PLANT GENOME ENGINEERING WITH SEQUENCE-SPECIFIC NUCLEASES: METHODS FOR EDITING DNA IN WHOLE PLANTS

A DISSERTATION SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA

Nicholas J. Baltes

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Advisor: Daniel F. Voytas

August 2014

Acknowledgements

I would like to express my sincere gratitude to everyone in the Voytas lab, both past and present. And I would like to specifically thank my advisor, Dan Voytas, for his support and encouragement.

Dedication

For my wife, Ashley, and my daughter, Adalyn.

ABSTRACT

The development and function of all living organisms, from bacteria to humans, is encoded within a universal blueprint–deoxyribonucleic acid (DNA). The ability to re-write this code of life promises great benefits, ranging from a better understanding of gene function to correcting genetic diseases. Therefore, there is high value for tools and techniques that enable genome editing in living cells.

Two revolutionary discoveries have facilitated the development of current tools and methodologies for genome engineering. The first came from studies in yeast and mice demonstrating that synthetic donor DNA–fragments of DNA containing homologous sequences to a chromosomal target–can recombine with the target through homologous recombination, thereby incorporating the information carried by the synthetic donor into the genome. Whereas these results established a method for editing genomic DNA, the absolute frequency of recombination was quite low, ranging from one correctly-modified cell in 10^4 to 10^7. The second discovery came from studies demonstrating that damaging DNA greatly increases recombination frequencies between like sequences. Together, these discoveries have led to a powerful method for efficiently editing DNA: delivering a donor DNA while simultaneously breaking the target locus.

In the last 20 years, multiple classes of enzymes have been developed that can be 'rewired' to recognize and break a DNA sequence of interest. These enzymes (herein referred to as sequence-specific nucleases) have proven to be powerful reagents for editing DNA in higher-eukaryotic cells. However, the ability to modify plant DNA does not solely depend on the activity of the genome engineering reagents. Instead, it also depends on the efficiency with which the genome engineering reagents are delivered, the cells they are delivered to, and the effectiveness of selecting (or screening) for cells with the desired modification.

Studies within this dissertation seek to develop novel methods for delivering genome engineering reagents to whole plants. First, in chapter 2, we focused our attention on geminiviruses—a large family of plant DNA viruses. Prior to these studies, geminiviruses were primarily used as vectors for virus-induced gene silencing or for protein expression; however, their circular DNA genomes, and their ability to replicate extrachromosomally, makes them an attractive vector for delivering genome engineering reagents (sequence-specific nucleases and donor molecules). Here, we were the first demonstrate their utility for editing plant DNA. Proof-of-concept experiments in *Nicotiana tabacum* established that replicons based on the bean yellow dwarf virus can indeed deliver genome engineering reagents to leaf cells, and that these modified cells could grow into calli and seedlings. Interestingly, we also observed an enhancement in homologous recombination in leaf cells, relative to our non-viral controls. This enhancement was due to a replicating donor molecule and by pleiotropic activity of the virus replication proteins.

In addition to DNA viruses, we explored the use of RNA viruses for the delivery of sequence-specific nucleases. Tobacco rattle virus—a plant RNA virus—is an attractive vector system for delivering foreign protein because it can systemically infect plants, it causes very mild symptoms, and it is strictly RNA which circumvents government regulation regarding the delivery of foreign DNA to plants. Previous studies demonstrated its utility by delivering sequence-specific nucleases to tobacco and petunia. Here, we show its use in the major model organism *Arabidopsis thaliana* (chapter 3). We first characterize virus movement throughout *Arabidopsis* plants showing its trafficking to rosettes, cauline leaves and floral tissue. We then demonstrate transient delivery of zinc-finger nucleases to leaf tissue by detecting mutations in cells where the virus has moved. Finally, we assessed the ability for tobacco rattle virus to facilitate mutagenesis of germline cells by screening next-generation seedlings for zinc finger nuclease-induced modifications. From ~10,000 seeds (from 16 different infected plants), we observed 5 seedlings with the modification event, suggesting that TRV can enter and facilitate mutagenesis of seed-progenitor cells.

The remaining of the dissertation focuses on expanding the utility of stable integration into plant genomes by applying this approach to additional plants, additional target genes, and additional genome modifications. We demonstrate targeted knockout of *Arabidopsis* genes that have not been previously knocked out by traditional mutagenesis methods, including T-DNA insertion or ethyl methyl sulfonate (appendix A). At the time, this was the first study to use this method to knockout a biologically-interesting and endogenous gene. This method was then applied in *Glycine max* to knockout several genes involved in RNA interference (chapter 5), and it was used in *Arabidopsis* to generate large chromosomal deletions, inversions and duplications (chapter 4). Taken together, studies conducted in this dissertation improve upon the methods and technologies for delivering reagents to whole plants.

TABLE OF CONTENTS

LIST OF TABLES **PAGE**

LIST OF FIGURES

CHAPTER 1

LIST OF PUBLICATIONS

1. **Baltes NJ,** Gil-Humanes J, Cermak T, Atkins PA, Voytas DF (2014). DNA replicons for plant genome engineering. *The Plant Cell* **26:** 151-63

2. Qi Y, Li X, Zhang Y, Starker CS, **Baltes NJ**, Zhang F, Sander JD, Reyon D, Joung KJ, Voytas DF (2013). Targeted deletion and inversion of tandemly arrayed genes in Arabidopsis thaliana using zinc finger nucleases. *G3* doi:10.1534/g3.113.006270. PMID: 23979943

3. Curtin S, Zhang F, Sander J, Haun W, Starker C, **Baltes NJ**, Reyon D, Dahlborg E, Goodwin M, Coffman A, Dobbs D, Joung K, Voytas DF, Stupar RM (2011). Targeted mutagenesis of duplicated genes in soybean with zinc finger nucleases. *Plant Physiology* **156:** 466-473 PMID: 21464476

CHAPTER 1

GENERAL INTRODUCTION

1.1 Significance

Severe societal challenges will be faced in the upcoming decades, including food shortages due to an increasing population. By the year 2050, world population is expected increase from the current 7 billion people to over 9 billion people. Counteracting an escalating food demand will require a 50% increase in agricultural production by 2030 (Ronald, 2011). Unfortunately, the amount of remaining arable land is limited, necessitating an increase in food production on currently-used land. Compounding these challenges are the predicted crop losses due to extreme temperatures, pest attacks, and pathogen outbreaks.

A powerful approach that may help overcome these challenges is to modify DNA sequences within plant chromosomes. For example, herbicide tolerance can be introduced by altering a few DNA bases within native plant genes (e.g., modification of the acetolactate synthase genes results in resistance to imidazolinone and sulphonylurea). Furthermore, plants can be engineered to have increased tolerance to environmental stresses (e.g., drought) and increased resistance to pathogens (e.g., viruses, fungi, bacteria, insects and nematodes). These, and additional crop improvements, suggest plant genome engineering will play a vital role in meeting the agricultural demands caused by an expanding population.

In addition to improving the genetics of crops, genome engineering can also be used to produce valuable plants or products for non-agricultural purposes. For example, there is great potential for plants to be used as bioreactors for pharmaceutical proteins. This process (referred to as molecular pharming) has many advantages over current production strategies, including the low cost of production, facile scalability, and reduced risk of contaminating human pathogens. However, to realize the potential benefits from these applications, it is critical that we generate effective tools and approaches for editing plant DNA.

1.2 Harnessing Cellular DNA Double Strand Break Repair Pathways for Introducing Desired Sequence Changes

One method to efficiently modify plant genomes involves introducing targeted DNA double-strand breaks at a locus of interest. Double-strand breaks are highly toxic lesions: a single break can arrest the cell cycle and, if left unrepaired, can lead to cell death. To preserve the integrity of their genomes, all living organisms have evolved pathways to repair genetic lesions. In general, cells have are two main repair pathways: non-homologous end joining and homologous recombination. Repair mechanisms of these two pathways can be exploited to introduce sequence changes within plant genomes.

1.2.1 Non-Homologous End Joining and Targeted Mutagenesis

Non-homologous end joining is the major pathway plant cells use to repair double-strand breaks and it is active in all phases of the cell cycle (Rothkamm et al., 2003). Currently, there are

believed to be two major forms of non-homologous end joining: canonical and alternative. Repair mechanics of the canonical pathway involve the direct rejoining of the two exposed DNA ends. Shortly after a double-strand break is generated, a complex of Ku70 and Ku80 proteins are recruited to the break site to protect and stabilize the exposed DNA ends. The Ku70/80 heterodimer then recruits DNA-PKcs (DNA-dependent protein kinase catalytic subunit). After possible processing by the MRN complex (MRE11-RAD50-NBS1) complex , DNA gaps are filled by DNA polymerases μ and λ, and the free DNA ends are ligated together by a complex of LIG4 (DNA ligase IV) and XRCC4 proteins (West et al., 2000) (Figure 1.1). Because non-homologous end joining is non-template directed, repair can result in insertions and deletions of nucleotides at the break site. On the other hand, alternative non-homologous end joining is a Ku-independent pathway, and frequently uses microhomology for repair. Here, exposed DNA ends are bound by PARP1 which then recruits MRN, CTIP and BRCA1 for subsequent end processing. The processed DNA ends are then ligated together by either LIG3 and XRCC1 or LIG1.

The error-prone nature of the non-homologous end joining pathway can be exploited to introduce sequence changes within genomes. If double-strand breaks are directed to a DNA sequence of interest within the host's genome, targeted mutagenesis can be facilitated (Lloyd et al., 2005). Furthermore, if breaks are directed to coding sequences within genes, targeted gene knockout can be facilitated (Zhang et al., 2010). For example, gene knockout can occur from frameshift mutations or deletions that remove nucleotides coding for essential amino acids. Lastly, if two or more breaks are directed to sequences within a single chromosome, one can facilitate multiplex targeted mutagenesis, targeted deletion or targeted inversion of intervening sequence (Qi et al., 2013a). And, on the other hand, if two or more breaks are made on different chromosomes, one can facilitate translocations (Piganeau et al., 2013). Taken as a whole, exploiting the repair mechanics of non-homologous end joining provides genome engineers with an approach to knockout genes, or delete or rearrange sequences within living cells (Figure 1.2).

1.2.2 Homologous Recombination and Gene Targeting

Homologous recombination is a central cellular process in prokaryotes and higher-eukaryotic cells. It is involved in several aspects of genome stability, including DNA double-strand break repair, segregation of chromosomes, resolution of stalled replication forks and telomere maintenance. Repair of DNA double-strand breaks by homologous recombination is template-directed and, therefore, is a quasi-error-free process. Repair begins with resection of the 5' DNA ends by the MRN complex together with additional proteins, including CtIP, EXO1 (DNA exonuclease I) and BLM (Bloom syndrome, RecQ helicase like), thereby exposing long stretches of single-stranded 3' DNA ends (Figure 1.1). These ends are coated and stabilized by RPA (replication protein A), followed by the replacement of RPA with RAD51 and BRCA2 (breast cancer type 2 susceptibility protein). The 3' DNA ends then search out and invade a homologous

sequence of DNA, resulting in D-loop formation. After invasion, there are two models that describe the subsequent extension and resolution process: the double-strand break repair model and the synthesis-dependent strand annealing model. According to the double-strand break repair model, both free 3' DNA ends invade a homologous sequence of DNA. The homologous DNA then serves as a template for DNA polymerases to extend both 3' ends (Szostak et al., 1983), thereby copying missing genetic information from the template molecule. Following extension, the invading strands can then be resolved yielding either crossover or non-crossover products (Figure 1-1). Early studies exploring the outcome of homologous recombination in plant cells demonstrated that, in some instances, one side of the DNA break was repaired by homologous recombination and the other by non-homologous end joining (Puchta, 1999; Puchta et al., 1996). Therefore, the double-strand break repair model was not appropriate for describing homologous recombination in plants. Currently, it is believed that the primary mechanism for homologous recombination in plants follows the synthesis-dependent strand annealing model (Puchta, 1999). Here, one single-stranded 3' DNA end invades a homologous sequence and is extended by host DNA polymerases. Following extension, the 3' DNA end is released and anneals to sequence present on the other (non-invading) strand. Extra DNA sequence, which creates flaps of DNA, is trimmed, and repair is completed when the single-stranded ends are ligated.

Frequencies of homologous recombination in plant cells are low, hindering the development of methods for gene targeting (i.e., recombination between a user supplied donor DNA molecule and a chromosomal locus). Whereas non-homologous end joining is considered a major repair pathway in most all plant cells, homologous recombination is, in general, limited to a small subset of plant cells, including those undergoing meiosis. Furthermore, repair of double-strand breaks by homologous recombination is largely determined by the phase of the cell cycle, predominating in the S and G2 phases when sister chromatids are present (Takata et al., 1998). Initial methods to implement gene targeting have used *Agrobacterium* to deliver T-DNA containing a donor molecule to plant cells (Offringa et al., 1990). Here, the target for modification was a defective neomycin phosphotransferase (*nptII*) gene that was stably integrated within the genome of *Nicotiana tabacum* plants. Repair of this defective gene through homologous recombination results in the production of NPTII protein, and consequently, cells become resistant to kanamycin. These cells can then divide and form calli (unorganized parenchyma cells) and eventually whole plants can be regenerated. Using this approach, the authors determined the gene targeting frequency (the number of true gene targeting events divided by the total number of times the T-DNA randomly integrated into the genome) was 3×10^{-4}. Additional studies in plants using either *Agrobacterium*-mediated transformation or direct gene-transfer to cultured cells obtained similar gene targeting frequencies, ranging from 10^{-4} to 10^{-6} (Risseeuw et al., 1995; Hanin et al., 2001; Hrouda and Paszkowski, 1994; Lee et al., 1990; Paszkowski et al.,

1988; Offringa et al., 1993). These results demonstrate that, whereas *Agrobacterium* T-DNA and plasmid DNA can be used as a substrate for homologous recombination, the frequency of gene targeting is low. Furthermore, these results stressed that, if gene targeting is to become routine practice in plants, significant enhancements in recombination frequencies are required.

1.3 Introducing targeted DNA Double Strand Breaks with Sequence-Specific Nucleases

One approach to increase the frequency of homologous recombination in plant chromosomes is to introduce a double-strand break at the target locus (Puchta et al., 1996). Here, and similar to the method described in the previous section (Offringa et al., 1990), a T-DNA harboring a donor molecule was delivered to plant cells for the repair of a defective reporter gene; however, in addition to delivering the donor molecule, plant cells were also delivered an enzyme that creates DNA breaks at the target locus (Figure 1-2). The authors demonstrated a significant increase in gene targeting frequencies (from ~10^{-6} to 10^{-3}). As a result of this and other studies, there has been a demand for reagents that can introduce targeted DNA breaks. Currently, there are four classes of sequence-specific nucleases: meganucleases, zinc-finger nucleases, transcription activator-like effector nucleases, and the clustered, regularly interspaced, short palindromic repeat (CRISPR)/CRISPR associated (Cas) system. All classes of nucleases can be engineered to recognize and cleave specific DNA sequences. Here we review the four classes of sequence-specific nucleases and describe their use in plant genome engineering.

1.3.1 Meganucleases

Meganucleases (also known as homing endonucleases) were the first class of sequence-specific nucleases used for plant genome engineering (Figure 1-3). Meganucleases were initially found to be encoded within introns of yeast genes, and their protein activity was found to mobilize their sequence to non-intron containing alleles (Jacquier and Dujon, 1985). This process is initiated by a meganuclease-induced double-strand break at the exon-exon junction within an intronless allele. The exposed DNA ends are then repaired by homologous recombination using information from the intron-containing allele, thereby creating an extra copy of the meganuclease-encoded intron. Some meganucleases also function as RNA maturases and facilitate the splicing of their intron (Delahodde et al., 1989).

Meganucleases are divided into five families based on sequence and structural motifs: LAGLIDADG (the most commonly used class for genome engineering), HNH, GIY-YIG, His-Cys box and PD-(D/E)XK. Unlike restriction enzymes found in bacteria, meganucleases have long recognition sequences (~14 to 40 nt for the LAGLIDADG family). This feature has enabled their use as genome engineering tools because they can bind and cleave a single site within the genome of higher-eukaryotic cells.

The amino acid sequence of meganucleases can be modified to redirect binding to a desired DNA sequence. However, relative to other sequence-specific nucleases, meganucleases are the most challenging class to redesign. This process is hindered by their non-modular protein structure. For example, within the LAGLIDADG family of meganucleases, the amino acids responsible for DNA binding share the same domain as those responsible for DNA cleavage (Prieto et al., 2007); therefore, attempting to alter the DNA binding sequence may affect their ability to create double-strand breaks. Furthermore, two to four consecutive nucleotides are recognized by a cluster of six to nine amino acids (Taylor et al., 2012), and a complex network of amino acids (up to 50) that help make the structure of the protein are also responsible for the indirect and direct DNA contacts. Methods to redirect targeting include in vitro evolution of coding sequences. Here, very large number of variants (as large as 5×10^{11}) are assayed for activity against desired target sequences (Takeuchi et al., 2014). Using this method, several highly-active meganucleases were engineered to recognize the *CFTR* gene within human cells.

Whereas meganucleases can be redesigned to target a sequence of interest, their use in plant genome engineering has been limited. Most applications have used the naturally-occurring I-*Sce*I or I-*Ceu*I and the meganuclease target sequence was inserted on a synthetic target site that was pre-integrated into the plant genome (Salomon and Puchta, 1998; Chilton and Que, 2003; Puchta et al., 1996; Ayar et al., 2012; Fauser et al., 2012). However, with advancements in methods for redirecting targeting (Takeuchi et al., 2014), and their relatively small size (~165 amino acids), meganucleases have potential to be powerful reagents for plant genome engineering.

1.3.2 Zinc-Finger Nucleases

Zinc-finger nucleases are chimeric fusion proteins that consist of a DNA-binding domain and DNA-cleavage domain (Figure 1-3). The DNA-binding domain is composed of a set of Cys_2His_2 zinc fingers (usually 3 to 6). Each zinc finger contacts primarily 3 bp of DNA and a set of 3 to 6 fingers recognizes 9 to 18 bp, respectively. The DNA-cleavage domain is derived from the cleavage domain of the *Fok*I restriction enzyme. *Fok*I activity requires dimerization; therefore, to site-specifically cleave DNA, two zinc-finger nucleases are designed in a tail-to-tail orientation (Kim et al., 1996).

Zinc-finger nucleases can be ‘rewired’ to recognize different DNA sequences. However, one challenge with redirecting targeting is the context-dependency of zinc fingers. For example, a zinc finger that recognizes GGG may not recognize this sequence when fused to other zinc fingers. As a result, modular assembly of zinc fingers has had limited success (Ramirez et al., 2008). One of the more successful methods for redirecting targeting involves generating a library of three zinc-finger variants from a pre-selected pool of zinc-finger monomers (Maeder et al., 2008). The resulting library of zinc-finger arrays can then be interrogated using a bacterial two-

hybrid screen, where binding of the zinc-finger array to a pre-determined sequence results in the expression of a selectable marker gene. This method has generated highly-active zinc-finger nuclease pairs for sites within animal and plant genomes.

Zinc-finger nucleases have been widely used for plant genome engineering. Plant species that have been modified using zinc-finger nucleases include, soybean (Curtin et al., 2011), *Arabidopsis* (Lloyd et al., 2005; Zhang et al., 2010; Osakabe et al., 2010; Even-Faitelson et al., 2011), tobacco (Townsend et al., 2009; Wright et al., 2005; Marton et al., 2010; Zhang et al., 2013), maize (Shukla et al., 2009; Ainley et al., 2013), and BY2 tobacco cell suspension (Cai et al., 2009). With their relatively small size (~300 amino acids per zinc-finger nuclease monomer), and the further advancements in methods for redirecting targeting (Sander et al., 2011), zinc-finger nucleases should continue to be an effective technology for editing plant DNA.

1.3.3 Transcription Activator-Like Effector Nucleases

Similar to zinc-finger nucleases, transcription activator-like effectors nucleases (TALENs) are fusion proteins, consisting of a DNA-binding domain and a DNA-cleavage domain (Figure 1-3). Whereas, the DNA-cleavage domain is the same between zinc-finger nucleases and TALENs (the catalytic portion of *Fok*I), the DNA binding domains are different. The TALEN DNA binding domain is derived from TALE proteins found in the plant pathogen *Xanthamonas.* These proteins are composed of direct repeats of 33-35 amino acids, and nearly all arrays found in *Xanthamonas* contain a final, half repeat, consisting of the first 20 amino acids from the normal repeat. Two amino acids within these repeats (position 12 and 13) are responsible for recognizing a single nucleotide base (these amino acids are referred to as repeat-variable diresidues; RVDs). When the TALE effector code was broken (i.e., the relationship between the RVD and corresponding target base) (Boch et al., 2009; Moscou and Bogdanove, 2009), the ability to redirect targeting, and their use as a genome engineering tool, was realized (Christian et al., 2010; Li et al., 2011; Mahfouz et al., 2011).

Redirecting TALEN targeting involves the modular assembly of repeat sequences containing the appropriate RVD corresponding to the nucleotide target. The most widely used RVDs and their nucleotide targets are HD, cytosine; NG, thymine; NI, adenine; NN, guanine and adenine; NS, adenine, cytosine, and guanine; N*, all four nucleotides. This one-to-one correspondence of a single RVD to a single DNA base has eliminated construction challenges due to context-dependency seen with zinc-fingers and meganucleases. However, one limitation when using TALENs is that the target sequence must have a thymine at the -1 position (Boch et al., 2009). Furthermore, the long and repetitive nature of TALENs puts strain on delivery methods where cargo capacity or stability is a limitation.

Several studies have demonstrated the usefulness of TALENs in different plant species, including *Arabidopsis* (Christian et al., 2013), tobacco (Zhang et al., 2013; Cermak et al., 2011),

barley (Wendt et al., 2013), rice (Li et al., 2012) and *Brachypodium* (Shan et al., 2013a). Taken together, the modular nature of TALE repeats, along with efficient methods for assembling repetitive DNA sequences (Cermak et al., 2011; Reyon et al., 2012), have enabled TALENs to become one of the premier tools for plant genome engineering.

1.3.4 Clustered, Regularly Interspaced, Short Palindromic Repeat (CRISPR)/CRISPR Associated Systems

The most recent addition to the sequence-specific nuclease family is the clustered, regularly interspaced, short palindromic repeat (CRISPR)/ CRISPR associated (Cas) system. CRISPR/Cas systems are normally present within bacteria and archaea, and provide these cells with an adaptive immune system against invading plasmids or viruses. Here, CRISPR/Cas systems function to destroy invading nucleic acids by introducing targeted DNA breaks (Garneau et al., 2010).

There are three major types of CRISPR/Cas system: Types I – III (Makarova et al., 2011). The Type II system has been adopted for genome engineering (Cong et al., 2013; Mali et al., 2013b). With this system, there are two components that enable targeted DNA cleavage: a Cas9 protein and an RNA complex consisting of a CRISPR RNA (crRNA; contains 20 nucleotides of RNA that are homologous to the target site) and a transactivating CRISRP RNA (tracrRNA). DNA cleavage by the Cas9 protein will occur at sequences homologous to the crRNA sequence and upstream of a protospacer-adjacent motif (PAM; e.g., NGG for *Streptococcus pyogenes* Cas9). For genome engineering purposes, the system can be reduced in complexity by fusing the crRNA and tracrRNA to generate a single-guide RNA (sgRNA).

