997) Genomic targeting of a bicistronic DNA fragment by Cre-mediated site-specific recombination. Gene 203:209–216

Kolot M, Silberstein N, Yagil E (1999) Site-specific recombination in mammalian cells expressing the Int recombinase of bacteriophage HK022—site-specific recombination in mammalian cells promoted by a phage integrase. Mol Biol Rep 26:207–213

Kondrak M, Meer I, Banfalvi Z (2006) Generation of marker- and backbone-free transgenic potatoes by site-specific recombination and a bi-functional marker gene in a non-regular one-border agrobacterium transformation vector. Transgenic Res 15:729–737

Kopertekh L, Schiemann J (2005) Agroinfiltration as a tool for transient expression of cre recombinase in vivo. Transgenic Res 14:793–798

Kopertekh L, Juttner G, Schiemann J (2004) Site-specific recombination induced in transgenic plants by PVX virus vector expressing bacteriophage P1 recombinase. Plant Sci 166:485–492

Kopertekh L, Schulze K, Frolov A, Strack D, Broer I, Schiemann J (2010) Cre-mediated seed-specific transgene excision in tobacco. Plant Mol Biol 72:597–605

Korenberg JR, Chen XN, Schipper R, Sun Z, Gonsky R, Gerwehr S, Carpenter N, Daumer C, Dignan P, Disteche C (1994) Down syndrome phenotypes: the consequences of chromosomal imbalance. Proc Natl Acad Sci U S A 91:4997–5001

Koshinsky HA, Lee E, Ow DW (2000) Cre-lox site-specific recombination between Arabidopsis and tobacco chromosomes. Plant J 23:715–722

Lauth M, Spreafico F, Dethleffsen K, Meyer M (2002) Stable and efficient cassette exchange under non-selectable conditions by combined use of two site-specific recombinases. Nucleic Acids Res 30:e115

Lee G, Saito I (1998) Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. Gene 216:55–65

Lewandoski M, Martin GR (1997) Cre-mediated chromosome loss in mice. Nat Genet 17:223–225

Li Z, Xing A, Moon B, Burgoyne S, Guida A, Liang H, Lee C, Caster C, Barton J, Klein T, Falco S (2007) A Cre/loxP-mediated self-activating gene excision system to produce marker gene free transgenic soybean plants. Plant Mol Biol 65:329–341

Li Z, Xing A, Moon BP, McCardell RP, Mills K, Falco SC (2009) Site-specific integration of transgenes in soybean via recombinase-mediated DNA cassette exchange. Plant Physiol 151:1087–1095

Li Z, Moon BP, Xing A, Liu Z-B, McCardell RP, Damude HG, Falco SC (2010) Stacking multiple transgenes at a selected genomic site via repeated recombinase-mediated DNA cassette exchanges. Plant Physiol 154:622–631

Liu H-K, Yang C, Wei Z-M (2005) Heat shock-regulated site-specific excision of extraneous DNA in transgenic plants. Plant Sci 168:997–1003

Logie C, Stewart AF (1995) Ligand-regulated site-specific recombination. Proc Natl Acad Sci U S A 92:5940–5944

Louwerse JD, van Lier MCM, van der Steen DM, de Vlaam CMT, Hooykaas PJJ, Vergunst AC (2007) Stable recombinase-mediated cassette exchange in arabidopsis using Agrobacterium tumefaciens. Plant Physiol 145:1282–1293

Luff B, Pawlowski L, Bender J (1999) An inverted repeat triggers cytosine methylation of identical sequences in arabidopsis. Mol Cell 3:505–511

Luo K, Duan H, Zhao D, Zheng X, Deng W, Chen Y, Stewart CN, McAvoy R, Jiang X, Wu Y, He A, Pei Y, Li Y (2007) 'GM-gene-deletor': fused loxP-FRT recognition sequences dramatically improve the efficiency of FLP or CRE recombinase on transgene excision from pollen and seed of tobacco plants. Plant Biotechnol J 5:263–374

Lutz KA, Corneille S, Azhagiri AK, Svab Z, Maliga P (2004) A novel approach to plastid transformation utilizes the phiC31 phage integrase. Plant J 37:906–913

Lyznik LA, Rao KV, Hodges TK (1996) FLP-mediated recombination of FRT sites in the maize genome. Nucleic Acids Res 24:3784–3789

Malchin N, Molotsky T, Yagil E, Kotlyar AB, Kolot M (2008) Molecular analysis of recombinase-mediated cassette exchange reactions catalyzed by integrase of coliphage HK022. Res Microbiol 159:663–670

Maqbool S, Christou P (1999) Multiple traits of agronomic importance in transgenic indica rice plants: analysis of transgene integration patterns, expression levels and stability. Mol Breed 5:471–480

Marenkova TV, Loginova DB, Deineko EV (2012) Mosaic patterns of transgene expression in plants. Russ J Genet 48:249–260

Martinez de Alba AE, Elvira-Matelot E, Vaucheret H (2013) Gene silencing in plants: a diversity of pathways. Biochim Biophys Acta 1829:1300–1308

Matsunaga E, Sugita K, Ebinuma H (2002) Asexual production of selectable marker-free transgenic woody plants, vegetatively propagated species. Mol Breed 10:95–106

McClintock B (1938) The production of homozygous deficient tissues with mutant characteristics by means of the aberrant mitotic behavior of ring-shaped chromosomes. Genetics 23:315–376

Medberry SL, Dale E, Qin M, Ow DW (1995) Intra-chromosomal rearrangements generated by Cre-lox site-specific recombination. Nucleic Acids Res 23:485–490

Mercer AC, Gaj T, Fuller RP, Barbas CF (2012) Chimeric TALE recombinases with programmable DNA sequence specificity. Nucleic Acids Res 40:11163–11172

Metzger D, Clifford J, Chiba H, Chambon P (1995) Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. Proc Natl Acad Sci U S A 92:6991–6995

Meyer P (2000) Transcriptional transgene silencing and chromatin components. Plant Mol Biol 43:221–234

Mlynárová L, Conner AJ, Nap J-P (2006) Directed microspore-specific recombination of transgenic alleles to prevent pollen-mediated transmission of transgenes. Plant Biotechnol J 4:445–452

Monetti C, Nishino K, Biechele S, Zhang P, Baba T, Woltjen K, Nagy A (2011) PhiC31 integrase facilitates genetic approaches combining multiple recombinases. Methods 53:380–385

Moon H, Abercrombie L, Eda S, Blanvillain R, Thomson J, Ow D, Stewart CN Jr (2011) Transgene excision in pollen using a codon optimized serine resolvase CinH-RS2 site-specific recombination system. Plant Mol Biol 75:621–631

Mouw KW, Rowland S-J, Gajjar MM, Boocock MR, Stark WM, Rice PA (2008) Architecture of a serine recombinase-DNA regulatory complex. Mol Cell 30:145–155

Nandy S, Srivastava V (2011) Site-specific gene integration in rice genome mediated by the FLP–FRT recombination system. Plant Biotechnol J 9:713–721

Nanto K, Ebinuma H (2008) Marker-free site-specific integration plants. Transgenic Res 17:337–344

Nanto K, Yamada-Watanabe K, Ebinuma H (2005) Agrobacterium-mediated RMCE approach for gene replacement. Plant Biotechnol J 3:203–214

Nanto K, Sato K, Katayama Y, Ebinuma H (2009) Expression of a transgene exchanged by the recombinase-mediated cassette exchange (RMCE) method in plants. Plant Cell Rep 28: 777–785