One limitation of the CRISPR/Cas system may be off-target cleavage (Fu et al., 2013; Cho et al., 2014). Target site recognition is facilitated through an RNA:DNA interaction (as opposed to a protein:DNA interaction used by meganucleases, zinc-finger nucleases, and TALENs). Redirecting Cas9 targeting involves modifying 20 nucleotides within the crRNA or sgRNA. Whereas these 20 nucleotides are used to direct Cas9 binding and cleavage, the system has been shown to tolerate mismatches, with a higher tolerance closer to the 5' end of the targeting sequence (Fu et al., 2013). Results from recent studies suggest the first 8-12 nucleotides, in addition to the PAM sequence, are most critical for target site recognition (Wu et al., 2014; Sternberg et al., 2014). To reduce off-targeting, several methods have been developed, including dual-nicking of DNA (Ran et al., 2013; Mali et al., 2013a), fusion of catalytically-dead Cas9 to *Fok*I (Guilinger et al., 2014; Tsai et al., 2014) and shortening of gRNA sequence (Fu et al., 2014).

Several plant species have been edited using CRISPR/Cas, including rice and wheat (Shan et al., 2013b), sorghum (Jiang et al., 2013) , tobacco (Li et al., 2013), and *Arabidopsis* (Feng et al., 2014; Li et al., 2013; Fauser et al., 2014). The simplicity of redirecting target

sequence, combined with the ease of expressing multiple sgRNAs for cleaving multiple sites within a genome, has resulted in the CRISPR/Cas system emerging as one of the most powerful and widely-used tools for genome engineering.

1.4 Conventional Methods for Delivering Reagents for Genome Editing to Plants

Whereas sequence-specific nucleases have enabled facile editing of genomes within living cells, editing plant DNA remains challenging. The probability of recovering a modified plant is not only dependent on the activity of the sequence-specific nuclease, but it is dependent on a number of additional factors, including the delivery method, the cellular conditions within the target cells, and the selection stringency. Here I describe commonly used methods for delivering genome engineering tools to plants. In general, there are two types of delivery methods: those that stably integrate reagents into plant genomes and those that transiently deliver reagents using *Agrobacterium* or direct plasmid delivery to protoplasts (Table 1).

1.4.1 Transient Delivery of Genome Engineering Reagents using *Agrobacterium* T-DNA

Agrobacterium is a broad host range soil bacteria and is one of the most widely used tools for transforming plants. When in contact with a plant cell, *Agrobacterium* injects a linear, single-stranded DNA molecule (T-DNA) into the plant cell's cytoplasm. This T-DNA then travels to the nucleus where it remains extrachromosomal, or integrates randomly into the host's genome. While extrachromosomal, coding sequences present on the non-integrated T-DNA molecules can be transcribed, leading to transient gene expression ~2-3 days after transformation (Kapila et al., 1997). If T-DNA is modified to harbor genome engineering reagents, both targeted mutagenesis and gene targeting can be facilitated.

The first demonstration of *Agrobacterium*-mediated genome editing in plants using sequence-specific nucleases was in 1996 (Puchta et al., 1996). Here, the target for repair was a defective *nptII* gene (containing an I-*Sce*I recognition site) that was stably integrated within the tobacco genome. Donor molecules were designed to restore *nptII* gene function. As a result of gene targeting, *nptII* gene expression would be reconstituted, and the respective cell would become resistant to kanamycin. Tobacco seedlings, carrying the defective *nptII* gene, were transformed with two T-DNAs, one encoding I-*Sce*I and the other harboring the corrective donor molecule. Transformed tissue was plated on regeneration media and kanamycin-resistant calli were regenerated. By comparing the number of true gene targeting events to the number of times the T-DNA integrated randomly into the tobacco genome, gene targeting frequencies were found to be between 2.2-18.3 x 10^{-3}, several orders of magnitude higher than the frequency without double-strand breaks (3.8 x 10^{-6}) (Hrouda and Paszkowski, 1994). This approach has been used to modify additional sequences within tobacco (Reiss et al., 2000; Li et al., 2013; Jiang et al.,

2013; Cai et al., 2009; Li et al., 2012), *Arabidopsis thaliana* (de Pater et al., 2009; Jiang et al., 2013) and sorghum (Jiang et al., 2013).

1.4.2 Stable Integration of Genome Engineering Constructs in Whole Plants

Constructs encoding sequence-specific nucleases (with or without donor molecules) can be randomly integrated into plant genomes. This strategy relies on gene modification to occur during the life of the plant with the anticipation that meiotic cells (or the precursor cells) will acquire a targeted modification, ultimately giving rise to seed with a heritable sequence alteration. This method may prove generally useful for plant genome engineering because a wide range of plant species can be transformed using *Agrobacterium* or biolistics.

To carry-out targeted mutagenesis, sequence-specific nucleases can be integrated into a plant's genome. One of the first examples demonstrating the effectiveness of this approach used the zinc-finger nucleases QQR to mutagenize a synthetic target site within *Arabidopsis* plants (Lloyd et al., 2005). In plants expressing the QQR zinc-finger nuclease, approximately 10% gave rise to seedlings with heritable mutations at the nuclease target site. Four years later, a similar approach was used to mutagenize *Arabidopsis*; however, here zinc-finger nuclease pairs were engineered to target sequence within the endogenous *ADH1* or *TT4* genes (Zhang et al., 2010); or the *ABAI4* gene (Osakabe et al., 2010). Plants harboring the *ADH1*- and *TT4*-targeting zinc-finger nucleases gave rise to mutant seed at a 69% and 33% frequency, respectively; and plants harboring the *ABAI4*-targeting zinc-finger nuclease had detectable mutations in floral tissue at a ~22% frequency (2 out of 9 plants). Additional studies have expanded upon this approach by using TALENs (Christian et al., 2013) and CRISPR/Cas (Feng et al., 2014; Fauser et al., 2014), and by mutagenizing additional plant species, including soybean (Curtin et al., 2011; Haun et al., 2014), tobacco (Petolino et al., 2010) *Arabidopsis* (Even-Faitelson et al., 2011) and barley (Wendt et al., 2013).

Gene targeting can be achieved by integrating sequence-specific nucleases and donor molecules into the plant's genome. Proof-of-concept experiments were demonstrated in *Arabidopsis* using I-*Sce*I and a donor molecule designed to repair a truncated beta-glucuronidase (*gus*) gene (Fauser et al., 2012). Here, the donor molecule was flanked with I-*Sce*I target sites; therefore, in plant cells expressing the meganuclease, the donor molecule will be liberated from the genome following two concerted I-*Sce*I-induced double-strand breaks. After screening seedlings from parent plants harboring all three genome engineering reagents (I-*Sce*I, donor molecule, and the disrupted *gus* target), the authors found that gene targeting occurred in 1.3% (189 blue seedlings/18,000 total seedlings) to 0.12% (18 blue seedlings/15,000 total seedlings) of seeds. Interestingly, the frequency of gene targeting was highly-dependent on the location where the donor molecule integrated: if the donor molecule was on the same chromosome as the target, gene targeting frequencies were significantly higher. A similar approach was used to modify *Zea*

mays DNA (Ayar et al., 2012); however, no modified seedlings were found from parental plants that contained the donor molecule, I-*Sce*I nuclease and target locus (0 out of 680 plants). To circumvent the challenge of transmitting a heritable germline mutation, the authors used tissue culture to regenerate plants from embryos isolated from immature kernels. From 2,356 embryos, seven modified plants were regenerated. Of these seven plants, two contained true gene targeting events, while the other five underwent recombination at the donor molecule locus, and not the target site (i.e., sequence from the target site was found at the locus containing the donor molecule). These results demonstrated that flanking the donor molecule with nuclease cut sites may not be necessary to facilitate gene targeting, but instead, may be a source of unwanted rearrangements.

Whereas stable integration of genome engineering reagents has proven to be effective for modifying plant DNA, there are several inefficiencies with this approach that may hinder its widespread use. First, the frequency of modification is highly variable across different transgenic plants, even when the genome engineering reagents are identical (Christian et al., 2013; Fauser et al., 2014; Zhang et al., 2010). This variability is most likely caused by epigenetic differences at the site of integration (affecting the expression of the sequence-specific nuclease) or by the spatial proximity of the donor molecule to the target locus (with respect to gene targeting). In addition, this method is also time-consuming and labor intensive. The minimum time to obtain a modified plant is the time it takes the plant to go from a seed to producing seed. For model organisms, including *Arabidopsis* and tobacco, this is approximately 3-4 months; however, for economically important crops, like wheat and corn, this is approximately 4 to 9 months (Petolino and Arnold, 2009). Furthermore, this number does not include the time it takes to generate the transgenic plant, to screen the seeds for modification, or to perform backcrosses to remove unwanted transgenes or off-target modifications.

1.4.3 Direct Delivery of Genome Engineering Reagents to Protoplasts

The highest frequencies of gene editing have been achieved using protoplasts–plant cells lacking cell walls. Protoplasts can be transformed at high frequencies (usually 50-90% of protoplasts receive DNA), which enables the efficient delivery of sequence-specific nucleases and donor molecules. Proof-of-concept experiments were demonstrated in tobacco protoplasts harboring a disrupted *gus:nptII* transgene where both *GUS* and *NPTII* were rendered nonfunctional by a 600 bp deletion encompassing essential regions within both genes (Wright et al., 2005). Transgenic protoplasts were transformed with plasmids encoding the zinc-finger nuclease Zif268:*Fok*I (the target site is present within *gus:nptII*) and a donor molecule designed to repair *the gus:nptII* transgene. After transfection, protoplasts were grown on media with plant hormones, and kanamycin-resistant calli were regenerated. A very high frequency of gene targeting was obtained (~10^{-1}; compare to 10^{-6} when a sequence-specific nuclease is not

delivered). Several years later, this approach was used to confer herbicide tolerance to tobacco plants. Here, the endogenous acetolactate synthase genes (*ALS SurA* and *SurB*) genes were modified with several nucleotide substitutions, resulting in amino acid changes known to confer resistance to imidazolinone and sulphonylurea herbicides (Townsend et al., 2009). Similarly, gene targeting frequencies were remarkably high (~0.4 x 10^{-1}). Additional studies have demonstrated the efficacy of this approach in tobacco protoplasts (Li et al., 2013; Zhang et al., 2013; Cermak et al., 2011), *Arabidopsis* protoplasts (Li et al., 2013; Cermak et al., 2011), wheat protoplasts (Shan et al., 2013b), *Brachypodium* protoplasts (Shan et al., 2013a) and rice protoplasts (Shan et al., 2013b, 2013a; Jiang et al., 2013).

In many instances, the goal of a genome engineering project is to obtain whole plants for phenotypic and genotypic characterization. Whereas protoplasts can be transformed at high efficiency, only a handful of plants can be effectively regenerated from protoplasts, and from the plants that can be regenerated, significant tissue culture expertise is required, which is something that not all plant research labs have. These limitations hinder the general-use of protoplast transformation for plant genome engineering.

1.4.4 Other Methods for Delivering Genome Engineering Reagents to Plants

Besides using *Agrobacterium* to transiently deliver or stably integrate genome engineering reagents into plants, or using direct-delivery to protoplasts, there are three other methods that have been adopted for delivering genome engineering tools to plant cells: whiskers-mediated transformation, biolistic bombardment and RNA viruses (Table 1-1). Whereas there are only a few examples using these methods, they have proven to be effective for editing DNA in corn, rice, tobacco and petunia.

1.4.4.1 Whiskers-Mediated Transformation

One challenge with delivering DNA (or RNA or protein) is the thick cell wall that surrounds each plant cell. Outside of using *Agrobacterium,* or chemicals/enzymes that remove or loosen the cell wall, DNA can be delivered by piercing small holes through the cell wall and plasma membranes. Whiskers-mediated transformation involves agitating plant cells with silicon carbide whiskers (10 to 80 μm in length, 0.6 μm wide) and DNA of interest. This agitation results in small, micron-sized punctures in the cell wall and membranes, thereby allowing macromolecules, including DNA, to enter into the cell's nucleus. Using this method, the crop species *Zea mays* was modified to have herbicide tolerance and to have reduced phytate levels (phytate is an anti-nutritional component of feed and is a pollutant) (Shukla et al., 2009). To this end, embryonic maize tissue cultures were transformed with plasmid DNA containing zinc-finger nuclease pairs targeting the endogenous *IPK1* gene (*IPK1* catalyzes the final step in phytate biosynthesis in maize seeds) along with a donor molecule designed to introduce the *PAT*

herbicide-tolerance gene into the *IPK1* coding sequence. Following whiskers-mediated transformation, herbicide-tolerant calli were regenerated and assessed for gene targeting events. From a total of 535 events generated from an autonomous PAT donor molecule, which can confer herbicide-tolerance when integrated both through gene targeting and illegitimate recombination, 100 plants contained true gene targeting events. Therefore, the gene targeting frequency was 1.9×10^{-1}.

1.4.4.2 Biolistic Bombardment

Similar to whiskers-mediated transformation, biolistic bombardment delivers DNA directly to plant cells by penetrating through the cell wall and membranes. Here, metal beads (usually gold or tungsten) are coated with DNA and placed inside a gene gun. The gene gun is then used to deliver the metal beads at a high velocity to whole plants, resulting in some plant cells receiving a metal bead in their nucleus. In addition to whole plants, this method can also be used to transform callus cells and cell cultures. Using this approach, genome engineering reagents were delivered to rice for the purpose of knocking out the phytoene desaturase gene (*PDS;* homozygous knockout mutations within *PDS* result in an albino phenotype) (Shan et al., 2013b). Here, metal beads were coated with plasmids harboring a plant codon-optimized Cas9 gene and sgRNAs targeting *PDS*. Rice calli were bombarded, and hygromycin-tolerant calli were grown into whole plants (the hygromycin-resistance gene was also present on the plasmids). Of ~200 transgenic plants, 16 were found to harbor somatic mutations, and 3 of these 16 plants had an albino and dwarf phenotype. Molecular analysis of the *PDS* gene revealed that all three plants contained homozygous mutations. Additional studies in maize showed similar results (Ainley et al., 2013). Here, plasmids harboring zinc-finger nucleases and donor molecules were bombarded into maize embryos. The embryos contained a pre-integrated 'trait-landing pad' harboring the corresponding nuclease target site. After selecting for cells containing the herbicide-resistant gene aryloxyalkanoat dioxygenase (*AAD1*; present on the donor molecule), the authors found that up to 5% of the regenerated plants contained the desired genome modification. These two studies in rice and maize demonstrated the usefulness of biolistic bombardment for delivery of genome engineering reagents.

1.4.4.2 RNA Viruses

Viruses have long been used for genome engineering in mammalian cells (Ellis and Bernstein, 1989; Lombardo et al., 2007; Wang and Taylor, 1993; Russell and Hirata, 1998); however, they have only recently been applied in plant cells. The first virus to be modified for the purpose of editing plant DNA was tobacco rattle virus (Marton et al., 2010). Tobacco rattle virus is an RNA virus and contains two genomes, RNA 1 and RNA 2. Heterologous sequence can be introduced into the virus genome by replacing two genes (2b and 2c) within the RNA 2 genome.

Heterologous sequences less than 3 kb will be replicated and systemically spread by the virus. Using this approach, tobacco rattle virus was modified to encode the zinc-finger nuclease QQR. The target site for this zinc-finger nuclease pair was a non-functional *gus* reporter within tobacco and petunia plants. Imprecise repair of the QQR target site by non-homologous end joining could reconstitute GUS expression. Within tobacco and petunia plants that were infected with this recombinant RNA virus, somatic cells were found to express GUS, and molecular characterization of the QQR target site revealed the presence of non-homologous end joining mutations. To generate whole plants harboring a stable mutation at the QQR target site, tissue from infected parental plants was placed onto media with growth hormones and mutated plants were regenerated. Whereas these results demonstrate the utility of tobacco rattle virus for delivering sequence-specific nucleases, there are several disadvantages that limit the general use of RNA viruses for genome editing: the relatively small cargo capacity of tobacco rattle virus (< 3 kb) restricts this method to delivering zinc-finger nucleases or meganucleases, and tobacco rattle virus is strictly RNA, and therefore cannot deliver donor molecules for gene targeting.

1.5 Geminiviruses and Their Use as Plant Vectors

Geminiviruses are a large family of plant viruses with circular single-stranded DNA genomes. The discovery that geminiviruses replicate their single-stranded DNA genomes through double-stranded intermediates using plant polymerases (in contrast to other plant viruses that use error-prone reverse transcriptases) led to the speculation that geminiviruses could serve as stable vectors for transient protein expression in plant cells. Furthermore, their broad host range makes them an attractive vector system for most plant species, including those that are hard to transform. Therefore, there may be potential for geminiviruses to deliver genome engineering reagents to plant cells. Here, we review the geminivirus life cycle and their use as plant vectors.

1.5.1 Geminivirus Biology

Geminiviruses–named after their twin particle morphology–are capable of infecting a wide range of plants, including species from both monocotyledonous and dicotyledonous groups. Geminiviruses are responsible for significant crop loss worldwide and have been documented to completely destroy fields containing corn, cassava and maize (Vanderschuren et al., 2007). African cassava mosaic virus (ACMV) alone is responsible for an economic loss of approximately US $2 billion per year (Varma and Malathi, 2003). Currently, geminiviruses are classified into seven genera: *Begomovirus, Mastrevirus, Curtovirus, Becurtovirus, Eragrovirus, Topocurvirus* and *Turncurtovirus* (Adams et al., 2013). Classification is based upon insect vector, host range and genome organization. Whereas most genera contain a single circular DNA genome of approximately 2.5 – 3.0 kb, some begomoviruses contain two genomes, termed DNA A and DNA B (Figure 1-4). Geminivirus genomes encode 4 – 7 proteins, and many of these proteins are

multifunctional. Geminiviruses do not encode polymerases and are heavily reliant on host proteins to complete their life cycle.

Geminivirus infection begins by mechanical delivery of virions (the virus genome bound to coat protein) to plant cells (usually phloem cells) by insect vectors, followed by their replication and systemic movement throughout the plant. Many geminiviruses remain phloem-limited; however, some have been found to invade surrounding mesophyll tissue (spongy and palisade parenchyma) as well as epidermal cells. Furthermore, geminiviruses do not invade meristematic tissues (Horns and Jeske, 1991). This observation helps to explain why geminiviruses are not found in the seed of infected plants, and how culturing the growing points of infected plants can result in the regeneration of virus-free plants.

Once delivered to plant cells by insect vectors, virions are transferred to the host cell's nucleus and replication proceeds. Little is known about the virion disassembly process (Kittelmann and Jeske, 2008). For example, it is unknown whether geminivirus DNA is uncoated within the cytoplasm or nucleus. Furthermore, it remains to be shown whether the virus travels through nuclear pores or requires the nuclear envelope to disintegrate. Once uncoated and inside the nucleus, the single-stranded DNA genome is converted into a double-stranded intermediate by host DNA polymerases. With respect to mastreviruses, complementary-strand synthesis requires the non-coding small intergenic region (SIR) found within the virus genome (Kammann et al., 1991; Shen and Hohn, 1991), and a short primer (~80 nucleotides) that binds to sequence within the SIR. This primer is used by host DNA polymerases to initiate complementary-strand synthesis, and is packaged within virions along with the mastrevirus genome (Donson et al., 1984). All other geminiviruses, which do not contain SIR sequence, require the de novo generation of primers (Saunders et al., 1992). Following complementary-strand synthesis, the double-stranded intermediate is packaged into minichromosomes (Pilartz and Jeske, 1992) and serves as a template for transcription of virus genes and for rolling-circle replication. Both transcription and rolling-circle replication require the non-coding large intergenic region (LIR; *Mastrevirus*), intergenic region (IR; *Curtovirus* and *Topocurvirus*) or common region (CR; *Begomovirus* with bipartite genomes). The LIR, IR and CR function as bidirectional promoters for transcription of complementary- and virion-sense genes. These regions also contain the origin of replication (DNA sequences that recruit virus-specific proteins for the purpose of initiating rolling-circle replication).

The only virus protein necessary for replication is the replication initiator protein (Rep; also referred to as the replication associated protein). In all genera, Rep coding sequence is present in the complementary-sense direction and transcription is initiated by the complementary-sense LIR, IR or CR promoter. Specifically, for mastreviruses, in the absence of Rep protein, the LIR promoter is strong in the complementary-sense direction, thereby promoting high expression of the Rep gene during the very early stages of infection. However, Rep protein negatively

regulates its own gene expression; therefore, accumulation of Rep results in decreased gene expression from the complementary-sense LIR promoter (Hefferon et al., 2006).

Geminiviruses replicate primarily by a rolling-circle mechanism, similar to some bacteriophages. The Rep coding sequence contains three motifs that are highly conserved with prokaryotic rolling-circle replication proteins (Koonin and Ilyina, 1992): Motif I (FLTYPxC; unknown function), Motif II (HLHxxxQ; thought to be involved in coordinating a divalent cation required for rolling-circle replication), and Motif III (VxDYxxK; the tyrosine is required for nicking the DNA to initiate rolling-circle replication (Laufs et al., 1995a)). Rolling-circle replication begins when Rep protein binds to a conserved Rep-binding site located within the LIR, IR or CR (Figure 1-5). Once there, Rep creates a single-strand nick at the top of a nearby hairpin structure and within a conserved 9-nt sequence (TAATATT**|**AC; the line indicates the site of the single-stranded nick; (Heyraud-Nitschke et al., 1995; Laufs et al., 1995b)). Following cleavage, Rep protein remains covalently bound to the 5' adenosine (i.e., Rep-5'-AC…), and the 3' thymine serves as a primer for host DNA polymerases. Once the DNA polymerases have replicated the entire geminivirus genome and return back to the conserved 9-nt sequence, Rep creates another single-strand nick, thereby releasing a unit-length copy of the geminivirus genome. The Rep protein that is covalently bound to the adenosine nucleotide then ligates the two DNA ends, resulting in an additional circular, single-stranded geminivirus genome (Heyraud-Nitschke et al., 1995; Laufs et al., 1995b).

Replication was thought to mainly proceed by a rolling-circle mechanism; however, recent evidence demonstrates that some begomoviruses and curtoviruses also replicate through recombination-dependent replication (Alberter et al., 2005; Preiss and Jeske, 2003). Here, free 3' ends of the single-stranded DNA genome invade a homologous double-stranded DNA sequence. This homologous sequence is then used as a template for synthesis and elongation of the invading 3' end. The free 3' ends are thought to be generated by either incomplete synthesis or from nucleolytic attacks. This method of replication may explain the high rates of intra- and inter-genome recombination observed with geminivirus infection (Padidam et al., 1999; Frischmuth and Stanley, 1998).

Paradoxically, geminiviruses rely on host proteins for replication, yet they only infect cells that have exited the cell-cycle and do not have the resources for DNA replication. To overcome this barrier, geminiviruses encode proteins that re-activate S phase in resting plant cells. These proteins are RepA (*Mastrevirus*) and Rep (all other geminivirus genera). RepA contains a highly-conserved LXCXE motif (known to enable binding to the human retinoblastoma protein). This observation led to the discovery that RepA binds to the plant retinoblastoma related protein (pRBR) and most likely relieves the G1/S checkpoint caused by pRBR/E2F repression (Liu et al., 1999). Interestingly, this LXCXE motif is also present within the *Mastrevirus* Rep protein (Rep and RepA proteins are produced from the same pre-mRNA with RepA being translated from the non-

spliced version of this transcript); however, Rep does not bind to pRBR. The inability of Rep to bind pRBR is thought to be due to Rep-specific C-terminal amino acids which mask the LXCXE binding activity (Liu et al., 1999). In other Rep proteins from the remaining geminivirus genera, the LXCXE motif is not present; instead, there are other motifs that facilitate binding to pRBR (Kong et al., 2000).