Obayashi H, Kawabe Y, Makitsubo H, Watanabe R, Kameyama Y, Huang S, Takenouchi Y, Ito A, Kamihira M (2012) Accumulative gene integration into a pre-determined site using Cre/loxP. J Biosci Bioeng 113:381–388

O'Gorman S, Fox D, Wahl G (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science 251:1351–1355

Ohtsuka M, Ogiwara S, Miura H, Mizutani A, Warita T, Sato M, Imai K, Hozumi K, Sato T, Tanaka M, Kimura M, Inoko H (2010) Pronuclear injection-based mouse targeted transgenesis for reproducible and highly efficient transgene expression. Nucleic Acids Res 38:e198

Olivares EC, Hollis RP, Calos MP (2001) Phage R4 integrase mediates site-specific integration in human cells. Gene 278:167–176

Olson L, Tien J, South S, Reeves R (2005) Long-range chromosomal engineering is more efficient in vitro than in vitro. Transgenic Res 14:325–332

Onouchi H, Yokoi K, Machida C, Matsuzaki H, Oshima Y, Matsuoka K, Nakamura K, Machida Y (1991) Operation of an efficient site-specific recombination system of Zygosaccharomyces rouxii in tobacco cells. Nucleic Acids Res 19:6373–6378

Ou H-L, Huang Y, Qu L-J, Xu M, Yan J-B, Ren Z-R, Huang S-Z, Zeng Y-T (2009) A ϕC31 integrase-mediated integration hotspot in favor of transgene expression exists in the bovine genome. FEBS J 276:155–163

Ow DW (2005) 2004 SIVB congress symposium proceeding: transgene management via multiple site-specific recombination systems. In Vitro Cell Dev Biol Plant 41:213–219

Peitz M, Pfannkuche K, Rajewsky K, Edenhofer F (2002) Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. Proc Natl Acad Sci U S A 99:4489–4494

Prorocic MM, Wenlong D, Olorunniji FJ, Akopian A, Schloetel J-G, Hannigan A, McPherson AL, Stark WM (2011) Zinc-finger recombinase activities in vitro. Nucleic Acids Res 39:9316–9328

Proudfoot C, McPherson AL, Kolb AF, Stark WM (2011) Zinc finger recombinases with adaptable DNA sequence specificity. PLoS One 6:e19537

Qin M, Bayley C, Stockton T, Ow DW (1994) Cre recombinase-mediated site-specific recombination between plant chromosomes. Proc Natl Acad Sci U S A 91:1706–1710

Radhakrishnan P, Srivastava V (2005) Utility of the FLP-FRT recombination system for genetic manipulation of rice. Plant Cell Rep 23:721–726

Ramirez-Solis R, Liu P, Bradley A (1995) Chromosome engineering in mice. Nature 378: 720–724

Raymond CS, Soriano P (2007) High-efficiency FLP and φC31 site-specific recombination in mammalian cells. PLoS One 2:e162

Rubtsova M, Kempe K, Gils A, Ismagul A, Weyen J, Gils M (2008) Expression of active Streptomyces phage phiC31 integrase in transgenic wheat plants. Plant Cell Rep 27: 1821–1831

Russell S, Hoopes J, Odell J (1992) Directed excision of a transgene from the plant genome. Mol Gen Genet 234:49–59

Russell J, Chang D, Tretiakova A, Padidam M (2006) Phage Bxb1 integrase mediates highly efficient. Biotechniques 40:460–464
Sang Y, Millwood RJ, Neal Stewart Jr C (2013) Gene use restriction technologies for transgenic plant bioconfinement. Plant Biotechnol J 11:649–658
Sarkis GJ, Murley LL, Leschziner AE, Boocock MR, Stark WM, Grindley NDF (2001) A model for the gamma delta resolvase synaptic complex. Mol Cell 8:623–631
Sauer B (1996) Multiplex crellox recombination permits selective site-specific DNA targeting to both a natural and an engineered site in the yeast genome. Nucleic Acids Res 24:4608–4613
Sauer B, Henderson N (1990) Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. New Biol 2:441–449
Schlake T, Bode J (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. Biochemistry 33:12746–12751
Schwikardi M, Dröge P (2000) Site-specific recombination in mammalian cells catalyzed by γδ resolvase mutants: implications for the topology of episomal DNA. FEBS Lett 471:147–150
Sclimenti CR, Thyagarajan B, Calos MP (2001) Directed evolution of a recombinase for improved genomic integration at a native human sequence. Nucleic Acids Res 29:5044–5051
Seibler J, Bode J (1997) Double-reciprocal crossover mediated by FLP-recombinase: a concept and an assay. Biochemistry 36:1740–1747
Semprini S, Troup TJ, Kotelevtseva N, King K, Davis JRE, Mullins LJ, Chapman KE, Dunbar DR, Mullins JJ (2007) Cryptic loxP sites in mammalian genomes: genome-wide distribution and relevance for the efficiency of BAC/PAC recombineering techniques. Nucleic Acids Res 35:1402–1410
Senecoff JF, Rossmeissl PJ, Cox MM (1988) DNA recognition by the FLP recombinase of the yeast 2 μ plasmid: a mutational analysis of the FLP binding site. J Mol Biol 201:405–421
Shao M, Kumar S, Thomson JG (2013) Precise excision of plastid DNA by the large serine recombinase Bxb1. Plant Biotechnol J
Siuti P, Yazbek J, Lu TK (2013) Synthetic circuits integrating logic and memory in living cells. Nat Biotechnol 31:448–452
Smith MCM, Brown WRA, McEwan AR, Rowley PA (2010) Site-specific recombination by φC31 integrase and other large serine recombinases. Biochem Soc Trans 38:388–394
Sreekala C, Wu L, Gu K, Wang D, Tian D, Yin Z (2005) Excision of a selectable marker in transgenic rice (Oryza sativa L.) using a chemically regulated Cre/loxP system. Plant Cell Rep 24:86–94
Srivastava V, Gidoni D (2010) Site-specific gene integration technologies for crop improvement. In Vitro Cell Dev Biol Plant 46:219–232
Srivastava V, Ow DW (2001) Single-copy primary transformants of maize obtained through the co-introduction of a recombinase-expressing construct. Plant Mol Biol 46:561–566
Srivastava V, Ow D (2002) Biolistic mediated site-specific integration in rice. Mol Breed 8:345–350
Srivastava V, Ow DW (2003) Rare instances of Cre-mediated deletion product maintained in transgenic wheat. Plant Mol Biol 52:661–668
Srivastava V, Ow DW (2004) Marker-free site-specific gene integration in plants. Trends Biotechnol 22:627–629
Srivastava V, Anderson OD, Ow DW (1999) Single-copy transgenic wheat generated through the resolution of complex integration patterns. Proc Natl Acad Sci U S A 96:11117–11121
Srivastava V, Ariza-Nieto M, Wilson A (2004) Cre-mediated site-specific gene integration for consistent transgene expression in rice. Plant Biotechnol J 2:169–179
Stoll SM, Ginsburg DS, Calos MP (2002) Phage TP901-1 site-specific integrase functions in human cells. J Bacteriol 184:3657–3663
Street VA, Goldy JD, Golden AS, Tempel BL, Bird TD, Chance PF (2002) Mapping of charcot-marie-tooth disease type 1C to chromosome 16p identifies a novel locus for demyelinating neuropathies. Am J Hum Genet 70:244–250
Takahashi T, Naito S, Komeda Y (1992) Isolation and analysis of the expression of two genes for the 81-kilodalton heat-shock proteins from Arabidopsis. Plant Physiol 99:383–390