Movement of geminivirus virions from cell-to-cell occurs through the plant's endogenous plasmodesmata pathways. Plasmodesmata are functionally analogous to gap junctions in human cells: they are cytoplasmic bridges connecting neighboring cells, and are used as passageways for signaling molecules (e.g., short interfering RNA). Plant viruses, both DNA and RNA, exploit these passageways for their movement. Normally, the diameter of a plasmodesma is much smaller than a virion; therefore, plant viruses have adopted several different strategies to overcome this barrier, including the widening of the plasmodesma opening, assembly of secondary plasmodesma, or stretching of their nucleic acid genomes to fit through the small opening (Waigmann et al., 2004). For example, the movement protein from bean dwarf mosaic virus (*Begomovirus*) dilates plasmodesmata openings (Noueiry et al., 1994) to enable passage of its ~2.6 kb circular DNA A and DNA B genomes. Yet, even with mechanisms to make travel through plasmodesma amenable, movement through these junctions imposes strict size requirements on geminivirus genomes: larger-than-wild type genomes cannot pass through the plasmodesmata; instead, these genomes recombine to smaller sizes before moving to neighboring cells (Gilbertson et al., 2003).

Geminivirus proteins, in addition to their role in genome replication and movement, redirect and reprogram many cellular processes. Microarray analysis of transcripts within infected leaves revealed 5,365 differentially regulated host genes (Ascencio-Ibáñez et al., 2008). These genes were associated with several core features, including DNA replication, cell cycle control, and DNA repair. Geminivirus proteins interact with numerous host proteins, including proliferative cell nuclear antigen (Bagewadi et al., 2004), RAD54 (Kaliappan et al., 2012), adenosine kinase (Buchmann et al., 2009), SINAC1/ATF1 (Selth et al., 2005), GRAB2 (Xie et al., 1999; Lozano-Duran et al., 2011), BRN2 (Lozano-Duran et al., 2011), PEAPOD2 (Lacatus and Sunter, 2009), JDK (Lozano-Duran et al., 2011), AS1 (Yang et al., 2008), ATHB12 (Park et al., 2011), RFC (Luque et al., 2002), RPA32 (Singh et al., 2008), RAD51 (Suyal et al., 2013a), MCM2 (Suyal et al., 2013b), GRIMP (Kong and Hanley-Bowdoin, 2002), RBR (Kong et al., 2000; Horváth et al., 1998; McGivern et al., 2005; Settlage et al., 2001), CYCD3;1 (Ascencio-Ibáñez et al., 2008), E2FB (Ascencio-Ibáñez et al., 2008), E2F (Egelkrout et al., 2001), IK/KRP (Lai et al., 2009), RKP (Lai et al., 2009), H3 (Kong and Hanley-Bowdoin, 2002; Zhou et al., 2011), SET7/9 (Lozano-Duran et al., 2011), NSI (McGarry et al., 2003; Carvalho and Lazarowitz, 2004; Carvalho et al., 2006; Lozano-Duran et al., 2011), NIG (Carvalho et al., 2008a), Importin alpha (Lozano-Duran et al., 2011; Chandran et al., 2012), deltaCOP (Lozano-Duran et al., 2011), SYTA (Lewis and

Lazarowitz, 2010), cpHSC70 (Krenz et al., 2010, 2012), NIK1,2,3 (Fontes et al., 2004), rpl10 (Rocha et al., 2008; Carvalho et al., 2008b), NsAK (Florentino et al., 2006), BAM1 (Lozano-Duran et al., 2011), LRR-RLK (Piroux et al., 2007), GRIK (Shen and Hanley-Bowdoin, 2006), SnRK1 (Shen et al., 2011), SnRK2.1 (Lozano-Duran et al., 2011), SK2/SISK (Lozano-Duran et al., 2011; Piroux et al., 2007; Dogra et al., 2009), CRL (Trejo-Saavedra et al., 2009), ATJ3 (Lozano-Duran et al., 2011), SCE1 (Castillo et al., 2004; Sánchez-Durán et al., 2011), UBA1 (Lozano-Duran et al., 2011), UBC3 (Dogra et al., 2009), CAND1 (Lozano-Duran et al., 2011), RHF2A (Lozano-Duran et al., 2011), ADK (Buchmann et al., 2009), CSN5 (Lozano-Durán et al., 2011), CUL1 (Lozano-Durán et al., 2011), SAMDC1 (Zhang et al., 2011), SGS3 (Muangsan et al., 2004; Glick et al., 2008), RDR6 (Blevins et al., 2006), DCL4 (Blevins et al., 2006), HEN1 (Akbergenov et al., 2006), GDU1,3 (Chen et al., 2010), PR1 (Ascencio-Ibáñez et al., 2008), ACD6 (Yang et al., 2013), GSTF14 (Yang et al., 2013), GLO1 (Lozano-Duran et al., 2011), SKL2 (Lozano-Duran et al., 2011), LeHT1 (Eybishtz et al., 2009), AT4CL1 (Lozano-Duran et al., 2011), AOC1 (Lozano-Duran et al., 2011), F14P1.1 (Lozano-Duran et al., 2011), RD21/CYP1 (Lozano-Duran et al., 2011; Bar-Ziv et al., 2012), PLP2 (Lozano-Duran et al., 2011), GroEL (Morin et al., 1999), HSP16 (Ohnesorge and Bejarano, 2009). Other changes that occur within infected plant cells include a global decrease in chromosomal cytosine methylation due to adenosine kinase inactivation (Buchmann et al., 2009), inhibition of the post-transcriptional gene silencing pathway (Wang et al., 2014; Vanitharani et al., 2004; Cui et al., 2005; Gopal et al., 2007; Zrachya et al., 2007; Luna et al., 2012; Sharma and Ikegami, 2010; Zhang et al., 2012), increased frequencies of homologous recombination (Richter et al., 2014), and induction of the endocycle (Ascencio-Ibáñez et al., 2008).

1.5.2 Engineering Geminiviruses as Plant Vectors

Whereas geminiviruses are potent plant pathogens, their ability to replicate extrachromosomally to high copy number has tremendous potential for use as plant vectors, and possibly for the delivery of genome engineering reagents. Most frequently, heterologous sequences are introduced into geminivirus genomes for the purpose of knocking-down host gene expression (virus-induced gene silencing) and for the expression of desired proteins. Both of these applications require the insertion of heterologous sequence into wild-type genomes. However, where is the best place to insert this sequence? Here, I discuss the different approaches for introducing foreign sequences into geminivirus genomes.

There are two approaches for modifying geminivirus genomes: the full virus and deconstructed virus strategies (Gleba et al., 2004, 2007; Figure 1-6). Under the full virus strategy, foreign DNA is embedded into the geminivirus genome such that all (or most) of the virus genes and *cis*-acting replicational elements remain functional. As a result, recombinant viruses maintain the ability to replicate by rolling-circle replication and to move from cell-to-cell within a plant.

Heterologous sequence can be introduced, for example, into a non-coding region that can tolerate extra nucleotides. Using this approach, short sequences (2 to 32 nt) were inserted into the SIR of maize streak virus (Shen and Hohn, 1991); and ~100 nt were inserted into the DNA B component of tomato golden mosaic virus immediately following the nuclear shuttle protein (Peele et al., 2001). In both cases, recombinant geminiviruses moved systemically throughout the plant.

Heterologous sequence can also be introduced as a coat protein gene replacement. Interestingly, the coat protein is not required for cell-to-cell movement in some bipartite begomoviruses (Gardiner et al., 1988). Using this approach, African cassava mosaic virus and tomato golden mosaic virus were modified to encode chloramphenicol and neomycin phosphotransferase, respectively (Hayes et al., 1988; Ward et al., 1988). Inoculation of plants with these recombinant viruses resulted in systemic spreading and protein expression. Furthermore, replacement of the coat protein gene within the tomato golden mosaic virus and cabbage leaf curl virus with 800 nt of sequence homologous to an endogenous plant gene resulted in systemic spreading of the virus and subsequent knock-down of host gene expression (Kjemtrup et al., 1998; Muangsan and Robertson, 2004). Unfortunately, the full virus strategy has limitations, including a relatively small cargo capacity. The largest sequence that can be introduced is ~800 nt (as a coat protein replacement in bipartite begomoviruses; Figure 1-6). Sequences greater than 800 bp, or those that cause the virus genome to be larger than wild-type, result in genome instability and reduced infectivity (Gilbertson et al., 2003).

Under the deconstructed virus strategy, useful features are retained and the limiting or undesired features are removed. For example, to circumvent genome size constraints, proteins involved in cell-to-cell movement are removed (e.g., movement protein, coat protein) and the *cis*- and *trans*-acting replicational elements are retained (e.g., LIR, SIR and Rep/RepA for mastreviruses). By removing the elements required for movement, the virus no longer has the pressure to overcome the size restrictions imposed by the plasmodesmata; however, for some applications, cell-to-cell movement is beneficial. In effort to complement the loss of movement, *Agrobacterium* or biolistics are used to deliver replicon vectors to plant cells. Using this approach, the bean yellow dwarf virus (*Mastrevirus*) was modified to deliver reporter proteins (Huang et al., 2009; Mor et al., 2003; Regnard et al., 2010; Huang et al., 2010), monoclonal antibodies (Huang et al., 2010) and vaccine antigens (Huang et al., 2009; Regnard et al., 2010). The same methodology was used to deconstruct the beat curly top virus (*Curtovirus*) into a replicon for expressing the green fluorescent protein in tobacco (Kim et al., 2007). Currently, there appears to be no upper size limit for geminivirus replicons (Laufs et al., 1990; Matzeit et al., 1991; Shen and Hohn, 1994; Huang et al., 2009). Although, as replicon size increases, replication efficiency decreases (Suárez-López and Gutiérrez, 1997). This was demonstrated by increasing the size of wheat dwarf virus (*Mastrevirus*) replicons from 4.56 kb to 4.92 kb, bombarding wheat suspension

culture cells with these vectors, and then performing a southern blot to detect virus genomes. The larger geminivirus replicons were present at a lower copy number, relative to an internal control (a second geminivirus replicon with a 4.27 kb genome).

Dissertation Objectives

The work presented in this dissertation focuses on developing and applying methods for editing DNA in whole plants. Editing DNA in whole plants is particularly attractive due to the relative ease of handling and growing the material; however, there are very few methods for delivering genome engineering to whole plants. Therefore, developing new methods and expanding upon existing methods is critical for the continued progression of the plant genome engineering field. In this work, I explore several different delivery methods for modifying a variety of plant species. First, in chapter 2 and appendix B, I explore the capacity of geminiviruses to deliver genome engineering reagents to plant cells. Here, I describe the development, characterization and implementation of bean yellow dwarf virus vectors (*Mastrevirus*) for the simultaneous delivery of sequence-specific nucleases and donor molecules in *Nicotiana tabacum*. I then, in chapter 3, explore the ability of tobacco rattle virus to deliver zinc-finger nucleases to *Arabidopsis*. And finally, in chapter 4 and 5 and appendix A, I describe our efforts to expand upon existing delivery methods (stable integration of reagents into plant genomes) by knocking out or deleting host genes within *Arabidopsis* and soybean.

Table 1-1. Summary of publications related to plant genome engineering with sequence-specific nucleases and their delivery methods

Delivery Method	Sequence-specific nuclease	Donor molecule	Plant	Target(s)	Reference
Stable integration	Zinc-finger nuclease	No	*Arabidopsis thaliana*	Transgene	Lloyd *et al.* 2005
	Zinc-finger nuclease	No	*Arabidopsis thaliana*	*ADH1; TT4*	Feng *et al. 2010*
	TAL effector nuclease	No	*Arabidopsis thaliana*	*ADH1; TT4; MAPKKK1;DSK2B; NATA2; GLL22*	Christian *et al.* 2013
	CRISPR/Cas9	No	*Arabidopsis thaliana*	*TT4; GAI; BRI1; JAZ1; CHLI; AP1*; transgene	Feng *et al.* 2014
	CRISPR/Cas9	No	*Arabidopsis thaliana*	Transgene	Fauser *et al.* 2014
	Zinc-finger nuclease	No	*Glycine max*	*DCL1a/b; DCL4a/b; RDR6a; HEN1a;* transgene	Curtin *et al.* 2011
	TAL effector nuclease	No	*Glycine max*	*FAD2-1A/B*	Haun *et al.* 2014
	TAL effector nuclease	No	*Hordeum vulgare*	*HvPAPhy_a*	Wendt *et al.* 2013
	Meganuclease	Yes	*Arabidopsis thaliana*	Transgene	Fauser *et al.* 2012
	Meganuclease	Yes	*Zea mays*	Transgene	Ayar *et al*. 2012
	Zinc-finger nuclease	No	*Arabidopsis thaliana*	*ABI4*	Osakabe *et al.* 2010
	Zinc-finger nuclease	No	Tobacco	*Transgene*	Petolino *et al.* 2010
Transient delivery (Agrobacterium T-DNA)	Meganuclease	Yes	Tobacco	Transgene	Puchta *et al.* 1996
	Meganuclease	Yes	Tobacco	Transgene	Reiss *et al.* 2000
	Zinc-finger nuclease	Yes	*Arabidopsis thaliana*	Transgene	de Pater *et al.* 2009
	Zinc-finger nuclease	Yes	Tobacco (suspension culture cells)	*CHN50;* Transgene	Cai *et al.*2009
	CRISPR/Cas9	No	Tobacco	*PDS3*	Li *et al.* 2013
	CRISPR/Cas9	No	*Arabidopsis thaliana*	*PDS3; RACK1a/b/c*	Li *et al.* 2013
	TAL effector nuclease	No	*Oryza sativa*	*OsSWEET14*	Li *et al.* 2012
	CRISPR/Cas9	No	Tobacco	Transgene	Jiang *et al.* 2013
	CRISPR/Cas9	No	*Arabidopsis thaliana*	Transgene	Jiang *et al.* 2013
	CRISPR/Cas9	No	*Sorghum*	Transgene	Jiang *et al.* 2013
Transient delivery (protoplasts)	Zinc-finger nuclease	Yes	Tobacco	Transgene	Wright *et al.* 2005
	Zinc-finger nuclease	Yes	Tobacco	*SurA/B*	Townsend *et al.* 2009
	CRISPR/Cas9	No	*Oryza sativa*	*SWEET14*	Jiang *et al.* 2013
	CRISPR/Cas9	No	Tobacco	*PDS3*	Li *et al.* 2013
	CRISPR/Cas9	No	*Arabidopsis thaliana*	*PDS3; FLS2*	Li *et al.* 2013
	TAL effector nuclease	Yes	Tobacco	*SurA/B*	Zhang *et al.* 2013
	TAL effector nuclease	No	*Arabidopsis thaliana*	*TT4; ADH1*	Cermak *et al.* 2011
	TAL effector nuclease	No	Tobacco	*SurB*	Cermak *et al.* 2011
	CRISPR/Cas9	No	*Triticum aestivum*	*MLO*	Shan *et al.* 2013

	CRISPR/Cas9	Yes	*Oryza sativa*	*PDS; BADH2;* Os02g23823; *MPK2*	Shan *et al.* 2013
	TAL effector nuclease	No	*Brachypodium*	*SPL; SBP*	Shan *et al.* 2013
	TAL effector nuclease	No	*Oryza sativa*	*DEP1; BADH2*	Shan *et al.* 2013
Whiskers	Zinc-finger nuclease	Yes	*Zea mays*	*IPK1*	Shukla *et al.* 2009
Biolistic Bombardment	CRISPR/Cas9	No	*Oryza sativa*	*PDS*	Shan *et al.* 2013
	Zinc-finger nuclease	Yes	*Zea mays*	Transgene	Ainley *et al.* 2013
RNA virus	Zinc-finger nuclease	No	Tobacco	Transgene	Marton *et al.* 2010
	Zinc-finger nuclease	No	*Petunia hybrida*	Transgene	Marton *et al.* 2010

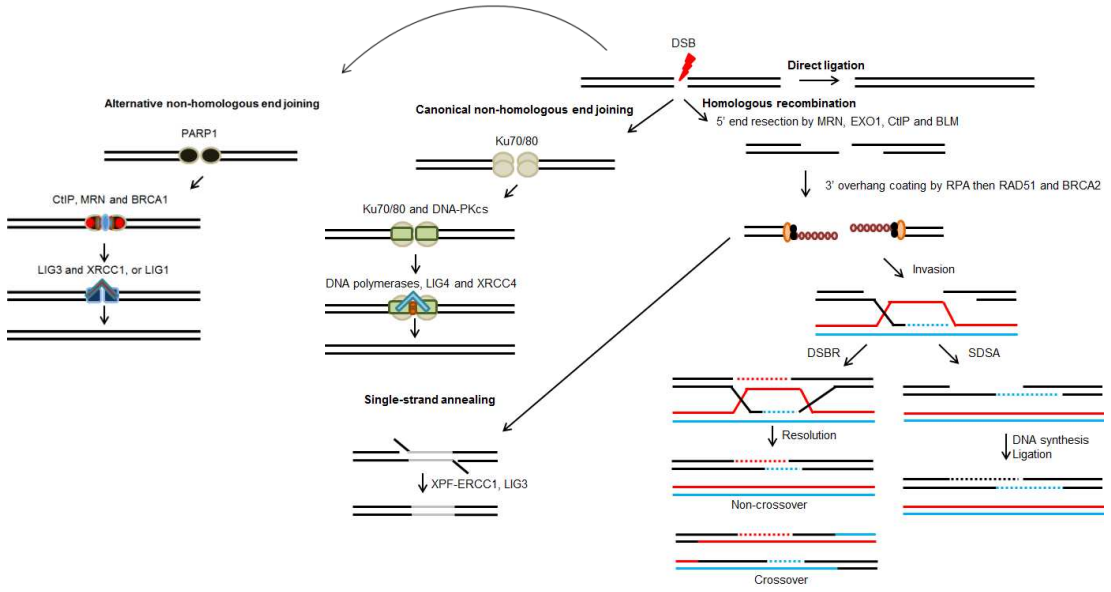

Figure 1-1. DNA double-strand break repair pathways. There are two pathways cells use to repair DNA double-strand breaks: non-homologous end joining (both canonical and alternative) and homologous recombination. In classical non-homologous end joining, the exposed DNA ends are recognized by the Ku70/80 heterodimer followed by the recruitment of DNA-PKcs. Gaps in DNA sequence are filled by polymerases γ and λ and then the DNA ends are ligated together by a LIG4 and XRCC4 complex. Repair by alternative non-homologous end joining involves PARP1 binding to the exposed DNA ends, followed by the recruitment of CtIP, MRN and BRCA1. Ligation of the DNA ends is carried out by either LIG3 and XRCC1 or LIG1. Repair of the double-strand break by homologous recombination begins with the resection of the DNA ends followed by coating of the exposed 3' ends with RPA. Rad51 is then recruited and facilitates invasion into a homologous sequence. Under the double-strand break repair (DSBR) model, both 3' ends invade a homologous sequence. Following extension of the 3' ends by host DNA polymerases, the Holliday junctions can be resolved to yield non-crossover or crossover products. Under the synthesis-dependent strand annealing (SDSA) model, the invading strand is displaced and then anneals to the non-invading strand. Double-strand breaks can also be repaired using direct-repeat sequences near the break site (single-strand annealing), resulting in the deletion of the sequence between the direct repeats. RPA, replication protein A; BRCA2, breast cancer type 2 susceptibility protein; Rad51, radiation sensitive 51 protein; MRN, MRE11-RAD50-NBS1; XRCC4, X-ray repair cross-complementing protein 4. Adapted from (San Filippo et al., 2008)

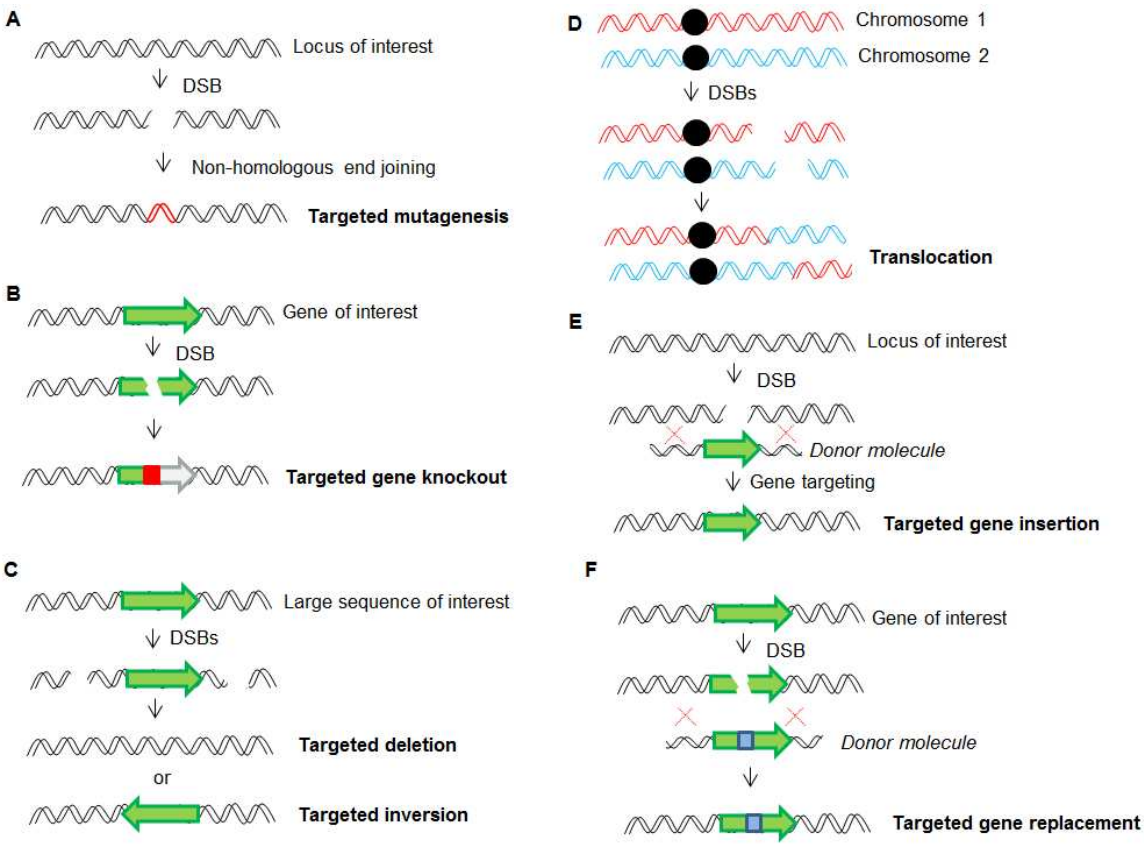

Figure 1-2. Plant genome engineering using DNA double-strand breaks. The repair mechanisms of DNA repair pathways can be exploited to introduce sequence changes into the host's genome. Repair of double-strand breaks by non-homologous end joining can result in deletions or insertions at the break-site. Therefore, targeting DNA breaks to a locus of interest, or a gene of interest, can facilitate (A) targeted mutagenesis or (B) targeted gene knockout, respectively. Targeting two breaks to two different sites within the same chromosome can facilitate (C) deletion or inversion of intervening sequence. Targeting two breaks to sites on two different chromosomes can result in translocations (D). Alternatively, double-strand breaks can stimulate recombination with a user-supplied donor molecule (gene targeting). Donor molecules can be designed to contain full genes or gene regulatory elements (targeted gene insertion; E), or point mutations for the purpose of making small changes within genes (targeted gene replacement; F). DSB, double-strand break.

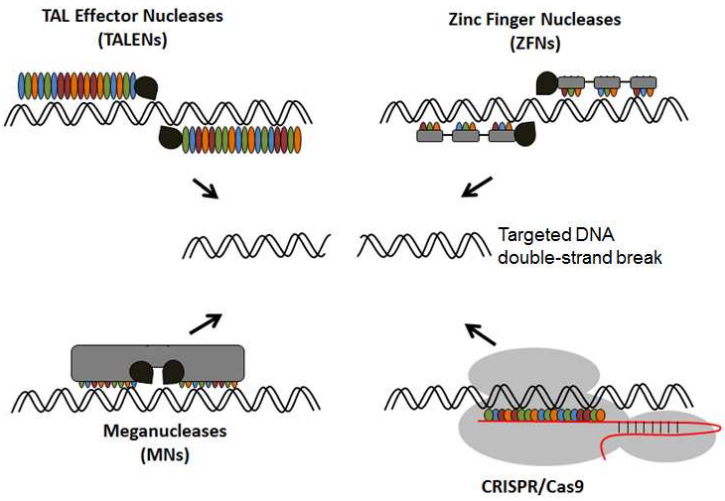

Figure 1-3. Classes of sequence-specific nucleases. Illustration of the four classes of sequence-specific nucleases: TAL effector nucleases, zinc-finger nucleases, meganucleases and CRISPR/Cas9. All classes of proteins can be 'rewired' to recognize and cleave desired DNA sequences. Whereas TAL effector nucleases, zinc-finger nucleases and meganucleases recognize target DNA through amino acids within the protein, the CRISPR/Cas9 system uses an RNA:DNA interaction (the guide RNA is represented by the red curved line).

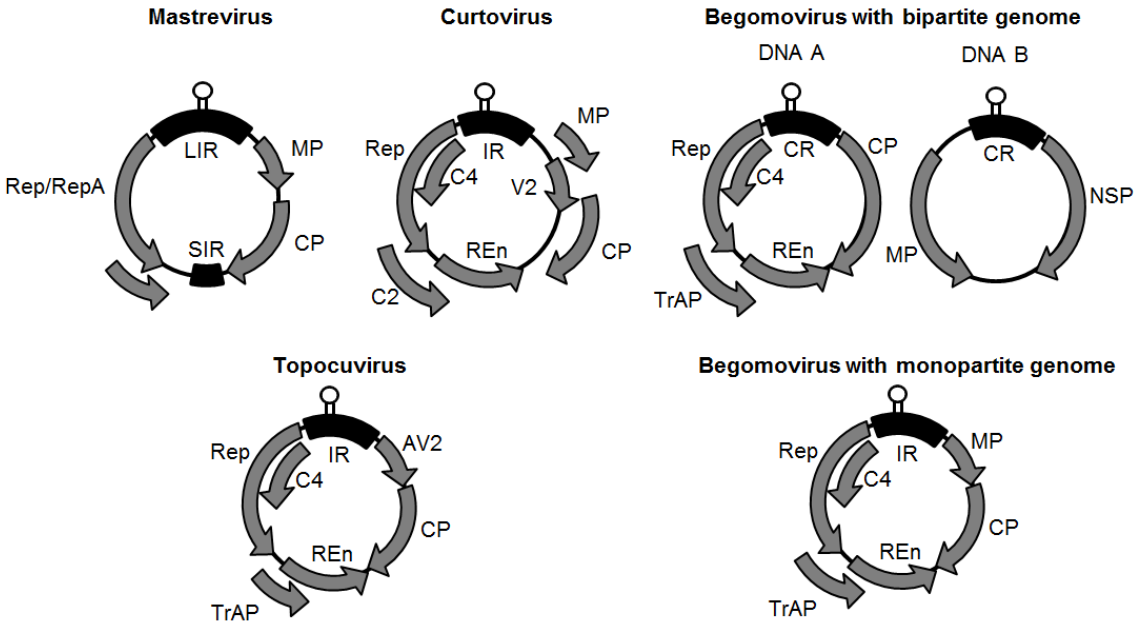

Figure 1-4. Genome organization of geminiviruses. Illustration of the genetic organization of geminiviruses within the genera *Mastreviruses, Curtoviruses, Topocuviruses* and *Begomoviruses* (both bipartite and monopartite). Black rectangles, non-coding sequences; gray arrows, coding regions; Rep, replication initiator protein; LIR, long intergenic region; SIR, short intergenic region; MP, movement protein; CP, coat protein; REn, replication enhancer gene; TrAP, transcription activator protein; NSP, nuclear shuttle protein.

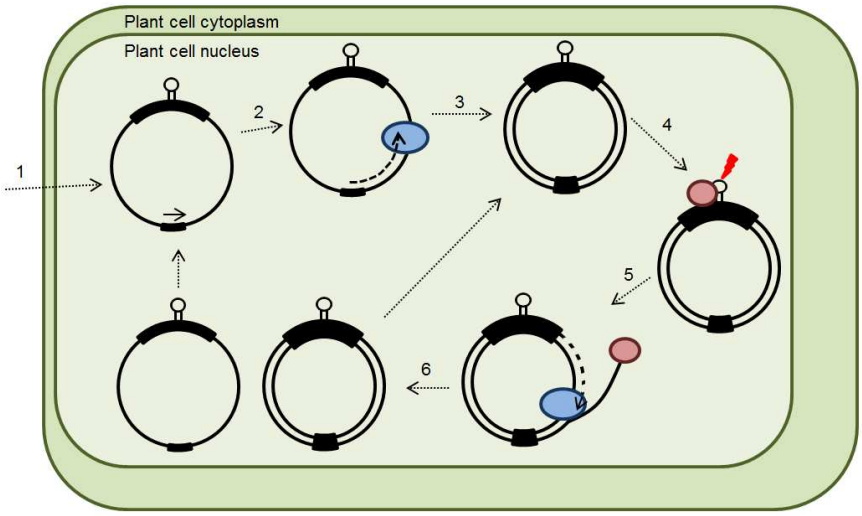

Figure 1-5. Rolling-circle replication of geminivirus genomes. (1) Geminivirus infection begins when a single-stranded virus genome enters the nucleus of a plant cell. The black curved rectangle at the top of the circular genome represents intergenic sequence: LIR, *Mastrevirus*; IR, *Curtovirus* and *Topocurvirus*; CR, *Begomovirus* with bipartite genomes. The small black rectangle at the bottom of the circular genome represents SIR (*Mastrevirus*). The lollipop structure represents the hairpin structure with the conserved nonanucleotide sequence (TAATATTAC). The arrow represents a small DNA primer (~80 nucleotides) bound to the SIR. (2,3) The single-stranded genome is then converted to a double-stranded intermediate by host DNA polymerases (blue circle). (4) Rep (red circle) binds to the origin of replication and creates a single-strand nick at the top of the hairpin structure. (5) The exposed 3' nucleotide is used as a primer for host DNA polymerases. DNA polymerases replicate the geminivirus genome, while Rep remains covalently linked to the 5' nucleotide. (6) When DNA polymerases return to the origin, Rep mediates the ligation of the single-stranded ends, thereby releasing a complete, single-stranded genome.

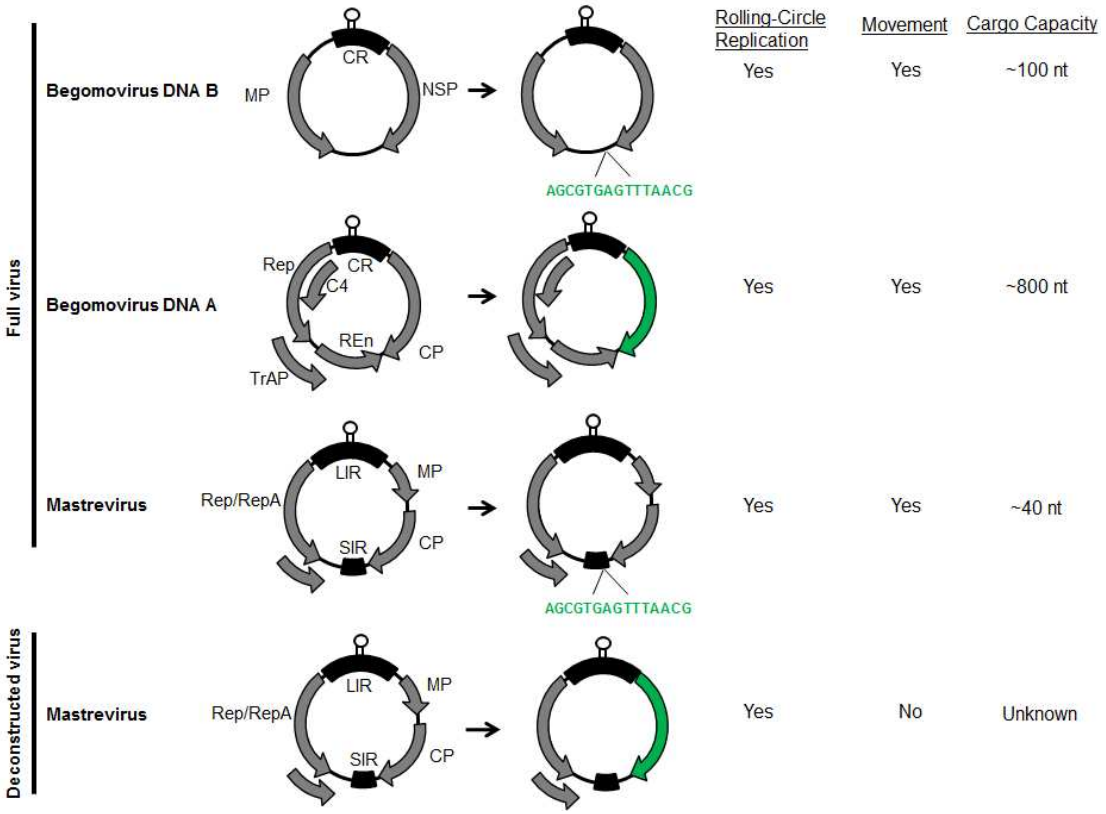

Figure 1-6 Strategies for converting geminiviruses into plant vectors. Geminivirus genomes can be modified to carry heterologous sequences. Under the full virus strategy, heterologous sequence is inserted as a coat protein gene replacement in bipartite begomoviruses, or it is inserted into regions that tolerate extra nucleotides. Viruses modified under this strategy retain cell-to-cell movement. Heterologous sequence can also be inserted by replacing movement protein (MP) and coat protein (CP) sequences within mastreviruses (deconstructed viruses). Cargo capacity refers to the approximate maximum number of nucleotides the virus can stably replicate (and move from cell to cell, if under the full virus strategy). Green text or arrows represent heterologous sequence.

CHAPTER 2

DNA REPLICONS FOR PLANT GENOME ENGINEERING

Reprinted with permission from Baltes, N. J., Gil-Humanes, J., Cermak, T., Atkins, P. A, & Voytas, D. F. (2014). DNA replicons for plant genome engineering. *The Plant Cell*, *26*(1), 151–63. doi:10.1105/tpc.113.119792; www.plantcell.org; Copyright American Society of Plant Biologists

ABRIDGEMENT

Sequence-specific nucleases enable facile editing of higher eukaryotic genomic DNA; however, targeted modification of plant genomes remains challenging due to ineffective methods for delivering reagents for genome engineering to plant cells. Here, we use geminivirus-based replicons for transient expression of sequence-specific nucleases (zinc-finger nucleases, ZFNs; transcription activator-like effector nucleases, TALENs; the CRISPR/Cas system) and delivery of DNA repair templates. In tobacco leaves (*Nicotiana tabacum*), replicons based on the bean yellow dwarf virus enhanced gene targeting frequencies one to two orders of magnitude over conventional *Agrobacterium* T-DNA. In addition to the nuclease-mediated DNA double-strand breaks, gene targeting was promoted by replication of the repair template and pleiotropic activity of the geminivirus replication initiator proteins. We demonstrate the feasibility of using geminivirus replicons to generate plants with a desired DNA sequence modification. By adopting a general plant-transformation method, plantlets with a desired DNA change were regenerated in less than six weeks. These results, in addition to the large host range of geminiviruses, advocate the use of replicons for plant genome engineering.

INTRODUCTION

Precise modification of plant DNA is of value for basic and applied biology, because it facilitates gene function studies and crop improvement. One method to modify DNA involves introducing targeted double-strand breaks. Double-strand breaks activate two main repair pathways, non-homologous end joining (NHEJ) and homologous recombination. These two pathways can be exploited to introduce sequence changes within the plant genome. Repair by NHEJ is often imprecise, leading to insertions or deletions of DNA at the cut site. Consequently, double-strand breaks directed to endogenous genes can facilitate targeted gene disruption or targeted mutagenesis (Lloyd et al., 2005; Zhang et al., 2010). Double-strand breaks can also be repaired with high-fidelity by homologous recombination. Homologous recombination normally involves the use of the sister chromatid or homologous chromosome for template-directed repair. If an exogenously-supplied repair template is present, gene replacement or targeted gene insertion can be carried out (referred to hereafter as gene targeting; Puchta et al., 1996; Bibikova et al., 2003).

Targeted double-strand breaks can be generated using sequence-specific nucleases. These nucleases include meganucleases (Smith et al., 2006), zinc-finger nucleases (ZFNs; (Kim et al., 1996), transcription activator-like effector nucleases (TALENs; Christian et al., 2010; Miller et al., 2011) and the clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas) system (Mali et al., 2013b; Cong et al., 2013). Using sequence specific-nucleases, plant genomes have been modified in many different ways, ranging from several

nucleotide substitutions (Townsend et al., 2009) to the targeted deletion of megabases of DNA (Qi et al., 2013a).

In plants, delivery of genome engineering reagents, including sequence-specific nucleases and repair templates, presents a barrier to efficiently achieving targeted genome modifications. For gene targeting, there are a few reports using *Agrobacterium* or physical means to deliver genome engineering reagents (Cai et al., 2009; Shukla et al., 2009); however, the highest frequencies of gene targeting have been achieved using protoplasts, plant cells lacking cell walls (Townsend et al., 2009; Wright et al., 2005; Zhang et al., 2010). Protoplasts can be transformed at high efficiency, but only a handful of plants can be effectively regenerated from protoplasts. Alternatively, constructs encoding nucleases and repair templates can be integrated into plant genomes (Fauser et al., 2012; Ayar et al., 2012). This so-called *in planta* genome engineering strategy relies on the gene targeting to occur during the life of the plant, with the anticipation that germline cells will be modified and ultimately give rise to seed with the desired sequence alterations.

In mammalian systems, effective delivery of genome engineering reagents has been achieved using viruses, including retroviruses (Ellis and Bernstein, 1989; Lombardo et al., 2007), adenoviruses (Wang and Taylor, 1993) and adeno-associated viruses (Russell and Hirata, 1998). In plants, the RNA virus, tobacco rattle virus, was used to deliver a ZFN to petunia (*Petunia hybrida*) and tobacco (*Nicotiana tabacum*) leaves, resulting in targeted modification of an integrated reporter gene in somatic cells (Marton et al., 2010). Whereas the cargo capacity of plant RNA viruses enables delivery of ZFNs or a meganuclease, it is not sufficient to deliver larger proteins, such as a pair of TALENs or Cas9 for the CRISPR/Cas system. Furthermore, RNA viruses cannot deliver DNA repair templates for gene targeting.

To develop an efficient and facile vector system for plant genome engineering, we focused our efforts on the geminiviruses. Geminiviruses are a large family of plant viruses with single-stranded, circular DNA genomes of approximately 2.5 - 3.0 kb. Once inside a host cell's nucleus, their single-stranded genome is converted to a double-stranded intermediate by host DNA polymerases. This double-stranded genome is then used as a template for transcription of virus genes and for rolling-circle replication. The only geminivirus protein necessary for replication is the replication initiator protein (Rep). Rep initiates rolling-circle replication by binding to a site within the large intergenic region (LIR, *Mastreviruses*), intergenic region (IR, C*urtoviruses* and *Topocuviruses*), or common region (CR, *Begomoviruses* with bipartite genomes). There it creates a single-strand nick within an invariant 9-nt sequence (TAATATTAC) located at the apex of a conserved hairpin structure. Following rolling-circle replication, single-stranded genomes are either converted back to a double-stranded intermediate, or encapsidated by coat protein to produce virions. These virions are then transported to adjacent cells through the plant's

endogenous plasmodesmata pathways.

Methods to engineer geminiviruses for protein expression include the full virus and deconstructed virus strategies (Gleba et al., 2004). Under the full virus strategy, geminiviruses retain most or all of the features required for replication and systemic infection of the host plant. Heterologous sequence can be introduced within the genome by replacing the coat protein gene (coat protein is not required for cell-to-cell movement in some bipartite begomoviruses; Gardiner et al., 1988). Using this approach, tomato golden mosaic virus and African cassava mosaic virus were modified to express neomycin phosphotransferase (NPT; Hayes et al., 1988) and chloramphenicol acetyltransferase (CAT; Ward et al., 1988), respectively.

Unfortunately, it may not be possible to deliver relatively large heterologous sequences while maintaining cell-to-cell movement and rolling-circle replication. Cell-to-cell movement through the endogenous plasmodesmata pathways imposes strict requirements on the geminivirus genome size (Gilbertson et al., 2003; Hayes et al., 1989; Etessami et al., 1989; Elmer and Rogers, 1990). Larger than wild type genomes frequently recombine to smaller sizes, and, as a result, are outcompeted during the infection process. Whereas most studies exploring genome size restrictions use bipartite begomoviruses, these constraints may also be experienced with other geminiviruses, including mastreviruses. For example, using the maize streak virus, approximately 3.2 kb of heterologous sequence was inserted into the small intergenic region. Insertion of sequence into this region does not disrupt virus replication or expression of virus proteins (Shen and Hohn, 1991). After delivery to leaf cells, the chimeric virus replicated efficiently; however, no systemic movement was observed (Shen and Hohn, 1994).

Under the deconstructed virus strategy, limiting or undesired virus functions are removed, and only the useful ‘blocks’ are kept. For example, to circumvent genome size constraints imposed by the plasmodesmata, movement protein and coat protein sequences are removed, whereas the *cis*- and *trans*-acting replicational elements are retained. Lack of cell-to-cell movement can be compensated by using *Agrobacterium* or particle bombardment to deliver viral vectors to cells. This approach has been successfully used to generate replicons based on the bean yellow dwarf virus for expression of reporter proteins, vaccine proteins and monoclonal antibodies (Mor et al., 2003; Huang et al., 2010; Regnard et al., 2010; Zhang and Mason, 2006; Huang et al., 2009). Currently, there is no obvious upper size limit for replicons that do not move from cell-to-cell (Huang et al., 2009; Matzeit et al., 1991; Laufs et al., 1990; Shen and Hohn, 1994). However, it may be that as replicon size increases, replication efficiency decreases (Suárez-López and Gutiérrez, 1997).

The aim of this study was to explore the use of geminiviruses for plant genome engineering. We demonstrate that DNA carried by geminiviruses can be used as a template for homologous recombination. We further demonstrate that geminivirus-based replicons enable

highly-efficient genome engineering. We show that this technology can be used to produce calli and plantlets with precise DNA sequence changes. And, lastly, we develop a simplified geminivirus vector system for efficient plant genome engineering.

RESULTS

Full Viruses for the Delivery of Repair Templates in *Arabidopsis*

We began our study by exploring the use of full viruses for the delivery of repair templates. Our target locus was the *Arabidopsis* (*Arabidopsis thaliana*) alcohol dehydrogenase (*ADH1*) gene. We modified the bipartite cabbage leaf curl virus (CaLCuV) DNA A genome to contain repair templates designed to introduce a unique 18 bp sequence within *ADH1* (see Supplemental Figure 1A online). After optimizing repair template length (Figure 2-1), *Arabidopsis* plants were infected with CaLCuV (harboring a repair template with 600 bp of total sequence) by particle bombardment (Figure 2-2). Within the genome of the infected plants was a ZFN pair that targets a site within the *ADH1* gene (Zhang et al., 2010). This ZFN pair was under the transcriptional control of an estrogen-inducible promoter (Zuo et al., 2000). As a result, double-strand breaks could be temporally controlled by spraying *Arabidopsis* plants with a solution containing estradiol (Figure 2-3). Recombinant CaLCuV genomes were given 14 days to replicate and spread before ZFN expression was induced. Seven days post spraying, genomic DNA was isolated and the *ADH1* gene was assessed for the presence of the unique 18 bp sequence. From 23 experiments, we observed a single plant where the gene targeting product could be detected by PCR (Figure 2-4). Whereas, these results suggest gene targeting with full viruses is possible, further optimization of this technology was warranted.

Our approach using full viruses has several potential drawbacks, including (i) the uncertainty of having both gene targeting reagents (ZFN protein and geminivirus DNA) in a nucleus at the same time and (ii) the variability in the concentration of these reagents within the same cell. For example, because we did not observe NHEJ-induced mutations within all *ADH1* loci, we could not ensure that double-strand breaks were being produced in all cells after spraying with estradiol. Furthermore, leaky expression of the ZFN pair from the XVE promoter may result in premature double-strand breaks and subsequent NHEJ-mutations that destroy the ZFN binding domain. Due to these uncertainties, we sought to develop a system where (i) geminivirus vectors could be efficiently delivered to target cells of interest and (ii) sequence-specific nucleases and repair templates are simultaneously delivered.

Deconstructed viruses for the Delivery of Sequence-Specific Nucleases and Repair Templates

We chose to deconstruct the mild strain of the bean yellow dwarf virus (BeYDV; Halley-

Stott et al., 2007) because it had previously been used to express heterologous proteins in plant cells (Regnard et al., 2010). BeYDV replication requires three viral elements: the *cis*-acting LIR, the short intergenic region (SIR), and the *trans*-acting replication-initiation protein (Rep) and RepA. Rep and RepA are expressed from the same precursor mRNA: Rep is produced from the spliced mRNA whereas the unspliced mRNA yields RepA (Wright et al., 1997). Eliminating coding sequences for the coat and movement proteins abolishes cell-to-cell movement, but effectively eliminates genome-size constraints and makes it possible to increase vector cargo capacity. To compensate for loss of cell-to-cell movement, we used *Agrobacterium* to deliver geminivirus constructs to cells. The *Agrobacterium* T-DNA was modified to harbor *cis*-acting sequences needed for replication in a LIR-SIR-LIR orientation (hereafter referred to as the LSL T-DNA; Figure 2-5A). Concomitant delivery of Rep protein should result in replicational release and amplification of circular geminivirus replicons (GVRs; Stenger et al., 1991).

Delivery of conventional T-DNA to plant cells by *Agrobacterium* can result in transient gene expression. By modifying T-DNA to harbor sequences for genome engineering reagents (sequence-specific nucleases and repair templates), both targeted mutagenesis and gene targeting can be carried out (Li et al., 2013; Reiss et al., 2000). To better distinguish between the effects of GVRs and the input T-DNA, we sought to design the LSL T-DNA such that expression of heterologous sequence is minimized when virus proteins are not present. We chose to position heterologous sequence downstream of the LIR in the virion-sense direction. The LIR functions as a bidirectional promoter and promoter strength in the virion-sense direction is weak in the absence of Rep and RepA proteins (Hefferon et al., 2006). To promote high gene expression within circularized GVRs, we positioned a duplicated 35S promoter (P35S) upstream of the LIR in the virion-sense direction. Splice donor and acceptor sequences were positioned, respectively, upstream and downstream of the LIR sequence that is predicted to form after circularization. High-protein expression may also occur from the transactivation of the virion-sense LIR promoter by the viral Rep proteins (Hefferon et al., 2006).