Takata Y, Kondo S, Goda N, Kanegae Y, Saito I (2011) Comparison of efficiency between FLPe and Cre for recombinase-mediated cassette exchange in vitro and in adenovirus vector production. Genes Cells 16:765–777

Thomason L, Calendar R, Ow D (2001) Gene insertion and replacement in Schizosaccharomyces pombe mediated by the Streptomyces bacteriophage C31 site-specific recombination system. Mol Genet Genomics 265:1031–1038

Thomson JG, Ow DW (2006) Site-specific recombination systems for the genetic manipulation of eukaryotic genomes. Genesis 44:465–476

Thomson JG, Rucker EB, Piedrahita JA (2003) Mutational analysis of loxP sites for efficient Cre-mediated insertion into genomic DNA. Genesis 36:162–167

Thomson JG, Yau YY, Blanvillain R, Chiniquy D, Thilmony R, Ow DW (2009) ParA resolvase catalyzes site-specific excision of DNA from the *Arabidopsis* genome. Transgenic Res 18:237–248

Thomson JG, Chan R, Thilmony R, Yau YY, Ow DW (2010) PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the Arabidopsis genome. BMC Biotechnol 10:17

Thomson J, Chan R, Smith J, Thilmony R, Yau Y-Y, Wang Y, Ow D (2012) The Bxb1 recombination system demonstrates heritable transmission of site-specific excision in Arabidopsis. BMC Biotechnol 12:9

Thorpe HM, Smith MCM (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc Natl Acad Sci U S A 95:5505–5510

Thorpe HM, Wilson SE, Smith MCM (2000) Control of directionality in the site-specific recombination system of the Streptomyces phage φC31. Mol Microbiol 38:232–241

Thyagarajan B, Guimaraes MJ, Groth AC, Calos MP (2000) Mammalian genomes contain active recombinase recognition sites. Gene 244:47–54

Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol Cell Biol 21: 3926–3934

Torres R, García A, Payá M, Ramirez JC (2011) Non-integrative lentivirus drives high-frequency cre-mediated cassette exchange in human cells. PLoS One 6:e19794

Tsanov KM, Nishi Y, Peterson KA, Liu J, Baetscher M, McMahon AP (2012) An embryonic stem cell-based system for rapid analysis of transcriptional enhancers. Genesis 50:443–450

Turan S, Zehe C, Kuehle J, Qiao J, Bode J (2013) Recombinase-mediated cassette exchange (RMCE)—a rapidly-expanding toolbox for targeted genomic modifications. Gene 515:1–27

Uemura M, Niwa Y, Kakazu N, Adachi N, Kinoshita K (2010) Chromosomal manipulation by site-specific recombinases and fluorescent protein-based vectors. PLoS One 5:e9846

Van Deursen J, Fornerod M, Van Rees B, Grosveld G (1995) Cre-mediated site-specific translocation between nonhomologous mouse chromosomes. Proc Natl Acad Sci U S A 92:7376–7380

Vergunst A, Hooykaas P (1998) Cre/lox-mediated site-specific integration of Agrobacterium T-DNA in Arabidopsis thaliana by transient expression of cre. Plant Mol Biol 38:393–406

Vergunst A, Jansen L, Hooykaas P (1998) Site-specific integration of Agrobacterium T-DNA in Arabidopsis thaliana mediated by Cre recombinase. Nucleic Acids Res 26:2729–2734

Vergunst AC, Schrammeijer B, den Dulk-Ras A, de Vlaam CMT, Regensburg-Tuïnk TJ, Hooykaas PJ (2000) VirB/D4-dependent protein translocation from Agrobacterium into plant cells. Science 290:979–982

Verweire D, Verleyen K, De Buck S, Claeys M, Angenon G (2007) Marker-free transgenic plants through genetically programmed auto-excision. Plant Physiol 145:1220–1231

Wang K, Moeller L (2008) Engineering with precision: tools for the new generation of transgenic crops. Bioscience 58:391–401

Wang Y, Chen B, Hu Y, Li J, Lin Z (2005) Inducible excision of selectable marker gene from transgenic plants by the Cre/lox site-specific recombination system. Transgenic Res 14: 605–614

Wang Y, Yau Y-Y, Perkins-Balding D, Thomson J (2011) Recombinase technology: applications and possibilities. Plant Cell Rep 30:267–285

Wunderlich FT, Wildner H, Rajewsky K, Edenhofer F (2001) New variants of inducible Cre recombinase: a novel mutant of Cre-PR fusion protein exhibits enhanced sensitivity and an expanded range of inducibility. Nucleic Acids Res 29:e47

Wurdig J, Flachowsky H, Hanke M-V (2013) Studies on heat shock induction and transgene expression in order to optimize the Flp/FRT recombinase system in apple (Malus X domestica Borkh.). Plant Cell Tiss Organ Cult 115:457–467

Yau Y, Wang Y, Thomson J, Ow D (2011) Method for Bxb1-mediated site-specific integration in planta. Methods Mol Biol 701:147–166

Zhang W, Subbarao S, Addae P, Shen A, Armstrong C, Peschke V, Gilbertson L (2003) Cre/lox-mediated marker gene excision in transgenic maize (Zea mays L.) plants. Theor Appl Genet 107:1157–1168

Zhang Y, Li H, Ouyang B, Lu Y, Ye Z (2006) Chemical-induced autoexcision of selectable markers in elite tomato plants transformed with a gene conferring resistance to lepidopteran insects. Biotechnol Lett 28:1247–1253

Zhang L, Wang L, Wang J, Ou X, Zhao G, Ding X (2010) DNA cleavage is independent of synapsis during streptomyces phage φBT1 integrase-mediated site-specific recombination. J Mol Cell Biol 2:264–275

Zheng B, Sage M, Cai W-W, Thompson DM, Tavsanli BC, Cheah Y-C, Bradley A (1999) Engineering a mouse balancer chromosome. Nat Genet 22:375–378

Zou X, Peng A, Xu L, Liu X, Lei T, Yao L, He Y, Chen S (2013) Efficient auto-excision of a selectable marker gene from transgenic citrus by combining the Cre/loxP system and ipt selection. Plant Cell Rep 32:1601–1613

Zuo J, Niu Q-W, Moller SG, Chua N-H (2001) Chemical-regulated, site-specific DNA excision in transgenic plants. Nat Biotechnol 19:157–161

Chapter 8
Developing CRISPR Technology in Major Crop Plants

Kunling Chen and Caixia Gao

Abstract Targeted genome modification (TGM) by sequence-specific nucleases (SSNs) is a powerful tool for elucidating gene function and improving crops. Very recently, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system-based RNA-guided endonucleases have been added to the SSN toolbox. TGMs generated by this system rely on a synthetic single guide RNA (sgRNA) to direct the Cas9 protein to cleave a predetermined DNA sequence. Unlike previous SSNs, CRISPR provides a simple, cost-effective and versatile approach to multiplex genome engineering. In this review, we describe the molecular mechanisms involved in the CRISPR system, and summarize and discuss the applications of this technology in plant genome engineering.

1 Introduction

With the completion of whole genome sequencing in an increasing number of crops, plant researchers face the challenge of converting the enormous amounts of sequence data into genetic information. The post-genome era requires the development of new affordable methods to elucidate gene function and develop new breeding technologies to enhance crop improvement. Targeted genome modification (TGM) technologies to generate sequence-specific genetic changes are long sought-after genome engineering tools (Curtin et al. 2012; Voytas 2013). In the past few years, engineered nucleases, including zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), have been widely used for TGM in many organisms (Carroll 2011; Joung and Sander 2012; Perez-Pinera et al. 2012; Porteus and Carroll 2005; Urnov et al. 2010; Sun and Zhao 2013). These engineered nucleases can be used to alter genomes at specific loci, and they make use of a targeting procedure in which proteins recognize specific DNA sequences. However these systems require a newly designed and synthesized nuclease for each

K. Chen, Ph.D. • C. Gao, Ph.D. (✉)
State Key Laboratory of Plant Cell and Chromosome Engineering,
Institute of Genetics and Developmental Biology, Chinese Academy of Sciences,
No. 1 West Beichen Road, Chaoyang, Beijing 100101, P. R. China
e-mail: cxgao@genetics.ac.cn

F. Zhang et al. (eds.), *Advances in New Technology for Targeted Modification of Plant Genomes*, DOI 10.1007/978-1-4939-2556-8_8

new target DNA sequence and reengineered ZFN or TALEN. Thus the assembly of ZFNs and TALENs is time consuming, costly, and requires experimental skill (Gaj et al. 2013).

Very recently, a revolutionary SSN technique based on the bacterial CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR associated) system has been developed for precise genome manipulation (Jinek et al. 2012; Pennisi 2013). The CRISPR system uses the Cas9 nuclease and an engineered single guide RNA (sgRNA) that specifically binds to the target sequence and recruits Cas9 to cleave the locus (Jinek et al. 2012; Cong et al. 2013; Mali et al. 2013a). Only a single customized gRNA is required to target a specific sequence, and Cas9 does not have to be reengineered for each new target site. The CRISPR system therefore promises to displace ZFNs and TALENs. Recently, the CRISPR technology was selected as one of the ten breakthroughs of the year by the journal Science (Genetic microsurgery for the masses 2013). Within just 1 year, applications of CRISPR-mediated genome modification were described in more than 50 publications using a wide range of hosts, such as human cells (Jinek et al. 2012, 2013; Cong et al. 2013; Mali et al. 2013a; Cho et al. 2013; Hou et al. 2013; Chen et al. 2013), rat (Hu et al. 2013; Li et al. 2013a), zebrafish (Auer et al. 2014; Chang et al. 2013; Hwang et al. 2013; Jao et al. 2013; Blackburn et al. 2013), fruit fly (Bassett et al. 2013; Gratz et al. 2013; Baena-Lopez et al. 2013), nematode (Friedland et al. 2013), mouse (Cong et al. 2013; Li et al. 2013a; Yang et al. 2013), yeast (Dicarlo et al. 2013), bacteria (Jiang et al. 2013a), plants (Mao et al. 2013; Miao et al. 2013; Nekrasov et al. 2013; Shan et al. 2013; Upadhyay et al. 2013; Xie and Yang 2013; Jiang et al. 2013b; Li et al. 2013b; Liang et al. 2014), and many other organisms. Mutation rates obtained with the CRISPR system are comparable to those observed with ZFNs and TALENs, or in many cases even higher in stably transformed plants (Feng et al. 2013). CRISPR has become one of the most useful tools for genome engineering since 2013. Like ZFNs and TALENs, the CRISPR system introduces double-strand breaks (DSBs) at appropriate genomic sites and triggers DNA repair by either error-prone nonhomologous end joining (NHEJ) or homologous recombination (HR) (Chen and Gao 2013; Wyman and Kanaar 2006) pathways. Therefore it can be used to generate modifications like gene knockouts, replacements, insertions, and chromosome rearrangements at targeted loci. Here, we review recent advances in CRISPR technology and the application of this technology to plant genomes. We also discuss its potential as a new plant breeding technology to overcome regulatory issues, as well as possible constraints on its expansion, and future prospects.

2 Development of CRISPR Technology

2.1 The CRISPR/Cas Defense System

The CRISPR/Cas system is a widespread bacterial and archaeal system of adaptive immunity against invading viruses and plasmids. Based on the core elements and sequences, three types of CRISPR/Cas systems have been identified (Barrangou 2013;

Makarova et al. 2011; Wiedenheft et al. 2012). The type I and type III systems require multiple proteins to form a large functional multi-Cas complex, whereas the type II system only needs a single Cas9 protein. Hence the latter is more easily adapted to other organisms for practical application.

The type II CRISPR locus from *Streptococcus pyogenes* SF370 contains a cluster of *cas* genes as well as two noncoding RNA elements—the trans activating CRISPR RNA (tracrRNA) and CRISPR array (Barrangou 2013; Deltcheva et al. 2011), in the order of tracrRNA-*cas* gene operon-CRISPR leader sequence-CRISPR arrays (Deltcheva et al. 2011; Barrangou 2012). The most characteristic feature of this system is the CRSIPR array of nearly identical DNA repeats, separated by regularly spaced sequences.

CRISPR/Cas is an RNA-guided DNA cleavage defense system. Once viruses and plasmids invade, host cells respond by integrating short fragments of the foreign DNA into the CRISPR loci as novel spacer sequences. Subsequently, tracrRNAs and a long pre-crRNA of repeat-spacer arrays are transcribed, and the tracrRNA hybridizes to each repeat sequence within the pre-crRNA to form a double-stranded RNA. After that, these dsRNAs are cleaved by endogenous RNase III and unknown nucleases associated with Cas9, and mature crRNAs are released (Deltcheva et al. 2011). The mature crRNAs consist of approximately 40 nt, generally made up of a 20 nt unique guide sequence originating from the spacer and a 19-22 nt 3′ handle derived from the repeat sequence. The crRNAs remain associated with the tracrRNA and Cas9 to form a ribonucleoprotein complex (Barrangou 2013). The targeted Cas9 cleaves the protospacer DNA in both strands—the HNH nuclease domain cleaves the complementary strand and the RuvC-like domain cleaves the noncomplementary strand. Precise cleavage of the nucleotide three-base pair upstream of protospacer adjacent motif (PAM) creates a blunt end. Once the crRNA pairs with the complementary protospacer sequence upstream of the PAM of the invader, a double-strand DNA R-loop is generated by the DNA strand displaced by the crRNA. Finally, the tracrRNA–crRNA-Cas9 ribonucleoprotein complex cleaves the protospacer DNA sequence (Barrangou 2013; Deltcheva et al. 2011).

2.2 CRISPR Genome Engineering

In 2012, the molecular machinery of the Type II CRISPR/Cas system derived from *Streptococcus pyogenes* was developed in the form of programmable RNA-guided endonucleases (RGENs) for genome editing by Jinek et al. (2012) (Fig. 8.1). These co-authors were the first to demonstrate that a sgRNA mimicking the dual-tracrRNA-crRNA molecule could direct sequence-specific cleavage as a ribonucleoprotein complex in vitro. The CRISPR/Cas system was thus dramatically simplified. Using sgRNA, Cas9 can introduce DSBs at any selected sites, and the only constraint is that the recognition sites must have a 3′-downstream NGG (PAM) sequence (Jinek et al. 2012). This major achievement provided the foundation for CRISPR genome engineering.