To investigate whether our GVRs replicate and express protein in plant cells, we modified LSL T-DNA plasmid to encode green fluorescent protein (GFP; pLSLGFP) or β-glucuronidase (GUS; pLSLGUS; Figure 2-5B). GVRs were delivered to tobacco leaf tissue by infiltrating a mixture of *Agrobacterium* strains containing pLSLGFP or pLSLGUS and a separate T-DNA expression plasmid encoding Rep and RepA (pREP). Both GUS and GFP expression were enhanced when pREP T-DNA was delivered (Figure 2-5B and 2-6). Background GUS activity seen in the tissue transformed with pLSLGUS is most likely due to activity of the virion-sense LIR promoter within *Agrobacterium* and plant cells.

To correlate the enhanced protein expression with GVR production, replicational release was evaluated by PCR using primers that would only amplify a product if circularization occurred. Strong amplification of the LIR sequence was observed only from samples in which

pREP T-DNA was delivered, confirming successful replicational release of replicons (Figure 2-5C and 2-7). In a separate experiment, we quantified the relative GVR gene copy number over a two-week time course using quantitative polymerase chain reaction (qPCR; Figure 2-8). We observed a maximum copy number of approximately 6,000 per single-copy gene from samples collected five days post infiltration. This is consistent with previous reports describing that mastreviruses replicate to very high copy number within plant cells (Timmermans et al., 1992). Together, these data illustrate that our GVRs replicate and express heterologous proteins in leaf cells. Furthermore, these results show that peak replicon and protein levels occur between three and five days post infiltration, which is consistent with previous reports using BeYDV-based replicons for protein expression (Zhang and Mason, 2006). In subsequent genome engineering experiments, we chose to sample tissues between five and seven days post infiltration. We reasoned that sampling on these days would allow sufficient time for the GVRs to reach maximum copy number and for peak protein expression. Furthermore, as DNA modifications are stable events, a targeted change occurring at three days post infiltration, for example, should be detectable seven days post infiltration.

Expression of Zinc-Finger Nucleases with Geminivirus Replicons

To demonstrate targeted mutagenesis using GVRs, an LSL T-DNA plasmid was modified to encode the Zif268:*Fok*I ZFN (pLSLZ; Figure 2-9A). Target sequence for this ZFN is present within a reporter gene that is stably integrated in *N. tabacum* plants (Wright et al., 2005). This reporter is a non-functional, translational fusion between GUS and neomycin phosphotransferase (NPTII; Datla et al., 1991). Leaf tissue from transgenic tobacco plants carrying the *gus:nptII* reporter was syringe infiltrated with an *Agrobacterium* strain containing pLSLZ (one side of a leaf) or coinfiltrated with a mixture of *Agrobacterium* strains containing pLSLZ and pREP (the other side of the leaf). Total DNA was extracted seven days post infiltration, replicational release was verified by PCR (Figure 2-9A), and the Zif268 target sequence was analyzed for mutations introduced by NHEJ. To this end, a 484 bp DNA fragment encompassing the Zif268 target sequence was amplified by PCR. Amplicons were digested with *Mse*I (an *Mse*I site is present in the spacer sequence between the two Zif268 binding sites), and bands were resolved by agarose gel electrophoresis (Figure 2-9B). Cleavage-resistant amplicons from leaf tissue transformed with both pLSLZ and pREP T-DNA were cloned and sequenced (Figure 2-9C). Six out of eight sequenced clones contained mutations at the ZFN target site. Five out of the six sequences had distinct NHEJ-induced mutations indicating that DNA in multiple leaf cells had been modified. We observed significantly higher levels of targeted mutagenesis in cells transformed with pLSLZ and pREP compared to pLSLZ alone (Figure 2-9D). Cleavage-resistant amplicons observed from tissue delivered pLSLZ T-DNA are most likely due to expression of Zif268:*Fok*I from the virion-

sense LIR promoter. These results indicate that GVRs can express ZFNs for creating targeted DNA double-strand breaks.

Expression of TALENs and components of the CRISPR/Cas system with Geminivirus Replicons

Having demonstrated that GVRs can deliver ZFNs to plant cells, we next tested if they were effective vehicles for delivering TALENs and CRISPR/Cas reagents. We constructed a pLSL vector that encodes the T30 TALEN pair (pLSLT; Figure 2-10A; Zhang et al., 2013), which targets a site within the tobacco acetolactate synthase (*ALS*) genes. A second vector was constructed that encodes Cas9 and synthetic guide RNA (sgRNA), also targeting *ALS* (pLSLC; Figure 2-10B). After delivery to plant cells, mutagenesis of endogenous targets was monitored by loss of a restriction enzyme site due to NHEJ-induced mutation at the site of cleavage. We observed an enhancement in mutagenesis when pREP was co-transformed, suggesting GVR-mediated expression of TALENs and CRISPR/Cas elements can be achieved (Figure 2-10A, C, E, F).

Geminivirus Replicons Enhance Frequencies of Gene Targeting

GVRs were next assessed for their ability to achieve gene targeting through the delivery of sequence-specific nucleases and repair templates. The target for modification was the defective *gus:nptII* gene, which can be repaired by correcting a 600 bp deletion that removes part of the coding sequences for both GUS and NPTII (Wright et al., 2005). An LSL T-DNA plasmid was constructed that encodes the Zif268*:Fok*I ZFN followed immediately by repair template sequence (designated *us:NPTII*) designed to restore *gus:nptII* function (pLSLZ.D; Figure 2-11A). Cells having undergone gene targeting will express a functional GUS protein, and, consequently, will stain blue when incubated in a solution with the GUS substrate X-Gluc. Random integration of the repair template or read-through transcription from virus promoters should not produce functional GUS protein due to 500 bp that is missing from the 5' end of the coding sequence. This was experimentally confirmed by transforming non-transgenic leaf tissue with pLSLZ.D and pREP T-DNA; no GUS activity was observed five days post infiltration (Figure 2-11B, Neg. Ctrl.). To compare the performance of GVRs with conventional T-DNA, we engineered a T-DNA vector with Zif268*:Fok*I sequence followed immediately by the *us:NPTII* repair template (p35SZ.D). Five days following infiltration, leaf tissue was stained in X-Gluc and assessed for cells having undergone gene targeting. We observed an enhancement in GUS activity in leaf tissue transformed with pLSLZ.D and pREP T-DNA (Figure 2-11B) relative to the p35SZ.D and pLSLZ.D controls. To verify repair of the reporter gene, PCR was performed on total DNA extracted from leaf tissue transformed with pLSLZ.D and pREP T-DNA. Primers were used that recognize sequence within the repaired reporter gene and approximately 1 kb downstream of the homology carried by the repair template (Figure 2-12). Amplification of a 2.3 kb product from 9 of the 11 leaf-samples

indicates the presence of the repaired *GUS:NPTII* transgene. Sequence analysis confirmed the presence of an *NPTII* repair template-transgene junction, consistent with repair of the reporter by gene targeting.

The relative enhancement in gene targeting was calculated by quantifying the density of blue sectors within transformed leaf tissue. A significant enhancement in gene targeting was observed in tissue transformed with pLSLZ.D and pREP T-DNA across multiple transgenic plant lines where the reporter gene is integrated at different chromosomal positions (Figure 2-11C; Wright et al., 2005). Notably, in plant line 1 we observed an enhancement in gene targeting greater than two orders of magnitude. Taken together, these results demonstrate that GVRs can deliver sequence-specific nucleases and repair templates for gene targeting, and that GVRs contribute features that result in the enhancement of gene targeting in somatic cells relative to conventional T-DNA delivery.

Double-Strand Breaks Enhance Gene Targeting

There are several features of GVRs that may promote gene targeting, including replication to high copy numbers, high protein expression and pleiotropic activities of Rep and RepA. To test these features, we transformed single leaves with two *Agrobacterium* samples and directly compared GUS activity resulting from gene targeting (Figure 2-13A). Pairing two samples on a single leaf reduced the variation in results due to environmental conditions (e.g., differences in leaf age, size and health). As T-DNA size affects transfer efficiency, we ensured that vectors being directly compared had similar T-DNA sizes. To determine the contribution of double-strand breaks to gene targeting, and to test the effectiveness of this assay, *Zif268:FokI* (897 bp) was replaced with *GFP* (912 bp). We observed a significant decrease (100 fold) in GUS activity when *Zif268:FokI* was removed (Figure 2-13B), which is consistent with previous reports describing a stimulatory effect of double-strand breaks on recombination (Puchta et al., 1996).

Replication of Repair Templates Enhances Gene Targeting

To determine if replication of GVRs contributes to gene targeting, we compared our GVR system (pLSLZ.D and pREP) to a replication-inactive system (p35SZ.D and pREP). We observed a significant decrease in blue sectors when the *cis*-acting replication elements were removed (Figure 2-13C, left graph). These results indicate that replication enhances gene targeting; however, we could not conclude whether replicating sequence-specific nucleases or repair templates contributed more towards gene targeting. Therefore, to address this issue, we explored the role of sequence-specific nuclease expression on GVR-mediated gene targeting.

One explanation for the observed high frequency of gene targeting is that GVRs mediate high expression of sequence-specific nucleases, and, as a result, double-strand breaks are more frequent. To test this hypothesis, we compared the level of NHEJ-induced mutations in leaf tissue

delivered Zif268:*Fok*I using GVRs (pLSLZ and pREP) or conventional T-DNA (p35SZ). We predicted that if expression of sequence-specific nuclease was enhanced using GVRs, we would observe higher levels of NHEJ-induced mutations. Our results refute this hypothesis (Figure 2-14A): we observed a lower level of NHEJ-induced mutations with GVRs, suggesting that replication of sequence-specific nucleases does not augment levels of gene targeting. As an alternative approach to assess the impact of nuclease amplification, we modified Rep and RepA to carry a mutation in the conserved ATPase domain (K219H for BeYDV). This mutation impairs replication (Desbiez et al., 1995; Figure 2-14B); however, other functions of Rep and RepA should remain active (e.g., nicking activity, ligation activity and transactivation of virion-sense transcription). As expected, when pREPK219H T-DNA was delivered to leaf tissue, LSL T-DNA was not replicated (Figure 2-14C); and we observed similar levels of targeted mutagenesis with pREPK219H as with pREP (approximately 10%; Figure 2-14D and E). Collectively, the results suggest that replication of Zif268*:Fok*I coding sequence does not significantly contribute to enhancing targeted mutagenesis.

To test the hypothesis that gene targeting is enhanced by replication of repair templates and not sequence-specific nucleases, we compared gene targeting frequencies of two GVR systems that either replicate sequence-specific nucleases or repair templates (Figure 2-13C, right graph). Here, we observed significantly higher levels of gene targeting when repair templates were replicated. Similarly, when we translocated repair templates from conventional T-DNA to GVRs, we observed substantially higher levels of gene targeting (Figure 2-15). These results indicate that although expression of sequence-specific nucleases is critical for gene targeting, replication of repair template sequence contributes more towards enhancing gene targeting.

Pleiotropic Activity of Rep and RepA Enhances Gene Targeting

Geminivirus Rep and RepA proteins interact with a multitude of host proteins. Specifically, RepA protein from mastreviruses interacts with Geminivirus RepA-binding protein 1 (GRAB1) and GRAB 2 (Xie et al., 1999) and the retinoblastoma-related protein (RBR; Horváth et al., 1998). It may be possible that expression of mastrevirus Rep and RepA in plant cells affects levels of homologous recombination. To test this hypothesis, we compared gene targeting levels in tissue delivered p35SZ.D T-DNA or p35SZ.D and pREP T-DNA. Here, we observed a significant increase in blue sectors when pREP was delivered (Figure 2-13D). To confirm that delivery of pREP T-DNA to plant cells results in the expression of both Rep and RepA, we used reverse transcription PCR to detect Rep and RepA transcripts. To avoid detecting RepA transcripts produced by *Agrobacterium*, we transfected protoplasts with the pREP T-DNA plasmid. Total RNA was extracted 24 hours post transfection, and we detected unspliced (RepA) and spliced (Rep) transcripts (Figure 2-16). These results demonstrate that pleiotropic activity of Rep and/or RepA promotes gene targeting. Taken together with the results in the previous two

sections, high-frequency gene targeting using GVRs is due to synergy between targeted double-strand breaks, replication of repair template sequence and pleiotropic activity of Rep and RepA.

Recovery of Calli and Plantlets with a Precise DNA Change

To demonstrate the feasibility of GVRs for generating fully developed plants with a desired sequence change, we regenerated cells harboring the corrected *GUS:NPTII* transgene. We chose to regenerate plants from leaf cells because our previous data demonstrate the usefulness of GVRs in whole leaves. Leaves were syringe infiltrated with two strains of *Agrobacterium* containing pLSLZ.D and pREP. Five days later, infiltrated leaves were removed from the plant, surface sterilized and placed on regeneration media with kanamycin (Figure 2-17). To track the progression of regenerating cells with the *GUS:NPTII* transgene, subsets of leaf discs were stained in X-Gluc solution 0, 7, 14, 21, 42 and 49 days after plating. We observed GUS expression from leaf discs stained between 0 and 21 days (Figure 2-17A, B, C, and D). After day 42, we observed shoots with GUS expression (Figure 2-17E and F). To confirm that tissue staining blue harbors the corrected *GUS:NPTII* transgene, we extracted genomic DNA from blue calli and performed PCR with primers designed to amplify sequence from the repaired *GUS:NPTII* transgene. Because our negative control samples (non-infiltrated tissue) did not produce calli or shoots, we chose to analyze genomic DNA from callus tissue that did not stain blue in the pLSLZ.D and pREP samples. We observed amplification of sequence from all four blue-stained tissues, and not from tissue that did not stain blue (Figure 2-17G). Sequencing of all four PCR products confirmed the presence of an *NPTII* repair template-transgene junction, consistent with repair of the reporter by gene targeting (Figure 2-17H). These findings demonstrate that GVRs can be used to introduce exact DNA changes in plant cells, and that these cells maintain the ability to regenerate.

Development of a Single-Component Geminivirus-Replicon System

Having demonstrated the usefulness of GVRs for genome engineering, and that there are virus-specific features that promote gene targeting, we sought to improve the design of the geminivirus vectors for efficient genome editing. To this end, we developed a single-component system that requires the delivery of only one T-DNA molecule to generate GVRs. Transfer of a single T-DNA to cells is more efficient than the transfer of two T-DNAs, which may result in a higher frequency of cells receiving a functional GVR. We modified the GVR sequence such that Rep and RepA coding sequences were placed downstream of the complementary-sense LIR promoter – the natural position of Rep and RepA coding sequences within mastrevirus genomes. Furthermore, we positioned the strong constitutive promoter (P35S) directly upstream of our reporter gene or sequence-specific nuclease of interest.

To test the effectiveness of this vector system for protein expression, we placed GFP coding sequence downstream of P35S (pLSLGFP.R; Figure 2-18A). We compared protein

expression of pLSLGFP.R to the dual-component GVR system (pLSLGFP and pREP). In general, we observed more uniform GFP expression with pLSLGFP.R (Figure 2-18B). This is most likely due to a higher number of cells receiving a functional GVR, and also from transient expression of GFP from the linear LSL T-DNA. To verify that replicational release had occurred, total genomic DNA was extracted from leaf tissue expressing GFP and PCR was performed using primers that were homologous to sequences within GFP and the Rep/RepA coding sequences. Strong amplification of LIR sequence indicated the presence of circular GVRs (Figure 2-7B).

To assess gene targeting with our single-component vector system, we placed *Zif268:FokI* and *us:NPTII* repair template sequence downstream of the 2x35S promoter and within the LIR borders (pLSLZ.D.R; Figure 2-19A; Figure 2-20). One side of a leaf was syringe infiltrated with *Agrobacterium* containing pLSLZ.D.R and the other with was infiltrated with a mixture of *Agrobacterium* containing pLSLZ.D and pREP. As an additional control, a separate leaf was infiltrated with p35SZ.D. Five days post infiltration, leaf tissue was stained in a solution containing X-Gluc (Figure 2-19B) and the density of blue spots was quantified (Figure 2-19C). We observed comparable levels of GUS activity using our single-component vector relative to the two-component system. In general, we noticed that tissue delivered pLSLZ.D.R had more uniform GUS activity. This observation may reflect a difference in the transfer efficiency of a single T-DNA molecule to that of two T-DNA molecules.

Having previously demonstrated that gene targeting is stimulated by replication of repair templates, and not sequence-specific nucleases, we predicted that positioning Zif268*:Fok*I coding sequence outside of the LIR borders (but still within the same T-DNA; pZLSLD.R; Figure 2-19A) would result in comparable gene targeting frequencies to pLSLZ.D and pREP. Our results support this hypothesis. Gene targeting frequencies of pZLSLD.R were similar to that of the two component system, pLSLZ.D and pREP (Figure 2-19C). Together, these results illustrate the usefulness of single-component vectors for the delivery of GVRs for recombination-based genome editing.

DISCUSSION

The potential of geminiviruses for gene targeting has long been recognized, principally because their DNA genomes could serve as repair templates, and they replicate to high copy number. We found the key to realizing this potential was the use of sequence-specific nucleases to stimulate recombination between the virus and the chromosomal target. In addition, replication of repair templates and pleiotropic activities of Rep and RepA further enhanced gene targeting; however, these factors were insufficient to achieve gene targeting in the absence of a targeted chromosome break. Together, these data stress the importance of using geminivirus technology in conjunction with sequence-specific nucleases.

Genome size constraints hindered the use of full viruses for genome engineering. One strategy to increase the cargo capacity of begomoviruses is to replace the coat protein gene with a desired heterologous sequence of up to 800 bp. This strategy permits the delivery of some genome engineering tools, including repair templates, meganucleases, ZFN monomers and possibly sgRNA for the CRISPR/Cas system. However, the cargo capacity is insufficient for expression of large heterologous sequences, including TALENs and Cas9. For large cargo, there is a high likelihood that larger-than-wild type genomes will recombine to smaller sizes. Virus genome size constraints are imposed by cell-to-cell movement through plasmodesmata and not by replication (Gilbertson et al., 2003). Therefore, to deliver large heterologous sequences, we deleted virus genes involved in cell-to-cell movement and retained the elements required for rolling-circle replication. This effectively eliminated constraints on cargo capacity and made it possible to simultaneously deliver sequence-specific nucleases and repair templates for gene targeting. In addition, this approach allowed us to maintain the ability for replicons to amplify to high-copy number by rolling-circle replication.

Compared to meganucleases and ZFNs, TALENs and CRISPR/Cas reagents are, in general, easier to construct and have increased efficacy. Furthermore, TALENs appear to have low frequencies of off-target cleavage (Mussolino et al., 2011), and CRISPR/Cas technology enables efficient multiplexed genome engineering (Cong et al., 2013; Wang et al., 2013). Therefore, it would be advantageous for new vectors for genome engineering to be compatible with TALEN and CRISPR/Cas technology. Our results indicate that GVRs accomplish this. By placing TALENs and Cas9/sgRNAs within GVRs, we achieved targeted mutagenesis of the endogenous *ALS* genes. For gene targeting, when the repair template is on the GVR, sequence-specific nucleases can be expressed using either GVRs or conventional T-DNA, both of which resulted in comparable levels of gene targeting. Taken together, these results suggest that not only are GVRs amenable for expression of TALENs and the CRISPR/Cas system, but they can be combined with other methods for delivery of sequence-specific nucleases such as T-DNA.

We observed that gene targeting was enhanced by pleiotropic activity of Rep and RepA. One hypothesis to explain these results derives from our understanding of how geminiviruses establish successful infection within host plants. Geminiviruses only infect non-dividing cells and encode proteins that promote entry into S-phase (Horns and Jeske, 1991; Lucy et al., 1996; Ascencio-Ibáñez et al., 2008; Kong et al., 2000). Progression into S-phase is thought to provide geminiviruses with necessary host factors required for replication (e.g., nucleotides, DNA polymerase). In monocots, S-phase entry is mediated by the mastrevirus RepA protein: RepA interacts with the plant's retinoblastoma-related protein (RBR) and sequesters RBR from E2F, thereby promoting cell-cycle progression into S-phase. The BeYDV RepA has been shown to interact with RBR from maize, and our results suggest that RepA stimulates tobacco cells to enter S-phase (Liu et al., 1999). The basis for this latter conclusion is that in higher-eukaryotic cells, the

frequency of homologous recombination increases as cells move out of a resting state and into S and G2 (Rothkamm et al., 2003). We believe promotion into S-phase by Rep/RepA underlies the stimulatory effect these proteins have in gene targeting.

Geminiviruses infect a wide range of plants, including both monocots and dicots. We modified the bean yellow dwarf virus, which belongs to the genus *Mastrevirus*. Most mastreviruses infect monocots, however, the bean yellow dwarf virus infects dicots, including tobacco species, *Datura stramonium,* tomato (*Lycopersicon esculentum*), French bean (*Phaseolus vulgaris*), and *Arabidopsis* (Liu et al., 1997). Whereas bean yellow dwarf virus replicons should function within this list of species, they are unlikely to replicate in plants that are not hosts to the wild type virus, including monocots. It may be possible to develop geminivirus replicon systems for monocots using other mastreviruses (e.g., wheat dwarf virus and maize streak virus). Both wheat dwarf virus and maize streak virus have long been used for protein expression in monocots and replicate to similar copy numbers as our bean yellow dwarf virus replicon system (Timmermans et al., 1992; Palmer et al., 1999). Thus, there is promise for GVRs to enable efficient genome engineering across many plant species.

METHODS

Vector construction

LSL T-DNA plasmids (pLSL, pLSLGUS, pLSLGFP, pLSLZ, pLSLT, pLSLC and pLSLZ.D) were constructed using the *Agrobacterium* binary vector pCAMBIA1300. LIR and SIR sequences were derived from the mild BeYDV strain (Halley-Stott et al., 2007) (GenBank Accession number DQ458791). The boundaries of the LIR and SIR sequences were defined following previously described methods (Mor et al., 2003). The cauliflower mosaic virus promoter (P35S) was positioned within the replicon and upstream of the LIR. To enable translation of coding sequence that is located downstream of the LIR, splice donor and acceptor sequences were positioned, respectively, upstream and downstream of the LIR sequence that is formed upon circularization. To enable facile cloning, pLSL was modified to harbor Gateway-compatible attR1 and attR2 sequences between the upstream LIR and SIR elements. Conventional T-DNA expression plasmids (pREP, pREPK219H, p35SZ and p35SZ.D) were constructed using the Gateway-compatible *Agrobacterium* binary vector pMDC32. Coding sequence for Rep and RepA was derived from the mild BeYDV strain (Halley-Stott et al., 2007).

Two entry vectors were constructed for cloning sequence-specific nuclease (or reporter protein) and repair template sequences into the pLSL and pMDC32 destination plasmids. The first entry vector (pNJB091) is a derivative of pZHY013 (Zhang et al., 2013) with attL2 sequence replaced with Nos-T and attR5 sequence. The second entry vector (pNJB080) is also a derivative of

pZHY013, but here the *Fok*I and attL1 sequences were replaced with the chloramphenicol resistance gene, ccdB gene and attL5 sequence.

To engineer pLSL vectors for the repair of the *gus:nptII* transgene, pNJB80 was modified to harbor repair template sequence (*us:NPTII*), and pNJB91 was modified to harbor the Zif268:*Fok*I ZFN coding sequence. To generate the *us:NPTII* sequence for cloning into pNJB80, pDW1269 (Wright et al., 2005) was used as a template in a PCR with the oligonucleotides 5'-CGCGGA**GCGGCCGC**AGCAGTCTTACTTCCATGATTTC and 5'-GATCGATCT**CTGCAG**CCCCGCGCGTTGGCCGATTCA. Bold letters indicate *Not*I and *Pst*I recognition sites for cloning into pNJB80. The resulting repair template contained 600 bp of corrective sequence flanked by 1 kb of homologous sequence to the *gus:nptII* reporter. To generate Zif268:*Fok*I sequence for cloning into pNJB91, pDW1345 (Wright et al., 2005) was used as a template in a PCR with the oligonucleotides 5'-GCCCTTCA**CCATGG**CTTCCTCCCCTCCAAAGAAAAAG and 5'-GAACGATCG**GACGTC**CTATTAAAAGTTTATCTCACCGTTA. Bold letters indicate the *Nco*I and *Aat*II restriction enzyme sites used for cloning into pNJB91. The resulting plasmids were used in a MultiSite Gateway (Life Technologies) recombination reaction with pLSL or pMDC32 to generate pLSLZ.D or p35SZ.D, respectively. pLSLZ and p35SZ were generated as described above; however, instead of using pNJB80 with the *us:nptII* sequence, we used a different pNJB80 vector which harbors approximately 3.0 kb of 'filler' DNA sequence.

To generate pLSLT, pNJB91 was modified to harbor the T30 TALEN pair sequence (Zhang et al., 2013). This sequence contains the two T30 TALEN ORF linked by a T2A ribosomal skipping sequence. The N- and C- terminal truncation for both TALENs was NΔ152/C63 (Miller et al., 2011). The T30 TALEN sequence was cloned from pZHY528 into pNJB91 using standard cloning techniques. The resulting plasmid was used in a Multi-Site Gateway recombination reaction with pLSL and pNJB80 containing the *us:NPTII* repair template to generate pLSLT.

To generate pLSLC, pNJB91 was modified to harbor a codon-optimized Cas9 gene from *Streptococcus pyogenes*, and pNJB80 was modified to harbor the *Arabidopsis thaliana* U6 RNA Pol III promoter followed by sgRNA sequence. The Cas9 gene was codon-optimized for dicotyledonous plants and synthesized (GenScript) with C- and N-terminal SV40 NLS signals and an N-terminal 3x FLAG-tag sequence. The U6:gRNA sequence was synthesized on gBlocks (Integrated DNA Technologies).

Single-component LSL vectors (pLSLGFP.R, pLSLZ.D.R and pZLSLD.R), where Rep/RepA coding sequence was positioned within the replicon, were constructed in T-DNA plasmids with the pCAMBIA backbone. Sequence of features within pLSLZ.D.R can be found in Supplemental Figure 14 online. Boundaries for the LIR and SIR were determined using the annotated sequence for the mild BeYDV strain (GenBank Accession number DQ458791).

Plant material

Tobacco (*Nicotiana tabacum* var. Xanthi) plants, both wild type and transgenic (Wright et al., 2005), were grown at 21^{o}C with 60% humidity under a 16 hour light and 8 hour dark cycle. Leaves (fully expanded upper leaves) from 4 - 6 week old tobacco plants were used for *Agrobacterium* transformation. When comparing gene targeting or targeted mutagenesis frequencies between different T-DNA molecules, one leaf per plant was infiltrated.

***Agrobacterium* transformation**

T-DNA vectors were introduced into *Agrobacterium* (*Agrobacterium tumefaciens* GV3101/pMP90) using the freeze-thaw method. Single transformed colonies were grown overnight at 28^{o}C in 3 mL of LB starter culture containing kanamycin (50 μg/ml) and gentamicin (50 μg/ml). The next day, 1 ml of starter culture was used to inoculate 50 mL of LB culture containing kanamycin (50 μg/ml), gentamicin (50 μg/ml), 10 mM 2-(N-morpholino) ethanesulfonic acid (MES), and 20 μM acetosyringone. After reaching an OD_{600} of approximately 1 (around 16 hours), cells were pelleted and resuspended to an OD_{600} of 0.2 using infiltration buffer (10 mM MES, 150 μM acetosyringone and 10 mM $MgCl_2$, pH 5.6). Resuspended cultures were incubated at RT for 2 – 4 hours. For experiments requiring the co-infiltration of two different *Agrobacterium* strains, cultures were mixed together in a 1:1 ratio, or, for three different strains, in a 1:1:1 ratio. When comparing two different *Agrobacterium* samples, the OD's of the solutions were adjusted accordingly to ensure equal concentrations of individual *Agrobacterium* strains. Leaves from tobacco plants were infiltrated with *Agrobacterium* using a 1 mL syringe. Immediately following infiltration, plants were watered and covered with a plastic dome to maintain high humidity. Plastic domes were removed approximately 24 hours post infiltration.

GUS Assay

To visualize cells expressing GUS, leaf tissue was vacuum infiltrated with X-Gluc solution (10 mM phosphate buffer, 10 mM EDTA, 1 mM ferricyanide, 1 mM ferrocyanide, 0.1% Triton X, 1 mM X-Gluc). Leaf tissue was then incubated in X-Gluc solution at 37^{o}C for approximately 24 hours. Chlorophyll was removed by incubating stained leaf tissue in 80% ethanol for approximately 1 week.

Tobacco Regeneration

Leaves from 4 – 5 week-old tobacco plants were infiltrated with pLSLZ.D and pREP. As a control for selection and regeneration of kanamycin resistant cells, additional leaves were infiltrated with pCAMBIA2301 (T-DNA that contains a kanamycin resistance and also the GUS gene). Two separate experiments were performed; for each experiment three to five leaves from different plants were infiltrated with each T-DNA or T-DNA pair. Five days post infiltration, leafs were

removed from the plant and surface sterilized by submersion in 10% bleach for approximately 20 minutes. Leaves were then washed three times in sterile water. Sterilized leaf tissue was cut into approximately 1 cm by 1 cm discs using a surgical blade. Leaf discs were plated onto regeneration media (4.4 g MS with vitamins, 1 mg/L 6-benzylaminopurine, 0.1 mg/L 1-naphthaleneacetic acid, 30 g sucrose, 8 g agar pH 5.7) containing kanamycin (100 μg/ml) and Timentin (200 μg/ml). As a negative control for tobacco selection, non-transformed leaf tissue was sterilized and plated. Discs were transferred to new plates weekly.

PCR-based detection of circularized GVRs and the *GUS:NPTII* transgene

To detect the presence of circularized GVRs, total DNA was extracted from plant tissue and used as a template for PCR. Oligonucleotides were designed to be homologous to DNA sequence located upstream and downstream of LIR sequence in the circular replicon (5'-GTTTCACTTCACACATTATTACTG and 5'-TGTTGAGAACTCTCGACGTCCTGC). To minimize strand transfers during amplification, PCR was performed using the Expand Long Template PCR System (Roche).

To detect the presence of a repaired *GUS:NPTII* transgene, total DNA was extracted from leaf or callus tissue and used as a template for PCR. Oligonucleotides were designed to be homologous to DNA sequence located within the corrective sequence on the repair template and sequence approximately 1 kb downstream of the region of homology harbored on the repair template (5'-GTCGGTGAACAGGTATGGAAT and 5'- CTACATACCTCGCTCTGCTAATC).

Detection of ZFN-, TALEN- and Cas9-induced mutations

Experiments that used GVRs to mediated expression of ZFNs, TALENs and CRISPR/Cas elements for targeted mutagenesis were performed similarly: one side of a leaf was delivered pLSL and the other side was infiltrated with pLSL and pREP. Five to seven days post infiltration, total DNA was extracted and assessed for mutations at the appropriate sequence-specific nuclease target site. To detect Zif268:*Fok*I-induced mutations, the *gus:nptII* target locus was amplified using primers that are homologous to sequences 270 bp upstream and 214 bp downstream of the *Mse*I site (present within the spacer region of the Zif268 binding sequences; 5'-AAGGTGCACGGGAATATTTCGCGC and 5'-GCCATGATGGATACTTTCTCG). The resulting amplicons were digested for 16 hours at 37^{o}C with *Mse*I and resolved by gel electrophoresis. Cleavage-resistant amplicons were gel purified, cloned into pJET1.2 and sequenced. To detect TALEN-induced mutations at the *SurA* and *SurB* loci, total DNA was first pre-digested with *Alu*I for 16 hours at 37^{o}C. Digested DNA was used as a template for a PCR with primers designed to amplify a 573 bp product encompassing the T30 binding site (5'-GACGGTGCAGAAAGTGAAGTA and 5'- TATGGCCCAGGAGTGTCTAA). The resulting amplicons were digested for 16 hours at 37^{o}C with *Alu*I and resolved by gel electrophoresis. Cleavage-resistant amplicons were gel purified, cloned into pJET1.2 and sequenced. To detect

Cas9-induced mutations, the target DNA sequence present within *SurA* and *SurB* was amplified by PCR (F 5'-TGACATCTGGTGGATTAGGAGC, R 5'-CAATCACATCCAACAAGTATGGC). The resulting 408 bp amplicons were digested with *Alw*I for 16 hours at 37□C. Cleavage resistant products were gel purified and cloned into pJET1.2 and sequenced.

Data collection and statistics

Following staining in X-Gluc, the density of blue sectors was quantified in leaf tissue. Three non-overlapping images were taken for each infiltrated region (leaf halves) using a camera attached to a stereoscope (Nikon). Each image captured 1.8 cm^2 of leaf tissue, totaling 5.4 cm^2 for each leaf half. Image locations were directed to capture the entire number of blue sectors for each leaf half. For samples where the total number of blue sectors exceeded 5.4 cm^2 of area, regions with the highest density of blue sectors were imaged. Leaves without detectable GUS activity were excluded from the data analysis. ImageJ software was used to calculate the number of blue sectors per image. P values in Figures 4, 5 and 6 were generated using a paired, two-tailed Student's t-test. P values in Figures 2 and 3 were generated using a two-tailed Student's t-test.

Generating Cabbage Leaf Curl Virus Vectors for Gene Targeting

Repair templates with homology to the *ADH1* locus were cloned into pCPCbLCVA.007. pCPCbLCVA.007 is a plasmid containing a partial tandem direct repeat of the cabbage leaf curl virus (CaLCuV) genome (Muangsan and Robertson, 2004). The template for amplifying homology arms was genomic DNA from *Arabidopsis* (*Arabidopsis thaliana* ecotype Columbia). Repair templates were designed to introduce a unique 18 bp modification sequence (5'-GAGCTCAGTACTGCATGC) within the *ADH1*-ZFN binding site.

Biolistic Transformation of Cabbage Leaf Curl Virus Vectors

To prepare Arabidopsis plants (homozygous for a transgene encoding the ZFN pair targeting *ADH1*; Qi et al. 2013, Zhang et al. 2010) for biolistic bombardment, seeds were dispensed onto the surface of soil in each of the four corners of 2.5 x 2.5 inch pots. Plants were grown at 22℃ for 2 weeks before removing the dome, and then grown for an additional 1-2 weeks with watering when needed. Watering was stopped approximately 7 days before bombardment. Plants were bombarded when they reached the six- to nine-leaf-stage (approximately three to four weeks).

Biolistic bombardment was carried out closely following previously described protocols (Muangsan and Robertson, 2004). Briefly, to prepare microprojectile particles for five bombardments, 5 μg of each plasmid DNA was added to a tube containing 50 μl of 60 mg/mL gold beads and briefly vortexed. 50 μl of 2.5 M $CaCl_2$ was directly added to the samples and immediately pipetted in and out of a tip to break up conglomerates. 20 μl of 0.1 M spermidine was added and the samples were immediately vortexed for 5 min. The samples were centrifuged at

10,000 rpm for 10 seconds and the supernatant was removed. The gold-bead pellet was resuspended in 250 μl of 100% ethanol and then centrifuged at 10,000 rpm for 10 sec. Supernatant was removed and the samples were resuspended in 65 μl of 100% ethanol. The particles were then stored on ice until bombardment. To prepare the assembly for the microprojectile particles, macrocarrier holders and macrocarriers were soaked in 95% ethanol, air-dried, and assembled. Resuspended particles (10 μl) was spotted onto the center of the macrocarrier and allowed to air dry. Biolistic bombardment was carried out using a PDS-1000 machine (Bio-Rad) in a laminar flow hood.

Sampling Parameters for Cabbage Leaf Curl Virus Experiments

After biolistic bombardment, plants were moved back to the growth chamber. For gene targeting experiments, plants displaying symptoms of virus infection at fourteen days post bombardment were sprayed with a solution containing estradiol (0.01% Silwet L-77, 20 μM β-estradiol). Plants were sprayed daily for one week. For experiments analyzing the stability of the repair templates, tissue was collected approximately 10 days post bombardment. For experiments analyzing the *ADH1* gene for targeted insertions, tissue was collected approximately seven days after beginning estradiol treatment. For both experiments, newly developed tissue was chosen for subsequent analysis; one rosette leaf and one cauline leaf were combined and total DNA was extracted from both tissues. In the cases where no cauline leaves were present, one rosette leaf was sampled.

Time points for DNA isolation, estradiol spraying and our leaf collection strategy were chosen based on data from previously reported studies analyzing CaLCuV infection in Arabidopsis (Ascencio-Ibáñez et al., 2008). It was demonstrated that Arabidopsis plants infected with CaLCuV have a strong presence of virus DNA in upper rosette leaves 12 days post infection. We chose to induce ZFN-expression in our plants at 14 days post bombardment in light of these observations, and also because symptoms of virus infection were more clear at 14 days post bombardment. Whereas our inoculation methods were very efficient (near 100% of bombarded plants were successfully infected as determined by symptoms that developed 10 – 14 days after bombardment) this latter condition allowed us to remove plants that were phenotypically normal (most likely not infected with CaLCuV). For each plant, we chose to sample one upper rosette leaf and one cauline leaf because CaLCuV was shown to accumulate within these tissues (Ascencio-Ibáñez et al., 2008). To induce ZFN expression, plants were exposed to estradiol for seven days. Seven days was chosen as it allows ample time for high gene expression to occur using the XVE promoter system (e.g., in two week old Arabidopsis plants, gene expression reaches a maximum approximately 24 hours post induction and high expression continues 96 hours post induction; Zuo et al., 2000).

In the experiments assessing the stability of 400 bp repair templates, a total of 32 different plants were bombarded. Using primers P1 and P2 (Supplemental Figure 1), we detected a 400 bp sequence in 30 of the 32 bombarded plants. In two samples we did not detect sequence. In the experiments assessing 600 bp repair templates, a total of 24 different plants were bombarded. We detected a 600 bp sequence in 23 of the 24 plants. In one sample we did not detect sequence. In experiments assessing 800 bp repair templates, a total of 12 plants were bombarded. We obtained a single product with a size of 800 bp in 5 of the 12 plants. In the other 7 plants, we detected one or multiple products smaller than 800 bp. In the experiments assessing the stability of 1,000 bp repair templates, a total of 12 different plants were bombarded. In all twelve plants, we did not detect repair template sequence.

Time Course of GFP Expression in Tobacco Leaf Discs

Leaves from 4 week-old tobacco plants were infiltrated with Agrobacterium containing pLSLGFP (one side of a leaf) or a mixture of Agrobacterium containing pREP and pLSLGFP (the other side of the leaf). One hour after infiltration, discs were excised from the leaf tissue. Leaf discs were maintained in 6-well plates in approximately 4 mL of water at 26℃, with 12-h photoperiod. Images shown in Supplemental Figure 5 are of a single leaf disc from both treatments (pLSLGFP or pLSLGFP and pREP). Images of the entire leaf disc were taken at the indicated time points using a camera mounted on an AZ100 (Nikon) confocal microscope.

Quantitative PCR to Detect GVRs

Leaves from 4 week-old tobacco plants were infiltrated with Agrobacterium containing pLSLGFP and a mixture of Agrobacterium containing pREP and pLSLGFP. One leaf was infiltrated with one sample. At each of the indicated time points (see Supplemental Figure 4), tissue was excised from at least two different leaves (from different plants) and total genomic DNA was extracted. Infiltrated leaves were sampled only once. Quantitative PCR was performed using the FastStart Universal SYBR Green Mix kit (Roche) on the LightCycler® 480 Instrument (Roche). Primers were designed to detect GFP coding sequence (5'- ATCCTCGGCCACAAGTTGGA and 5'- GTGGCGGGTCTTGAAGTTGG), circularization (5'- GAGATGAGCACTTGGGATAGGTAAGG and 5'- CTGCAAACAATACACAACAAAGACAATG) and Rep coding sequence (5'- TCCGACACCCCAGCCTCTAC and 5'- TTGCTTCCACAATGGGACGA). The *Fbox* gene was used as reference gene to normalize the gene copy number (5'- GGCACTCACAAACGTCTATTTC and 5'- ACCTGGGAGGCATCCTGCTTAT). Three technical replications were performed for each sample. Expression data were standardized using the

Microsoft Excel Qgene template as previously described (Muller et al., 2002). Error bars represent s.e.m. of at least two different biological replicates.

Protoplast Isolation and Transfection

Rep and RepA transcripts were detected by RT-PCR after transfection of protoplasts with pREP. Leaf tissue was digested and protoplasts were isolated followed protocols as previously described (Zhang et al., 2013) with slight modifications. Protoplasts from tobacco (*Nicotiana tabacum*) were isolated from approximately 4-week-old plants. Approximately 10 leaves were cut into 1- to 2-mm strips using a razor. Leaf strips were then incubated in an enzyme solution (1.0% cellulose R10, 0.25% macerozyme R10, 0.45 M mannitol, 20 mM MES, 20 mM KCl, 10 mM $CaCl_2$) for 14 hours at 25˚C with shaking (20 rpm). The enzyme-plant mixture was then passed through a cell strainer (70 μm) moistened with 2 ml of washing buffer (0.45 M mannitol and 10 mM $CaCl_2$, pH 5.8) and cells were collected in a perti dish. The protoplast solution was divided into two parts and each half was transferred to a 15 mL tube containing 8 mL of 0.55 M sucrose solution. The sample was centrifuged at 1,000 g for 5 min. The protoplasts, found floating above the sucrose solution were then transferred to a 50 mL tube containing 10 mL of washing buffer and centrifuged again (100 g for 5 min at room temperature). The supernatant was removed, and the protoplast pellet was resuspended in a solution containing 0.4 M mannitol, 15 mM $MgCl_2$ and 4 mM MES to a final density of 10^6 cells per mL.

Purified protoplasts were transfected with pREP T-DNA plasmid. To this end, 30 μL of plasmid (15 μg for each plasmid) was incubated with 200,000 cells (200 μL). This solution was gently mixed with 230 μL of 40% PEG-Calcium transformation buffer (40% PEG, 0.2 M Mannitol and 100 mM $CaCl_2$). After 20 min incubation at room temperature, 900 μL of washing buffer was added to the reaction. Protoplasts were pelleted by centrifuging at 250 g for 5 minutes. The cells were resuspended in 800 μL of washing buffer and centrifuged again at 150 g for 5 min. The cells were resuspended in 1 mL of washing buffer. The suspended protoplasts were stored in a six well plate for 24 hours at 25˚C in the dark.

RNA isolation and RT-PCR

Transfected cells (600,000 in total for each sample) were pelleted 24 hours post transfection by centrifuging at 13,000 g for 5 minutes. Total RNA was extracted with Trizol (Invitrogen) following manufacturer's protocols. Isolated RNA was treated with DNase I (ambion). RNA (500 ng) was converted to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Rep and RepA transcripts were amplified using primers 5'-AACTCACACCTTTTCTTATTTTCT and 5'-TTTTCCATATTTAGGGTTGACAGT (Rep and RepA), and 5'-AACTCACACCTTTTCTTATTTTCT and 5'-ATTGAGCTTGTTGGTATGAG (RepA).

Accession Numbers

Sequence data from this article can be found in the *Arabidopsis* Genome Initiative or GenBank/EMBL databases under the following accession numbers: *ADH1* AT1G77120.1; mild strain of BeYDV, DQ458791; *SurA,* X07644.1, *SurB*, X07645.1. pCPCbLCVA.007, AY279345; pCPCbLCVB.002, AY279344.

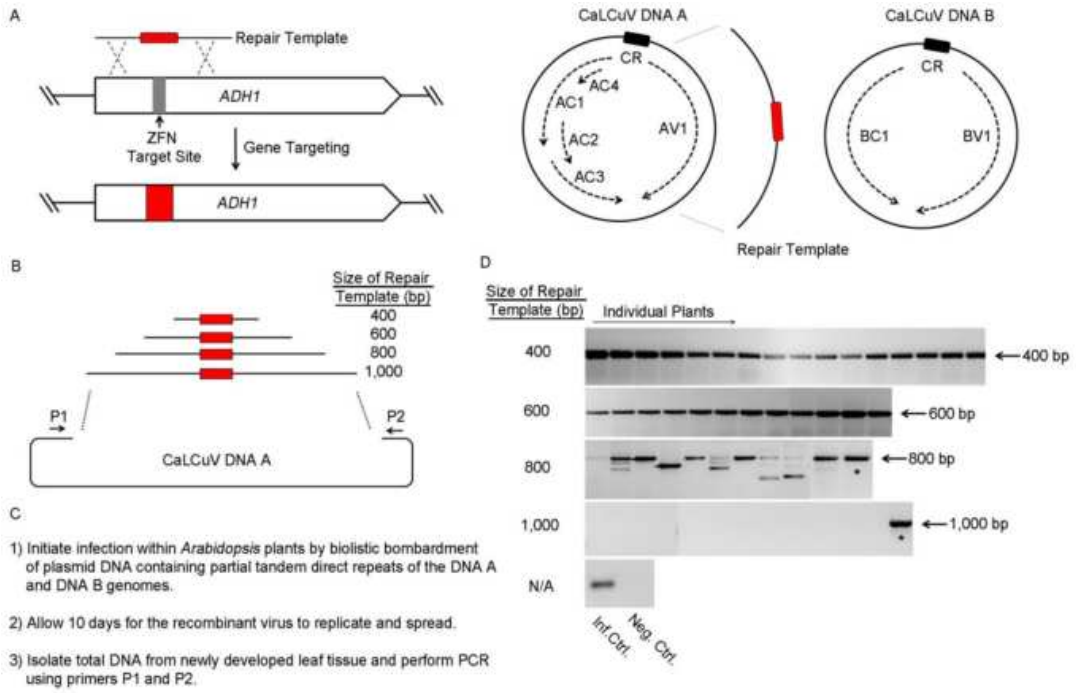

Figure 2-1. Supplemental Figure; Engineering Cabbage Leaf Curl Virus for Delivery of Repair Templates in *Arabidopsis*. (A) The bipartite geminivirus CaLCuV was modified to deliver repair templates to Arabidopsis plants. The target for modification was the alcohol dehydrogenase gene (*ADH1*). Repair templates were designed to introduce a unique 18 bp sequence within the *ADH1* coding sequence (left illustration). The red box represents the 18 bp modification sequence. To generate double-strand breaks within the ADH1 coding sequence, we stably integrated an *ADH1*-targeting ZFN pair into the Arabidopsis genome. The target site for this ZFN pair is illustrated with a gray box. To enable the delivery of repair template sequence, the coat protein gene from CaLCuV DNA A was replaced with repair template sequence (right illustration).CR, common region. (B) To determine the maximum-size repair template that CaLCuV can stably replicate and spread throughout Arabidopsis plants, repair templates ranging from 400 to 1,000 bp in total length were cloned into the DNA A genome. P1 and P2 refer to primers designed to amplify repair templates. (C) Methodology for assessing the stability of repair template sequences carried by CaLCuV. (D) Molecular analysis of repair templates carried by infectious viruses. We observed repair- template instability when the total length was 800 bp or greater. This is indicated by the additional shorter bands relative to our positive control (plasmid DNA). Each lane is a different plant that was bombarded with CaLCuV vectors. Asterisks indicate the PCR-positive controls (plasmid DNA was used in replace of total DNA). The infection control (Inf. Ctrl) is from plants bombarded with control vectors pCPCbLVCA.008 and pCPCbLCVB.002 (Muangsan and Robertson, 2004). This PCR was expected to yield a single product of approximately 400 bp. The negative control (Neg. Ctrl) is from plants bombarded with gold beads not containing plasmid DNA.