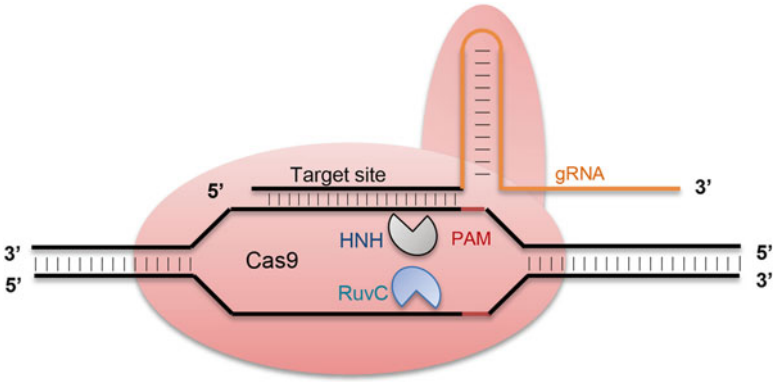

Fig. 8.1 The schematic of CRISPR structure. The CRISPR/Cas9 system consists of a sgRNA and the Cas9 nuclease. One nucleotide of the sgRNA interacts with one nucleotide of the DNA target site. The enzyme Cas9 is guided to the target DNA by sgRNA that contains a sequence that matches the sequence to be cleaved, which is demarcated by PAM sequences. Cas9 has two active sites that each cleave one strand of a double-stranded DNA molecule

By early 2014, precise DNA cleavage had been made at endogenous genomic loci by CRISPR in human cells (Cong et al. 2013; Mali et al. 2013a; Cho et al. 2013) and mice cells (Cong et al. 2013). In addition, targeted insertions were obtained by homologous recombination (HR) in *AAVS1* (Mali et al. 2013a) and *EMX1* (Cong et al. 2013). More importantly, another attractive feature of the CRISPR system was demonstrated the ability to edit several loci at the same time by introducing multiple sgRNAs simultaneously (Cong et al. 2013). This is difficult to achieve using ZFNs and TALENs. Later, in vivo genome editing by CRISPR was confirmed in zebrafish (Hwang et al. 2013), then extended to many other organisms (Pennisi 2013). In these reports, the cleavage efficiencies achieved with CRISPRs were equal to or better than those obtained with TALENs (Cong et al. 2013; Mali et al. 2013a; Cho et al. 2013; Hwang et al. 2013).

Along with the increased understanding of Cas9, researchers have expanded the application of CRISPR to include regulating genomic processes by, for example, nicking one strand of DNA to facilitating HR (Mali et al. 2013a), activate or repress gene expression (Bikard et al. 2013; Cheng et al. 2013; Maeder et al. 2013a; Perez-Pinera et al. 2013; Qi et al. 2013). From the results with ZFNs or TALENs, it was already known that breaking one strand could enhance DNA repair by the HR pathway, and Cong et al. tested site-directed mutagenesis with Cas9 following one strand cleavage of either the RuvC- or HNH-motif, and these strand-specific nickases were shown to efficiently increase the HR rate in mouse and human cells (Cong et al. 2013). Transcriptional regulation in eukaryotes involves a balance between activation and repression regulated by transcription factors and epigenetic modifications. A catalytically impaired Cas9 (dCas9) lacking endonuclease activity can be fused to different effectors. CRISPR associated with dCas9 can thereafter perform other tasks than cleaving DNA sequences. In *E. coli*, gene repression was achieved simply by expressing a dCas9 along with a sgRNA (Qi et al. 2013). SgRNA and

dCas9 can bind together to form a DNA recognition complex. Guided by sgRNA, the dCas9 can block transcription start sites or interfere with transcript elongation, thus repressing gene expression. This technology, called CRISPR interference (CRISPRi), can be used to repress multiple target genes by expressing dCas9 along with multiple sgRNAs (Qi et al. 2013). When dCas9 was fused to the KRAB domain of Kox1, the dCas9-KRAB fusion protein stimulated CRISPRi silencing, and knocked down CD71 and CXCR4 expression in human cells (Gilbert Luke et al. 2013). CRISPR activation (CRISPRa) technology involves fusing a dCas9 to activator domains, such as VP64 or p65, and was shown to be effective in activating the expression of single or multiple genes in human cells (Maeder et al. 2013a; Perez-Pinera et al. 2013; Gilbert Luke et al. 2013; Mali et al. 2013b). Activation and repression of transcription of RNA polymerase (RNAP) has also been demonstrated in bacteria using dCas9 (Bikard et al. 2013), and recently, targeted histone methylation or demethylation, and DNA demethylation, have been obtained using TALEs or ZFs fused to histone methyltransferases, histone deacetylases (Konermann et al. 2013; Mendenhall et al. 2013), or the TET1 catalytic domain (Maeder et al. 2013b). These results imply that epigenome editing by epigenetic CRISPR (epiCRISPR) will be realized in the near future using the same strategy.

2.3 Maximizing the Specificity of CRISPR

In theory, the specificity of CRISPR bestowed by the first 20 nt of the crRNA is enough to guarantee specificity, although only the seed region of the 13 bp adjacent to the PAM is required to specify the binding of Cas9 to the target site. Mismatch tolerance, particularly at the 5′-end of crRNA, has been frequently observed in mammalian cells (Jinek et al. 2012; Cong et al. 2013; Fu et al. 2013; Hsu et al. 2013; Pattanayak et al. 2013; Cradick et al. 2013; Cho et al. 2014) and plants (Shan et al. 2013; Upadhyay et al. 2013; Xie and Yang 2013). Current work suggests that the CRISPR system may not yet have adequate specificity to completely displace its ZFN and TALEN forebears, the main obstacle being the high off-target rate. In mammalian systems, some off-target effects were even more frequent than the on-target effects (Fu et al. 2013). Only limited off-target data are available for rice (Shan et al. 2013; Xie and Yang 2013), but fortunately, the off-target efficiency was lower than the perfect matches (Xie and Yang 2013).

Some approaches have been proposed to reduce the off-target rate: (1) The discovery in other species of archaea or bacteria of a Cas9 protein with a longer PAM. For example, NmCas9 from *Neisseria meningitidis* 8013 binds to a 24 nt protospacer sequence with a 5′-NNNAGAA-3′ PAM, and confers additional specificity (Hou et al. 2013); (2) Titration of the sgRNA and Cas9 levels to control the sgRNA/Cas9 reaction precisely (Hsu et al. 2013; Pattanayak et al. 2013); (3) Bioinformatic prediction to aid in rational target site selection and sgRNA design (Hsu et al. 2013; Pattanayak et al. 2013); (4) Artificial engineering paired mCas9 nickase function like ZFNs or TALENs by introducing two single strand breaks in both strands (Pattanayak et al. 2013; Cho et al. 2014).

3 CRISPR Technology in Plants

The first reports describing applications of the CRISPR system to plants were published in the August 2013 issue of Nature Biotechnology (Nekrasov et al. 2013; Shan et al. 2013; Li et al. 2013b). Three months later seven groups had employed this RNA-guided genome engineering system in more plant species (Mao et al. 2013; Miao et al. 2013; Upadhyay et al. 2013; Xie and Yang 2013; Jiang et al. 2013b; Liang et al. 2014; Feng et al. 2013). Up to now, model plants including *Arabidopsis* (Mao et al. 2013; Jiang et al. 2013b; Li et al. 2013b; Feng et al. 2013) and tobacco (Nekrasov et al. 2013; Jiang et al. 2013b; Li et al. 2013b) and crop plants including rice (Mao et al. 2013; Shan et al. 2013; Xie and Yang 2013; Jiang et al. 2013b; Feng et al. 2013), wheat (Shan et al. 2013; Upadhyay et al. 2013), maize (Liang et al. 2014), and sorghum (Jiang et al. 2013b) have been involved in this technology.