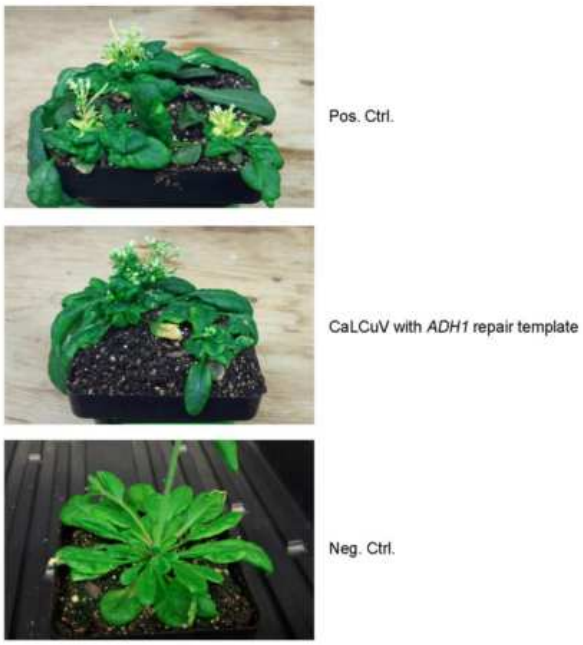

Figure 2-2. Supplemental Figure; Phenotype of *Arabidopsis* Plants Infected with Cabbage Leaf Curl Virus. *Arabidopsis* were bombarded with CaLCuV vectors and infection was monitored. The images are of plants approximately three weeks post bombardment. The positive control (Pos. Ctrl.) is an image of plants bombarded with control vectors pCPCbLCVA.008 and pCPCbLCVB.002. The DNA A vector pCPCbLCVA.008 contains homologous sequence to the *ChlI* as a coat protein gene replacement. As expected, plants infected with this virus displayed yellow or white in newly developed tissue. Additionally, plants displayed symptoms including leaf curling and uneven leaf surface. Plants bombarded with CaLCuV vectors harboring repair template sequences also displayed symptoms of virus infection. This phenotype was used as a marker for successful infection. The negative control (Neg. Ctrl.) image is of a plant bombarded with gold beads.

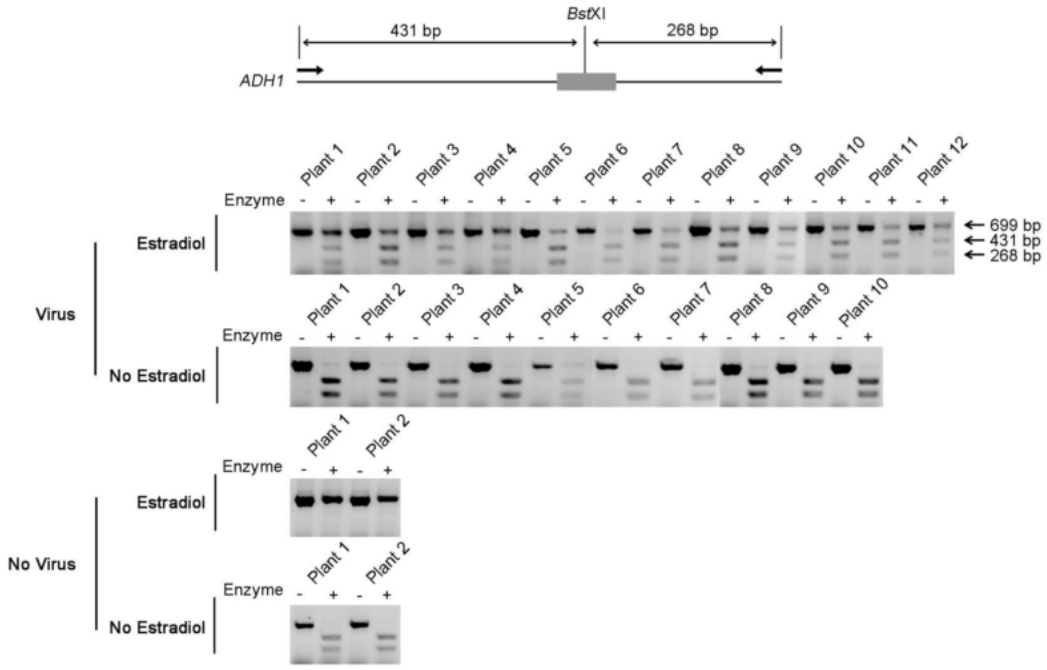

Figure 2-3. Supplemental Figure; Temporal Control of Targeted Double-Strand Breaks in Whole Plants. The estrogen-inducible ZFN pair targeting *ADH1* (Zhang *et al.* 2010) was integrated into the *Arabidopsis* genome and assessed for its gene expression to be temporally controlled with estradiol. Three to four week old *Arabidopsis* plants were bombarded with CaLCuV vectors containing 400 or 600 bp repair templates. Plants displaying symptoms of CaLCuV infection (~14 days post bombardment) were sprayed with a solution containing estradiol. Seven days post induction, genomic DNA was extracted and the *ADH1* loci were assessed for non-homologous end joining mutations. Genomic DNA was first predigested with *Bst*XI. The digested DNA was then used in a PCR reaction with primers designed to amplify the ZFN target site within *ADH1.* The resulting amplicons were disgested with *Bst*XI and resolved by gel electrophoresis. Cleavage-resistant product in samples where the plant was exposed to estradiol indicates that ZFN expression was successfully induced. The No Virus samples were of plants bombarded with gold beads without CaLCuV vectors. Plants not exposed to estradiol were not sprayed with the estradiol solution. Genomic DNA from the control experiments (plants without virus infection and plants not exposed to estradiol) were sampled (and sprayed, if applicable) on the same day as the experimental samples.

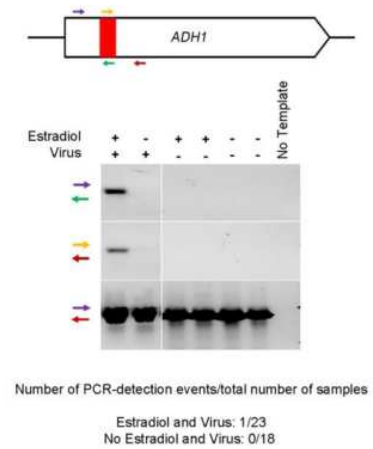

Figure 2-4. Supplemental Figure; PCR-Based Detection of Geminivirus-Mediated Gene Targeting. Genomic DNA from *Arabidopsis* plants infected with CaLCuV harboring repair templates was assessed for the presence of the 18 bp modification within the *ADH1* gene. Illustration of the *ADH1* gene and the primers designed to detect loci that underwent gene targeting (top).

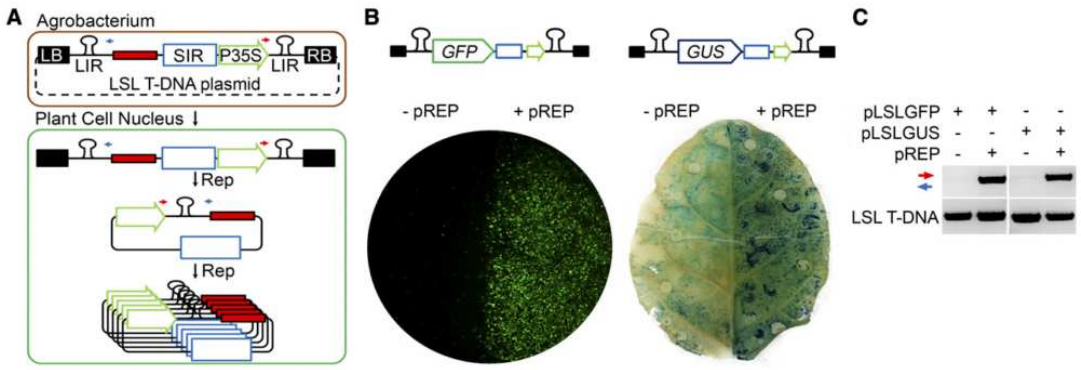

Figure 2-5. Development of Geminivirus Replicons for Protein Expression in Tobacco. (A) Approach to deliver GVRs to plant cells. Rep is delivered using a separate T-DNA expression plasmid (pREP); not shown. LB, left T-DNA border. RB, right T-DNA border. SIR, small intergenic region. LIR, large intergenic region. The red rectangle indicates the region where heterologous sequence is cloned. P35S, 2x35S promoter from the cauliflower mosaic virus. Splice donor and acceptor sequences flank the LIR within circularized replicons; not shown. Blue and red arrows indicate primer binding sites for PCR to detect circularized GVRs. **(B)** Images of leaf tissue expressing GFP and GUS. As a negative control for GVR-mediated protein expression, leaf tissue was infiltrated with a single strain of *Agrobacterium* containing the LSL T-DNA plasmid. Tissue delivered pLSLGFP was imaged three days post infiltration. Tissue delivered pLSLGUS was stained in a solution with X-Gluc seven days post infiltration. **(C)** PCR-based detection of circularized GVRs within plant cells. The PCR control (LSL T-DNA) used primers designed to amplify sequence within linear LSL T-DNA.

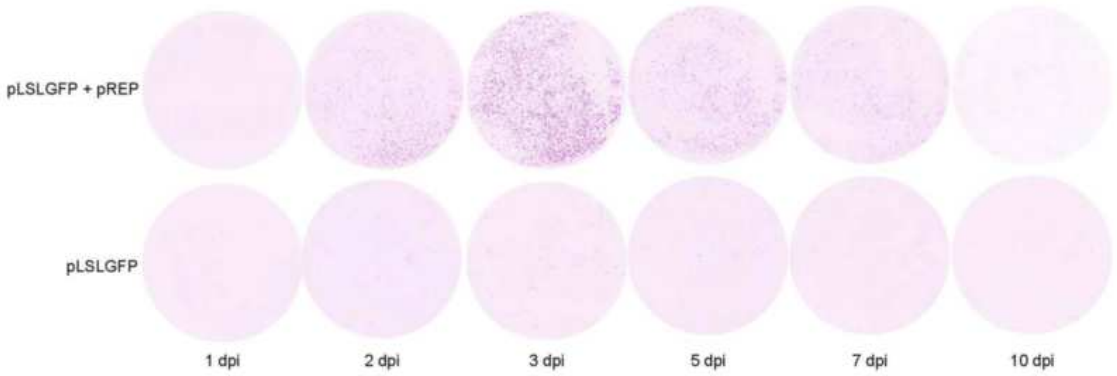

Figure 2-6. Supplemental Figure; Time Course of Geminivirus Replicon-Mediated Expression of Green Fluorescent Protein. Tobacco leaves were infiltrated with two strains of *Agrobacterium* containing pLSLGFP and pREP (top row) or a single strain of *Agrobacterium* containing pLSLGFP. Following infiltration, leaves were cut into discs and placed in water. Images were taken using a camera mounted on a confocal microscope (Nikon A1R-A1). Images are of a single leaf disc. Image colors were inverted to better visualize cells expressing GFP.

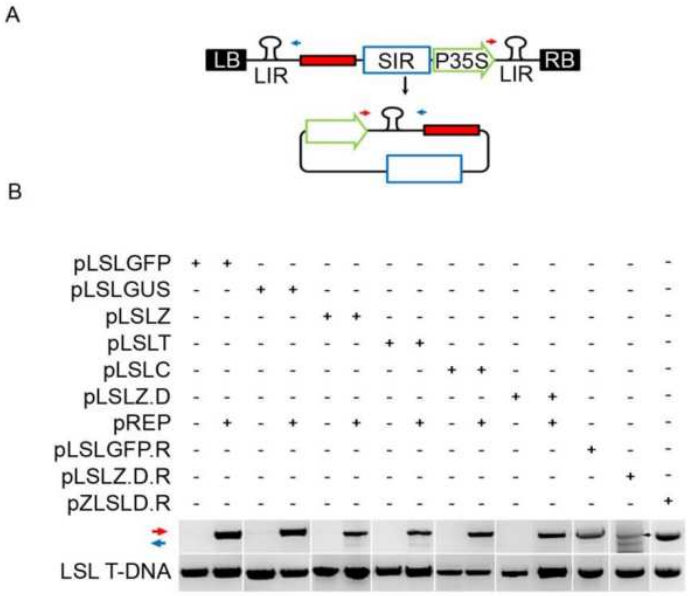

Figure 2-7. Supplemental Figure; PCR-Based Detection of Circularized Geminivirus Replicons. (A) Illustration of replicational release. LSL T-DNA (top), and circularized replicons (bottom). **(B)** PCR-based detection of circularized replicons. LSL T-DNA refers to a control PCR designed to detect sequence present within the LSL T-DNA. Replicational release was assessed in the tissue three to seven days post infiltration. Detection of circular GVRs with samples pLSLGFP.R, pLSLZ.D.R and pZLSLD.R used primers homologous to sequence within Rep and RepA and to sequence within *GFP, Zif268:FokI* or the *us:NPTII* repair template, repectively.

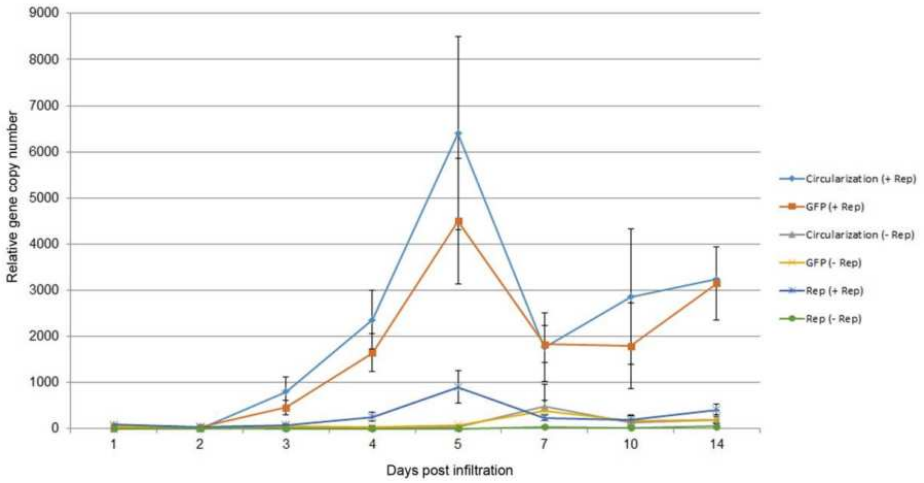

Figure 2-8. Supplemental Figure; Time Course of Geminivirus Replicon Gene Copy Number. Quantitative PCR was performed to determine the relative gene copy number of GVRs over two weeks. Tobacco leaf tissue was infiltrated with two strains of *Agrobacterium* containing pLSLGFP and pREP (these samples are indicated in the figure legend as + Rep) or a single strain of *Agrobacterium* containing pLSLGFP (- Rep). Total DNA was isolated from different leaves at days 1, 2, 3, 4, 5, 7, 10 and 14 days post infiltration. Gene copy numbers were normalized to an internal control (*F-box*). Error bars represent s.e.m. of at least two biological replicates. Primers were used that detect circularized GVRs, GFP coding sequence and Rep coding sequence; and labelled in the figure legend as circularization, GFP and Rep, respectively.

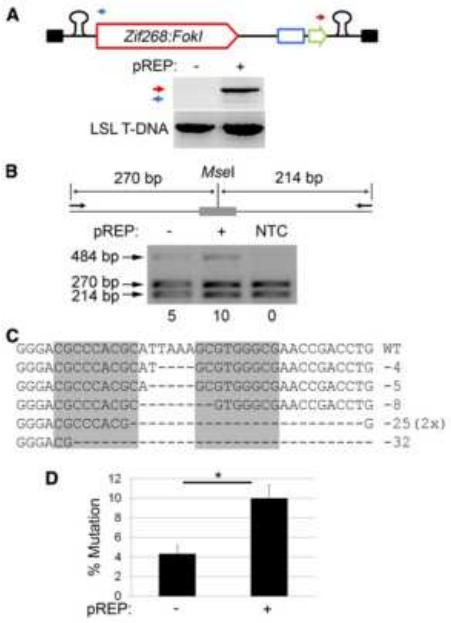

Figure 2-9. Geminivirus Replicon-Mediated Expression of ZFNs for Targeted Mutagenesis. **(A)** Illustration of LSL T-DNA encoding Zif268:*Fok*I (top). PCR-based detection of circularized GVRs within plant cells (bottom). **(B)** PCR-based detection of ZFN-induced mutations at the Zif268 target sequence. Numbers beneath the gel image indicate the percentage of cleavage-resistant amplicons. NTC, non-transformed control. **(C)** Sequences of individual amplicons containing NHEJ-induced mutations. Letters in front of gray background indicate the Zif268 binding sequence. **(D)** Quantification of NHEJ-induced mutations. Error bars represent s.e.m. of three experiments. P* < 0.05.

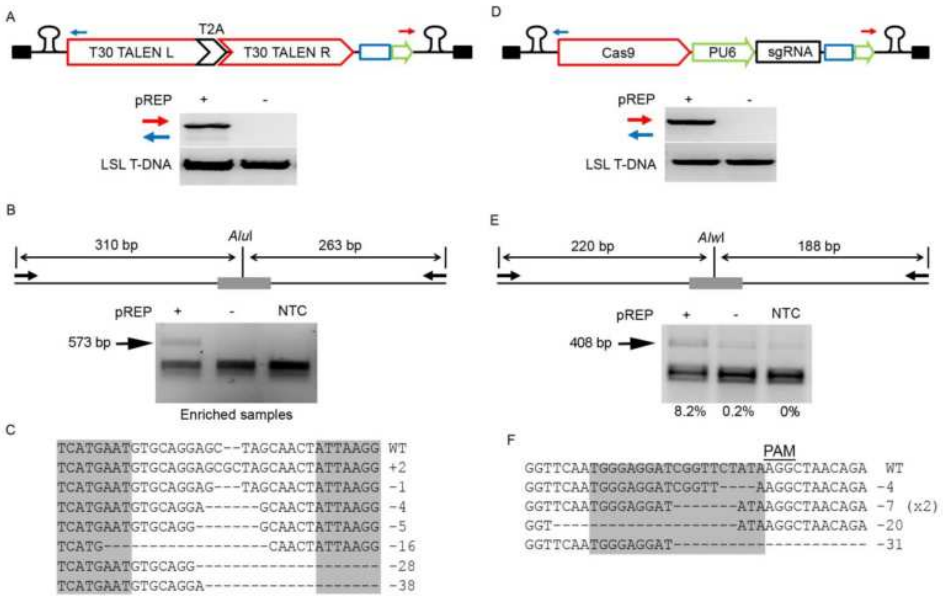

Figure 2-10. Supplemental Figure; Expression of TALENs and Components of the CRISPR/Cas System using Geminivirus Replicons. (A) Illustration of an LSL T-DNA encoding the T30 TALEN pair (top; pLSLT). PCR-based detection of circularized GVRs within plant cells (bottom). T2A, ribosomal skipping sequence. **(B)** Detection of TALEN-induced mutations at the acetolactate synthase (*ALS*) genes. Tobacco leaf tissue was syringe infiltrated with two strains of *Agrobacterium* containing pLSLT and pREP or a single strain of *Agrobacterium* containing pLSLT. Five days post infiltration, genomic DNA was isolated and digested with *Alu*I. Digested DNA was used as a template in a PCR designed to amplify the T30 target site within *SurB.* The resulting amplicons were digested with *Alu*I and bands were separated by gel electrophoresis. NTC, non-transformed control. **(C)** Sequence results from cleavage-resistant amplicons in the sample transformed with pLSLT and pREP T-DNA. **(D)** Illustration of an LSL T-DNA harboring a plant codon optimized Cas9 (*Streptococcus pyogenes*). Single guide RNA (sgRNA) was placed downstream of the *Arabidopsis* U6 promoter (PU6; Top; pLSLC). PCR-based detection of circularized GVRs within plant cells. **(E)** Detection of Cas9-induced mutations at *ALS*. Tobacco leaf tissue was syringe infiltrated with two strains of *Agrobacterium* containing pLSLC and pREP, or a single strain of *Agrobacterium* containing pLSLC. Five days post infiltration, genomic DNA was isolated and used as a template in a PCR designed to amplify the Cas9:sgRNA target site within *ALS.* The resulting amplicons were digested with *Alw*I and bands were separated by gel electrophoresis. The percent of cleavage resistant product seen in the NTC was subtracted from all three samples. **(F)** Sequence results from cleavage-resistant amplicons in the sample transformed with pLSLC and pREP T-DNA. PAM, protospacer adjacent motif.

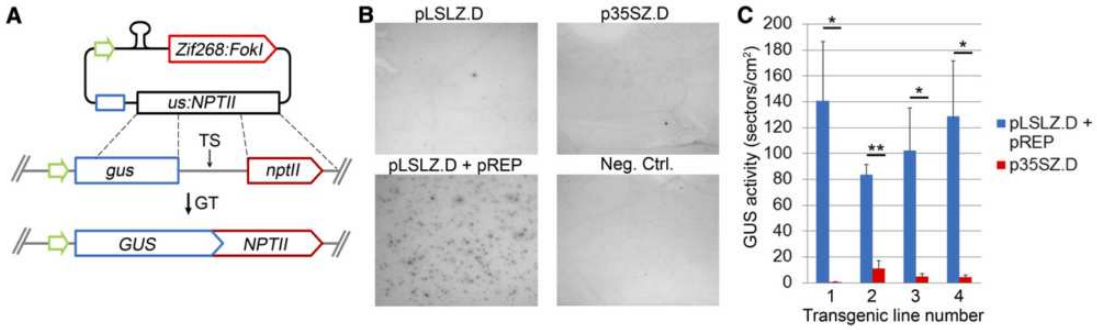

Figure 2-11. Geminivirus Replicons Promote High-Frequency Gene Targeting. (A) Approach to repair a non-functional *gus:nptII* gene through homologous recombination. GT, gene targeting. TS, Zif268:*Fok*I target site. **(B)** Images of leaf tissue stained in a solution with X-Gluc. To better visualize stained cells, chlorophyll was removed from leaf tissue and the green and blue channels were removed from the image. The Neg. Ctrl. image is of non-transgenic leaf tissue transformed with two strains of *Agrobacterium* containing pLSLZ.D and pREP. **(C)** Relative frequencies of gene targeting using GVRs (pLSLZ.D + pREP, blue bars) and conventional T-DNA (p35SZ.D, red bars). The X-axis represents four different plant lines with the *gus:nptII* gene integrated at different chromosomal positions. Error bars represent s.e.m. of at least three biological replicates. P* < 0.05. P** < 0.005.