3.1 Improving CRISPR Technology in Plants

The application of CRISPR technology to plants was improved mainly by developing efficient CRISPR genome editing systems and fast validation systems. In order to obtain an efficient CRISPR system, the usual strategy is to optimize Cas9 and select an appropriate RNA polymerase III promoter for sgRNA expression. Belhaj et al. have summarized CRISPR improvements in details (Belhaj et al. 2013). On the other hand, various transient assay systems have been developed in a number species for CRISPR validation, because obtaining targeted genome-modified plants is very time consuming.

Cas9 has been optimized for the codon frequencies of different eukaryotes including rice (Miao et al. 2013; Shan et al. 2013), a variety of plants (Jiang et al. 2013b; Li et al. 2013b), *Chlamydomonas reinhardtii* (Jiang et al. 2013b), and humans (Mao et al. 2013; Upadhyay et al. 2013; Xie and Yang 2013; Jiang et al. 2013b), and even a wild spCas9 (Jiang et al. 2013b) has been validated in different plants (Table 8.1). The optimized Cas9 are provided with the appropriate nuclear localization signals (NLS) for correct subcellular localization. In plants, constitutive expression of Cas9 has been achieved under the control of the 35sCaMV (Nekrasov et al. 2013; Shan et al. 2013; Xie and Yang 2013; Jiang et al. 2013b), OsAct1 (Jiang et al. 2013b), 35SPPDK (Li et al. 2013b), or UBQ (Mao et al. 2013) promoters.

The sgRNA, another core element of CRISPR, is also generally modified to enhance its expression using the U3 or U6 RNA polymerase III promoter from native species such as *Arabidopsis*, rice, or wheat (Table 8.1). The design of the sgRNA depends on the target sequence. A specific 20 bp target sequence adjacent to a PAM sequence $(N)_{20}$NGG has been widely used in mammalian genome editing, but a variety of $(N)_{19\text{-}20}$NGG sequences are acceptable in plants (Miao et al. 2013; Shan et al. 2013; Feng et al. 2013). Research in plants has also shown that the sgRNA does not need to provide a strict match. For example, it is not required for the U6 promoter in $G(N)_{19\text{-}20}$ or for the U3 promoter in $A(N)_{19\text{-}20}$, which need the "G" or "A" to define the transcription start site (Shan et al. 2013).

Table 8.1 Targeted genome modification using CRISPR in plants

Plant species	Cas9 optimization	sgRNA promoter	Targeted genes name	Outcome	Modification type	Reference
Arabidopsis	Human or plant condon optimized Cas9	*AtU6*	*AtPDS3*, *AtFLS2*, *AtRACK1b*, *AtRACK1c*	Single or multiple gene mutagenesis	NHEJ	Li et al. (2013b)
	Human condon optimized Cas9	*AtU6*	*AtCHL1*, *AtCHL2*, *AtTT4*	Single or multiple gene mutagenesis, large deletion	NHEJ	Mao et al. (2013)
	Human condon optimized Cas9	*AtU6*	*AtBRI1*, *AtJAZ1*, *AtGAI*	Gene mutagenesis	NHEJ	Feng et al. (2013)
	Chlamydomonas reinhardtii condon optimized Cas9	*AtU6*	mGFP	Gene mutagenesis	NHEJ	Jiang et al. (2013b)
Tobacco	Cas9 variant	*AtU6*	*NbPDS*	Gene mutagenesis	NHEJ	Nekrasov et al. (2013)
	Plant condon optimized Cas9	*AtU6*	*NbPDS*	Gene mutagenesis, gene replacement	NHEJ, HR	Li et al. (2013a, b)
	Chlamydomonas reinhardtii condon optimized Cas9	*AtU6*	mGFP	Gene mutagenesis	NHEJ	Jiang et al. (2013a, b)
	Human condon optimized Cas9	*N.A.*	NbPDS	Gene mutagenesis	NHEJ	Upadhyay et al. (2013)

Table 8.1 (continued)

Plant species	Cas9 optimization	sgRNA promoter	Targeted genes name	Outcome	Modification type	Reference
Rice	Rice condon optimized Cas9	*OsU3*	*OsPDS*, *OsBADH2*, *Os02g23823*, *OsMPK2*,	Gene mutagenesis, gene replacement	NHEJ, HR	Shan et al. (2013)
	Rice condon optimized Cas9	*OsU3*	*OsCAO1*, *OsLAZY1*	Gene mutagenesis	NHEJ	Miao et al. (2013)
	Human condon optimized Cas9	*OsU6*	*OsROC5*, *OsSPP*, *OsYSA*	Gene mutagenesis	NHEJ	Feng et al. (2013)
	Chlamydomonas reinhardtii condon optimized Cas9	*AtU6*	*OsSWEET14*, *OsSWEET11*	Gene mutagenesis	NHEJ	Jiang et al. (2013a, b)
	Human condon optimized Cas9	*OsU6*	*OsMYB1*	Gene mutagenesis	NHEJ	Mao et al. (2013)
	Human condon optimized Cas9	*OsU3*, *OsU6*	*OsMPK5*	Gene mutagensis	NHEJ	Xie and Yang (2013)
Wheat	Rice condon optimized Cas9	*TaU6*	*TaMLO*	Gene mutagenesis	NHEJ	Shan et al. (2013)
	Human condon optimized Cas9	*N.A.*	*TaPDS TaINOX*	Single or multiple gene mutagenesis, large deletion	NHEJ	Upadhyay et al. (2013)
Maize	Rice condon optimized Cas9	*ZmU3*	*ZmIPK*	Gene mutagenesis	NHEJ	Liang et al. (2014)
Sorghum	Monocot condon optimized Cas9	*OsU6*	*mDeRED*	Gene mutagenesis	NHEJ	Jiang et al. (2013a, b)

N.A. not available

For fast validation of CRISPR, transient expression systems have been developed to assess in vivo efficiency of CRISPR in a short period of time. The protoplast-based transient system has been used to test NHEJ-mediated mutagenesis and HR-mediated gene replacement in *Arabidopsis*, tobacco, rice, and wheat. Moreover, targeted gene replacements in protoplasts have reached an efficiency of 18.8 %, or 42 % in *Arabidopsis* (Mao et al. 2013; Feng et al. 2013), 9 % in tobacco (Li et al. 2013b), and 6.9 % in rice (Shan et al. 2013). Other transient systems that have been examined include leaf agroinfiltration for *Arabidopsis* (Li et al. 2013b) and tobacco (Nekrasov et al. 2013; Jiang et al. 2013b), embryo transformation for sorghum (Jiang et al. 2013b), and suspension cells for wheat (Upadhyay et al. 2013). Mutation detection methods used widely include PCR/RE, which can detect disruption of a preserved restriction enzyme site in the targeted sequence, and Surveyor assays based on T7 endonuclease or Cel1.

In planta, the mutation frequency can reach 89 % in *Arabidopsis* (Mao et al. 2013), and 91.6 % in rice (Miao et al. 2013). In addition to point mutations, large deletions in *AtTT4* resulting in removal of a 230-bp fragment have been obtained (Mao et al. 2013), and multiplex gene disruptions have been created in *Arabidopsis* (Mao et al. 2013; Li et al. 2013b), rice (Miao et al. 2013), and wheat (Upadhyay et al. 2013).

3.2 Examples of Applications to Crop Plants

To help the reader understand the applications to crops, we summarize achievements in three major food crops as follows.

3.2.1 Rice

Up to now, 13 endogenous rice genes have been disrupted by five different groups using CRISPR (Table 8.1). Our group established a CRISPR genome editing system in rice and wheat using rice codon-optimized SpCas9, with an NLS attached at both ends and a proOsU3-sgRNA (Shan et al. 2013). This system has been used to mutate four rice genes (*OsPDS*, *OsBADH2*, *Os02g23823*, and *OsMPK2*) in rice protoplasts with efficiencies from 26.5.5 to 38.0 %. Stably regenerated mutants of *OsPDS* and *OsBADH2* were also obtained by bombarding Cas9 along with the corresponding sgRNAs into rice calli. In addition, three biallelic mutations of *OsPDS* had the typical albino and dwarf *OsPDS* null phenotype (Fig. 8.2). Furthermore, a single-stranded oligo with a KpnI+EcoRI site was introduced into *OsPDS*; this proved for the first time that HR-dependent targeted gene insertions can be engineered by CRISPR in rice. Through a bioinformatic analysis, we found that there are ~3,183,497 and ~566,367 sequences specifically targetable by sgRNAs in the rice genome and rice cDNAs, respectively. This finding implies that nearly all transcripts can be targeted by CRISPR, with an average of nine targets per cDNA.

Fig. 8.2 The dwarf and albino phenotype of a rice OsPDS knockout produced by the CRISPR/Cas9 system

The Zhu group has obtained knockout mutants of the rice genes *OsROC5*, *OsSPP*, and *OsYSA* by transforming an Agrobacterium vector containing HspCas9 and proOsU6-sgRNA cassettes into rice calli (Feng et al. 2013). In the T0 generation, 10 % of the *ysa* mutants had the expected albino leaf phenotype. One line with an *OsROC5* mutation and two lines with *OsYSA* mutations harbored homozygous or biallelic mutations (Feng et al. 2013). This group reported a mutation frequency of about 50 % in *OsMYB1* in another publication (Mao et al. 2013).

The Qu group has developed Gateway™ binary T-DNA vectors for co-expression of spCas9 and sgRNA (Miao et al. 2013) and tested them with sgRNAs or dual-crRNA:tracrRNAs targeting either *OsCAO1* or *OsLAZY1*. The mutagenesis efficiency generated by the sgRNA constructs was higher than that of dual-crRNA:tracrRNA constructs. The mutation frequencies of *OsCAO1* and *OsLAZY1* in the T1 generation reached 83.3 % and 91.6 %, respectively, using sgRNAs. Also, knockout mutants with biallelic mutations of both genes displayed the expected phenotypes.

The Weeks group has focused on testing CRISPR technology with reporter system in monocot and dicot plants. They targeted the promoter region of the bacterial blight susceptibility genes, *OsSWEET14* and *OsSWEET11*, in rice protoplast. Meanwhile, Cas9/sgRNA activity on an out-of-frame DsRED2 was also identified in immature sorghum embryos (Jiang et al. 2013b).

Yang's group has created targeted knockouts in the rice disease resistance gene *OsMPK* using CRISPRs (Xie and Yang 2013). Their whole genome analysis predicted that specific gRNAs could be designed to target more than 90 % of rice genes.

3.2.2 Wheat

As a hexaploid plant with a giant genome of 17 Gb genome, wheat is the most complicated food crop. Large genome size, polyploidy, and lack of an efficient genetic transformation system are substantial barriers to generate TGMs. Targeted genome editing was not even available in wheat until CRISPR technology appeared. The wheat TaMLO gene, an important gene implicated in powdery mildew resistance, was disrupted in wheat protoplasts with a proTaU6-sgRNA and rice codon-optimized Cas9 (Shan et al. 2013). Upadhyay et al. reported that two wheat endogenous genes *TaINOX* and *TaPDS* were disrupted with a mutation efficiency of 18–22 % in suspension cells, and multiplex CRISPR RNA targeting two different positions and two separate genes was also studied (Upadhyay et al. 2013).

3.2.3 Maize

Our group has recently confirmed CRSPR genome editing in maize with the same rice-codon optimized Cas9 and a proZmU3-sgRNA. Two sgRNAs targeting the *ZmIPK* gene in maize protoplasts achieved frequencies of 16.4 % and 19.1 %, respectively. The CRISPR/Cas system induced targeted mutations in *Zea mays* protoplasts with an efficiency (13.1 %) similar to that obtained with TALENs (9.1 %) (Liang et al. 2014).

4 Perspective and Conclusions

CRISPR-mediated genome modification is rapidly becoming a powerful tool for genome engineering. This new technology can be developed as a fast, efficient genome editing and genome regulation approach for functional genomics research as well as genetic modification of crops. Some developing directions need to be considered.

4.1 Developing CRISPR as a New Plant Breeding Technique for Next Generation Crop Improvement

To support the demands of an increasing human population, new breeding techniques (NBTs) that can rapidly, efficiently, and precisely produce innovative varieties with improved traits are greatly needed (Lusser et al. 2012). TGM technology based on CRISPR is a sophisticated method to modify crop genomes, and is one of the NBTs holding the greatest promise for crop improvement (Lusser et al. 2012; Chen and Gao 2014; Podevin et al. 2012, 2013; Waltz 2012; Pauwels et al. 2014).

In contrast to classical transgenic breeding, this precise manipulation of genomes may overcome some of the constraints associated with transgene-based breeding techniques. For example, modified plants with only frame-shift mutations due to small insertions or deletions induced by NHEJ, or with sequence-specific point mutations resulting from homologous recombination (HR), are indistinguishable from those introduced by traditional breeding methods (Lusser et al. 2012; Chen and Gao 2014). Therefore, these products are similar to natural variants, or those produced by physical or chemical mutagenesis in conventional breeding (Podevin et al. 2013; Kuzma and Kokotovich 2011). The method can be used to create allelic variation and the products may fall outside existing definitions and regulations affecting GM crops (Lusser et al. 2012; Podevin et al. 2012).

Furthermore, this technology may avoid introducing novel genetic elements and proteins; for example, TGM plants regenerated from protoplasts or suspension cells can be high precision modified crop varieties with marker gene-independent and antibiotic-free genomes. Therefore this technology has the advantage over transgenic procedures of reducing potential risks and thus of enjoying greater public acceptance (Lusser et al. 2012; Podevin et al. 2012).

Of course, CRISPRs have other advantages as powerful NBTs. Firstly, CRSIPR is easy to implement and only an inexpensive commercial synthesis of sgRNAs is needed. Secondly, there are many potential target sites; these only need an NGG PAM, and there is an average of one PAM every 10 bp in the rice genome (Xie and Yang 2013). Thirdly, target cleavage is not affected by DNA methylation, which means that cleavage should be effective at many more loci (Hsu et al. 2013). Finally, the ability to perform multiplex genome editing by CRISPR creates the possibility of altering several key factors related to an important trait, or modifying several traits at the same time. This could help to understand a whole metabolic pathway, or signal transduction or transcript regulation network, and so should be a powerful aid in compound breeding.

Thus CRISPR-mediated TGM is an exciting direction for both research and application in crop improvement.

4.2 Developing CRISPR for Regulating Plant Genomes

Cas9 can be recruited to form a transcrRNA-crRNA-Cas9 complex to target any chosen site and thus provides opportunities for altering gene regulation by fusing dCas9 to other functional domains. CRISPRa which up-regulates the transcription of endogenous genes is a powerful tool for obtaining accurate gene overexpression in the right location. Similarly, CRISPRi can be used to create mutants with down-regulated expression of selected genes, and especially multiple homologous gene knockdown mutants. Neither CRISPRa nor CRISPRi has yet been tested in plants, but it is important to do this. Furthermore, demethylation or methylation of histones or DNA has so far only been reported using ZFNs and TALENs, but we believe that epiCRISPR will very soon have an impact on epigenetic research in plants.

In conclusion, CRISPR is a promising genome editing method, though it is still far from mature in plants. Many improvements are still needed for effective CRISPR application. In any event, we believe that this rapid, versatile technique will not only ensure a bright future for plant genomic research, but may also in the near future become a routine NBT once its specificity has been improved.

References

(2013) Genetic microsurgery for the masses. Science 342:1434–1435.
Auer TO, Duroure K, De Cian A, Concordet JP, Del Bene F (2014) Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. Genome Res 24: 142–153
Baena-Lopez LA, Alexandre C, Mitchell A, Pasakarnis L, Vincent JP (2013) Accelerated homologous recombination and subsequent genome modification in Drosophila. Development 140:4818–4825
Barrangou R (2012) RNA-mediated programmable DNA cleavage. Nat Biotechnol 30:836–838
Barrangou R (2013) CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip Rev RNA 4:267–278
Bassett AR, Tibbit C, Ponting CP, Liu JL (2013) Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep 11:220–228
Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods 9:39
Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41:7429–7437
Blackburn PR, Campbell JM, Clark KJ, Ekker SC (2013) The CRISPR system–keeping zebrafish gene targeting fresh. Zebrafish 10:116–118
Carroll D (2011) Genome engineering with zinc-finger nucleases. Genetics 188:773–782
Chang N, Sun C, Gao L et al (2013) Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res 23:465–472
Chen K, Gao C (2013) TALENs: customizable molecular DNA scissors for genome engineering of plants. J Genet Genomics 40:271–279
Chen K, Gao C (2014) Targeted genome modification technologies and their applications in crop improvements. Plant Cell Rep 33(4):575–583. doi:10.1007/s00299-013-1539-6, Epub 2013 Nov 24
Chen B, Gilbert LA, Cimini BA et al (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas System. Cell 155(7):1479–1491
Cheng AW, Wang H, Yang H et al (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res 23:1163–1171
Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31:230–232
Cho SW, Kim S, Kim Y et al (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res 24:132–141
Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823
Cradick TJ, Fine EJ, Antico CJ, Bao G (2013) CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res 41:9584–9592
Curtin SJ, Voytas DF, Stupar RM (2012) Genome engineering of crops with designer nucleases. Plant Genome 5:42–50

Deltcheva E, Chylinski K, Sharma CM et al (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471:602–607
Dicarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343
Feng Z, Zhang B, Ding W et al (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell Res 23:1229–1232
Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA (2013) Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat Methods 10:741–743
Fu Y, Foden JA, Khayter C et al (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31:822–826
Gaj T, Gersbach CA, Barbas Iii CF (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31:397–405
Gilbert Luke A, Larson Matthew H, Morsut L et al (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154:442–451
Gratz SJ, Cummings AM, Nguyen JN et al (2013) Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194:1029–1035
Hou Z, Zhang Y, Propson NE et al (2013) Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc Natl Acad Sci U S A 110:15644–15649
Hsu PD, Scott DA, Weinstein JA et al (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31:827–832
Hu X, Chang N, Wang X et al (2013) Heritable gene-targeting with gRNA/Cas9 in rats. Cell Res 23:1322–1325
Hwang WY, Fu Y, Reyon D et al (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31:227–229
Jao LE, Wente SR, Chen W (2013) Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc Natl Acad Sci U S A 110:13904–13909
Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013a) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31:233–239
Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013b) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41:e188
Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821
Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J (2013) RNA-programmed genome editing in human cells. eLife 2:e00471
Joung JK, Sander JD (2012) TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14(1):49–55
Konermann S, Brigham MD, Trevino A et al (2013) Optical control of mammalian endogenous transcription and epigenetic states. Nature 500:472–476
Kuzma J, Kokotovich A (2011) Renegotiating GM crop regulation. EMBO Rep 12:883–888
Li D, Qiu Z, Shao Y et al (2013a) Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nat Biotechnol 31:681–683
Li JF, Norville JE, Aach J et al (2013b) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol 31:688–691
Liang Z, Zhang K, Chen K, Gao C (2014) Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. J Genet Genomics 41:63–68. doi:10.1016/j.jgg.2013.12.001
Lusser M, Parisi C, Plan D, Rodríguez-Cerezo E (2012) Deployment of new biotechnologies in plant breeding. Nat Biotechnol 30:231–239
Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK (2013a) CRISPR RNA-guided activation of endogenous human genes. Nat Methods 10:977–979
Maeder ML, Angstman JF, Richardson ME et al (2013b) Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat Biotechnol 31:1137–1142

Makarova KS, Haft DH, Barrangou R et al (2011) Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9:467–477
Mali P, Yang L, Esvelt KM et al (2013a) RNA-guided human genome engineering via Cas9. Science 339:823–826
Mali P, Aach J, Stranges PB et al (2013b) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 31: 833–838
Mao Y, Zhang H, Xu N, Zhang B, Gao F, Zhu JK (2013) Application of the CRISPR-Cas system for efficient genome engineering in plants. Mol Plant 6:2008–2011
Mendenhall EM, Williamson KE, Reyon D et al (2013) Locus-specific editing of histone modifications at endogenous enhancers. Nat Biotechnol 31:1133–1136
Miao J, Guo D, Zhang J et al (2013) Targeted mutagenesis in rice using CRISPR-Cas system. Cell Res 23:1233–1236
Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol 31:691–693
Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol 31:839–843
Pauwels K, Podevin N, Breyer D, Carroll D, Herman P (2014) Engineering nucleases for gene targeting: safety and regulatory considerations. N Biotechnol 31:18–27
Pennisi E (2013) The CRISPR craze. Science 341:833–836
Perez-Pinera P, Ousterout DG, Gersbach CA (2012) Advances in targeted genome editing. Curr Opin Chem Biol 16:268–277
Perez-Pinera P, Kocak DD, Vockley CM et al (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods 10:973–976
Podevin N, Devos Y, Davies HV, Nielsen KM (2012) Transgenic or not? No simple answer! EMBO Rep 13:1057–1061
Podevin N, Davies HV, Hartung F, Nogué F, Casacuberta JM (2013) Site-directed nucleases: a paradigm shift in predictable, knowledge-based plant breeding. Trends Biotechnol 31:375–383
Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. Nat Biotechnol 23:967–973
Qi LS, Larson MH, Gilbert LA et al (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183
Shan Q, Wang Y, Li J et al (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31:686–688
Sun N, Zhao H (2013) Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. Biotechnol Bioeng 110:1811–1821
Upadhyay SK, Kumar J, Alok A, Tuli R (2013) RNA guided genome editing for target gene mutations in wheat. G3 3:2233–2238
Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646
Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. Annu Rev Plant Biol 64:327–350
Waltz E (2012) Tiptoeing around transgenics. Nat Biotechnol 30:215–217
Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. Nature 482:331–338
Wyman C, Kanaar R (2006) DNA double-strand break repair: all's well that ends well. Annu Rev Genet 40:363–383
Xie K, Yang Y (2013) RNA-guided genome editing in plants using a CRISPR-Cas system. Mol Plant 6:1975–1983
Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R (2013) One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 154:1370–1379

Index

F. Zhang et al. (eds.), *Advances in New Technology for Targeted Modification of Plant Genomes*, DOI 10.1007/978-1-4939-2556-8