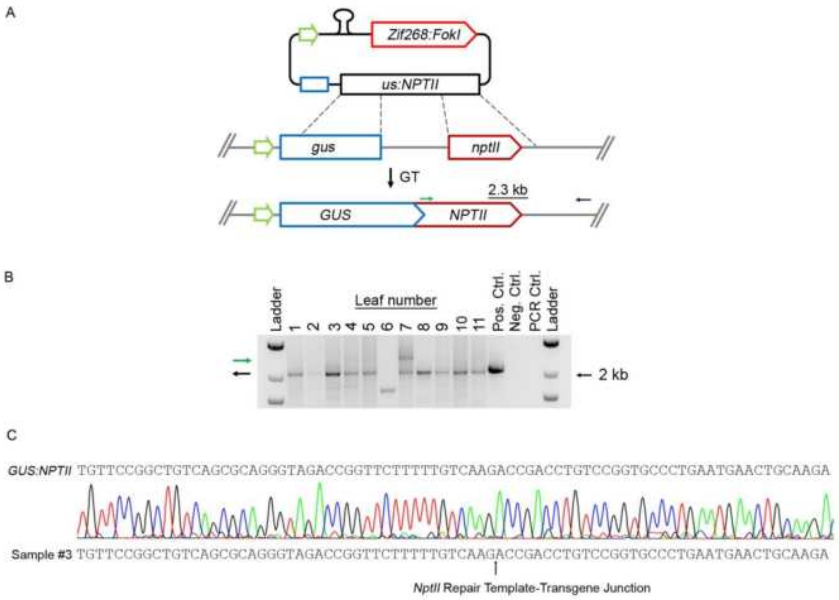

Figure 2-12. Supplemental Figure; PCR-Based Detection of the Repaired *GUS:NPTII* Transgene in Tobacco Leaf Cells. (A) Illustration of the approach to repair the *gus:nptII* transgene. GT, gene targeting. **(B)** Molecular analysis of leaf tissue delivered pLSLZ.D and pREP T-DNA for the repair of the *gus:nptII* transgene. Pos. Ctrl. used plasmid DNA containing the corrected *GUS:NPTII* transgene. PCR ctrl. used water in replace of genome DNA. Each lane represents a different leaf transformed with pLSLZ.D and pREP. Genomic DNA was extracted five days post infiltration. **(C)** DNA sequence of amplicons from leaf number 3.

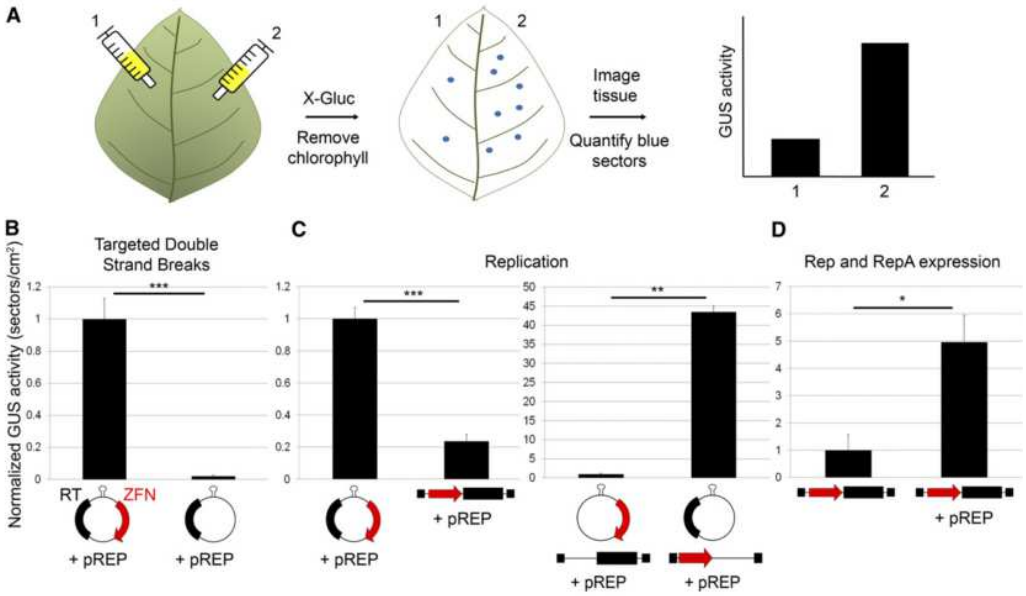

Figure 2-13. Synergism between Double-Strand Breaks, Replication of Repair Templates and Pleiotropic Activity of Rep and RepA on Gene Targeting. (A) Illustration of the approach to directly compare gene targeting frequencies between two *Agrobacterium* samples. **(B)** Effect of targeted double-strand breaks on gene targeting. As indicated, red arrows represent coding sequence for Zif268:*Fok*I; black boxes depict the *us:NPTII* repair templates (RT); small black boxes represent the left and right T-DNA borders. Error bars represent s.e.m. of six experiments. P*** < 0.001. Data were normalized to sample 1. **(C)** Effect of replicating ZFN and repair template sequence on gene targeting (left graph). Error bars represent s.e.m. of four experiments. P*** < 0.001. Effect of replicating repair template sequence on gene targeting (right graph). Error bars represent s.e.m. of three experiments. P** < 0.005. Data were normalized to sample 1 in both graphs. **(D)** Effect of Rep and RepA expression on gene targeting. Measure of center is mean. Error bars represent s.e.m. of four experiments. P* < 0.05. Data were normalized to sample 1.

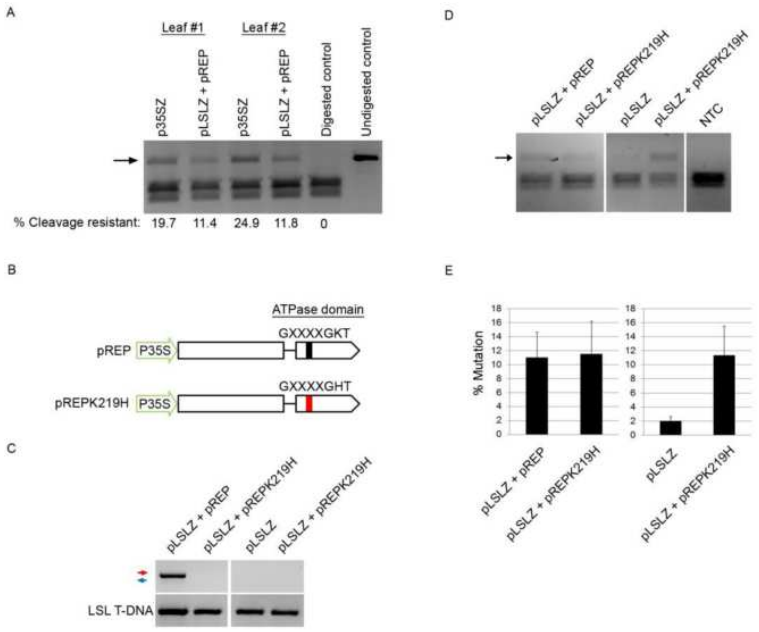

Figure 2-14. Supplemental Figure; Replication of Sequence-Specific Nucleases Does Not Enhance Targeted Mutagenesis. (A) PCR-based detection of ZFN-induced mutations at the Zif268 target site. Numbers beneath the gel image indicate the percentage of cleavage resistant amplicons. p35SZ refers to a T-DNA plasmid harboring P35S:*Zif268:FokI.* The black arrow points to cleavage-resistant product. **(B)** Illustration of the Rep/RepA coding sequence within pREP. A mutation was introduced into the Rep and RepA coding sequence within the conserved ATPase domain (K219H). This K219H mutation is predicted to result in the loss of replication activity (Desbiez *et al.,* 1995). **(C)** Replicational release of GVRs was assessed following leaf infiltration with *Agrobacterium* containing pREP or pREPK219H. Leaves were infiltrated with a mixture of *Agrobacterium* containing pLSLZ and pREP (one side of a leaf), and pLSLZ and pREPK219H (the other side of a leaf). Additional leaves were infiltrated with *Agrobacterium* containing pLSLZ and a mixture of *Agrobacterium* containing pLSLZ and pREPK219H. **(D)** Genomic DNA was assessed for the presence of NHEJ-induced mutations at the Zif268 target site. **(E)** Percentage of NHEJ-induced mutations at the Zif268 target. Error bars represent s.e.m. of three biological replicates. P values for the left and right graphs are 0.72 and 0.12, respectively.

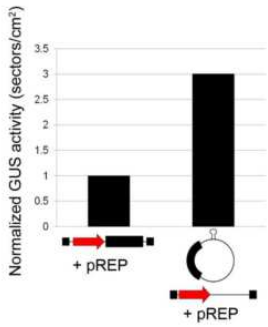

Figure 2-15. Supplemental Figure; Exploring the Role of Repair Templates for Geminivirus Replicon-Mediated Gene Targeting. The impact of replicating repair templates on gene targeting was assessed. Data were normalized to sample 1 (p35SZ.D and pREP). Data represents results from one leaf.

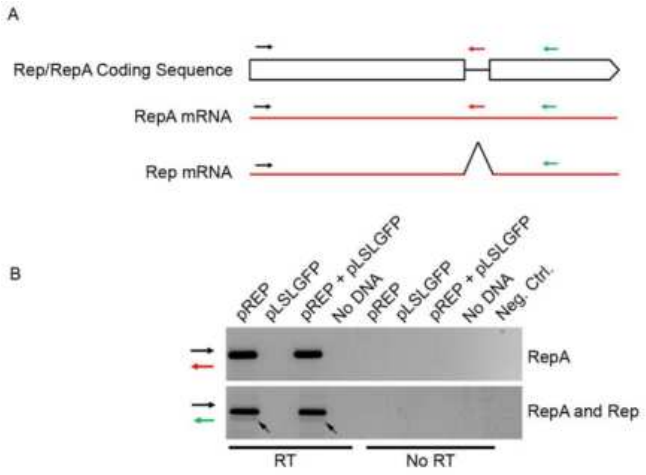

Figure 2-16. Supplemental Figure; Detection of Rep and RepA Transcripts in Tobacco Protoplasts. (A) Illustration of the Rep and RepA coding sequence (top) and mRNA products. The intron within Rep and RepA coding sequence is depicted with a black line. Primer binding sites are indicated with colored arrows. Intron size is 86 bp. **(B)** RT-PCR of Rep and RepA transcripts. Protoplasts were transfected with plasmids pREP, pLSLGFP, pREP and pLSLGFP, or water. RNA was extracted 24 hours post transfection. RT, reverse transcriptase. Neg. Ctrl. water was used in replace of RNA. Two bands were present in the bottom gel image. The upper band is the unspliced mRNA (RepA) and the bottom band is the spliced mRNA (Rep).

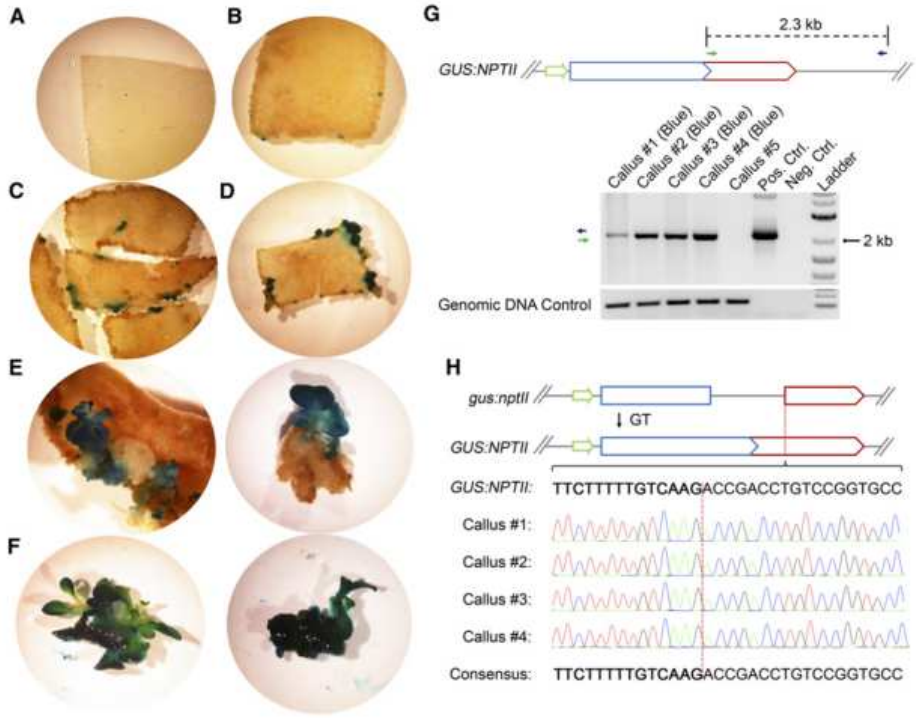

Figure 2-17. Regeneration of Cells with the Repaired *GUS:NPTII* Transgene. **(A)** Image of a leaf disc stained in X-Gluc 0 days after plating. **(B)** Image of a leaf disc stained in X-Gluc 7 days after plating. **(C)** Image of leaf discs stained in X-Gluc 14 days after plating. **(D)** Image of a leaf disc stained in X-Gluc 21 days after plating. **(E)** Shoots with GUS activity 42 days after plating. **(F)** Shoots with GUS activity 49 days after plating. **(G)** PCR-based detection of the repaired *GUS:NPTII* transgene. Total DNA was extracted from four calli that stained blue. As a negative control, genomic DNA was extracted from callus that did not stain blue (Callus #5). Primers were designed to be homologous to sequence within the 600 bp modification sequence and approximately 1 kb downstream of the homologous DNA region carried by the repair template. The Genomic DNA Control used primers designed to amplify sequence within the endogenous *F-box* gene. **(H)** DNA sequences of amplicons generated from calli #1 – 4.

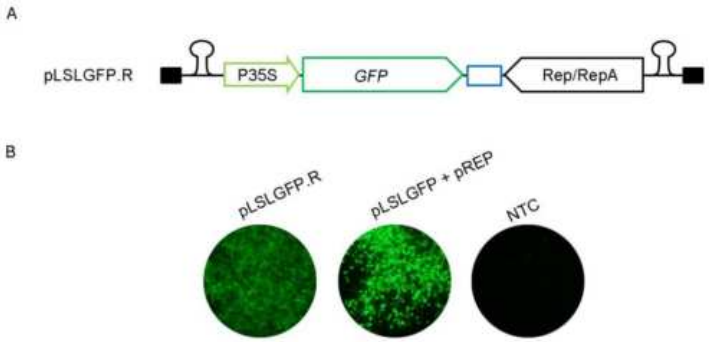

Figure 2-18. Supplemental Figure; Expression of Green Fluorescent Protein with Single-Component Geminivirus Replicon Vectors. (A) Illustration of the single-component GVR vector for expression of GFP. **(B)** Images of leaf tissue expressing GFP. One side of a leaf was infiltrated with *Agrobacterium* containing pLSLGFP.R and the other side was infiltrated with a mixture of *Agrobacterium* containing pLSLGFP and pREP. The non-transformed control (NTC) is an image of leaf tissue not infiltrated with *Agrobacterium.* Images were taken three days post infiltration, at the same magnification and exposure.

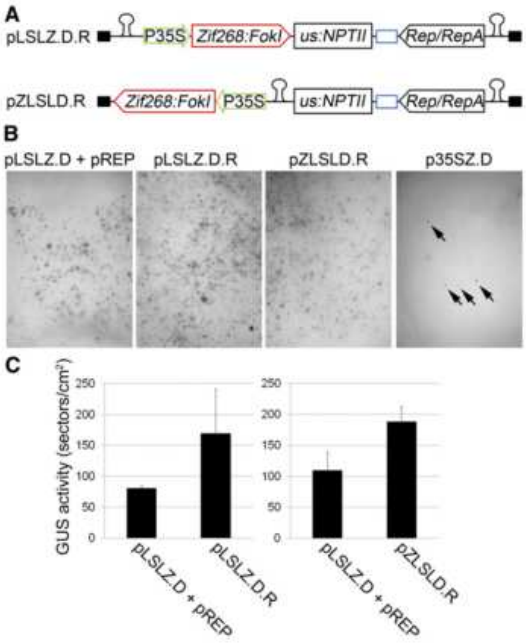

Figure 2-19. Single-Component Geminivirus Vectors Enable Efficient Genome Editing. (A) Illustration of single-component GVR vectors. The blue box represents SIR sequence. **(B)** Selected images of leaves transformed with pLSLZ.D and pREP, pLSLZ.D.R, pZLSLD.R or p35SZ.D. Green and blue channels were removed from the images to better visualize GUS staining. **(C)** Quantification of GUS activity in leaf tissue delivered GVR constructs. The left graph represents data from three leaves transformed with pLSLZ.D and pREP on one side and pLSLZ.D.R on the other side. The right graph represents data from two leaves transformed with pLSLZ.D and pREP on one side and pZLSLD.R on the other. Error bars represent s.e.m.. P values for both graphs were > 0.05.

>Upstream LIR
TAGCAGAAGGCATGTTGTTGTGACTCCGAGGGGTTGCCTCAAACTCTATCTTATAACCGGCGTGGAGGCATGGAGGC
AGGGGTATTTTGGTCATTTTAATAGATAGTGGAAAATGACGTGGAATTTACTTAAAGACGAAGTCTTTGCGACAAGGG
GGGGCCCACGCCGAATTTAATATTACCGGCGTGGCCCCCCCTTATCGCGAGTGCTTTAGCACGAGCGGTCCAGATTT
AAAGTAGAAAATTTCCCGCCCACTAGGGTTAAAGGTGTTCACACTATAAAAGCATATACGATGTGATGGTATTTGATG
GAGCGTATATTGTATCAGGTATTTCCGTTGGATACGAATTATTCGTACGACCCTC
>P35S
GGTCAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAA
TTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAA
GATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTG
CCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTC
AAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTC
AGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTAT
CTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCAT
CGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGAC
GTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTAT
CCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACCTCGACTCTAGAGGATCCC
>Zif268:FokI:NosT
ATGGCTTCCTCCCCTCCAAAGAAAAAGAGAAAGGTTCTTGAATTGCCTTACGCTTGTCCAGTTGAGTCTTGTGATAGA
CGTTTTAGTAGGTCAGACGAACTCACTAGGCATATTAGAATCCACACAGGACAGAAGCCTTTCCAGTGTAGAATCTGT
ATGAGGAACTTTAGTCGTAGCGATCATTTGACTACACATATCAGAACCCATACTGGTGAGAAACCATTCGCATGTGAC
ATCTGTGGTCGTAAGTTCGCAAGAAGTGATGAAAGGAAGCGTCATACAAAGATCCATCTTCGTCAGAAGGAACTAGT
GAAATCTGAATTGGAAGAGAAGAAATCTGAACTTAGACATAAATTGAAATATGTGCCACATGAATATATTGAATTGATT
GAAATCGCAAGAAATTCAACTCAGGATAGAATCCTTGAAATGAAGGTGATGGAGTTCTTTATGAAGGTTTATGGTTATC
GTGGTAAACATTTGGGTGGATCAAGGAAACCAGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTGA
TCGTTGATACTAAGGCATATTCAGGAGGTTATAATCTTCCAATTGGTCAAGCAGATGAAATGCAAAGATATGTCGAAG
AGAATCAAACAAGAAACAAGCATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCAGTAACAGAATTTAAGTT
CTTGTTTGTGAGTGGTCATTTCAAAGGAAACTACAAAGCTCAGCTTACAAGATTGAATCATATCACTAATTGTAATGGA
GCTGTTCTTAGTGTAGAAGAGCTTTTGATTGGTGGAGAAATGATTAAAGCTGGTACATTGACACTTGAGGAAGTGAGA
AGGAAATTTAATAACGGTGAGATAAACTTTTAATAGGACGTCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGA
TTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTA
ATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAA
ATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAATTGATCCCCCCTCGA
CAGCT
>us:NPTII repair template
AGCAGTCTTACTTCCATGATTTCTTTAACTATGCCGGAATCCATCGCAGCGTAATGCTCTACACCACGCCGAACACCT
GGGTGGACGATATCACCGTGGTGACGCATGTCGCGCAAGACTGTAACCACGCGTCTGTTGACTGGCAGGTGGTGGC
CAATGGTGATGTCAGCGTTGAACTGCGTGATGCGGATCAACAGGTGGTTGCAACTGGACAAGGCACTAGCGGGACT
TTGCAAGTGGTGAATCCGCACCTCTGGCAACCGGGTGAAGGTTATCTCTATGAACTGTGCGTCACAGCCAAAAGCCA
GACAGAGTGTGATATCTACCCGCTTCGCGTCGGCATCCGGTCAGTGGCAGTGAAGGGCGAACAGTTCCTGATTAAC
CACAAACCGTTCTACTTTACTGGCTTTGGTCGTCATGAAGATGCGGACTTGCGTGGCAAAGGATTCGATAACGTGCT
GATGGTGCACGACCACGCATTAATGGACTGGATTGGGGCCAACTCCTACCGTACCTCGCATTACCCTTACGCTGAAG
AGATGCTCGACTGGGCAGATGAACATGGCATCGTGGTGATTGATGAAACTGCTGCTGTCGGCTTTAACCTCTCTTTA
GGCATTGGTTTCGAAGCGGGCAACAAGCCGAAAGAACTGTACAGCGAAGAGGCAGTCAACGGGGAAACTCAGCAAG
CGCACTTACAGGCGATTAAAGAGCTGATAGCGCGTGACAAAAACCACCCAAGCGTGGTGATGTGGAGTATTGCCAAC
GAACCGGATACCCGTCCGCAAGGTGCACGGGAATATTTCGCGCCACTGGCGGAAGCAACGCGTAAACTCGACCCGA
CGCGTCCGATCACCTGCGTCAATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCTCTTTGATGTGCTGT
GCCTGAACCGTTATTACGGATGGTATGTCCAAAGCGGCGATTTGGAAACGGCAGAGAAGGTACTGGAAAAAGAACTT
CTGGCCTGGCAGGAGAAACTGCATCAGCCGATTATCATCACCGAATACGGCGTGGATACGTTAGCCGGGCTGCACT
CAATGTACACCGACATGTGGAGTGAAGAGTATCAGTGTGCATGGCTGGATATGTATCACCGCGTCTTTGATCGCGTC
AGCGCCGTCGTCGGTGAACAGGTATGGAATTTCGCCGATTTTGCGACCTCGCAAGGCATATTGCGCGTTGGCGGTA
ACAAGAAAGGGATCTTCACTCGCGACCGCAAACCGAAGTCGGCGGCTTTTCTGCTGCAAAAACGCTGGACTGGCAT
GAACTTCGGTGAAAAACCGCAGCAGGGAGGCAAACAACGCAGGGAGGCAAACAATGATATCACAACTCTCCTGACG
CGTCATCGTCGGCTACAGCCTCGGGAATTGCTACCTAGCTCGAGCAAGATCCAAGGAGATATAACAATGGCTTCCTC
CTGGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCA
CAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGTAGACCGGTTCTTTTTGTCAAGAC
CGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTACC
TTGCGCTGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGA
TCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTG
ATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCT
TGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAG
AATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGC
TTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATAT
TGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGC
ATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGATAACCGCGGAGAGCTCGAATTTCCCCGATCGTTCAAACATTT
GGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAG
CATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTA
ATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCG
TGGGCCCGTCCAGTTCCATTCGCCCTATAGTGAGTCGTATTAAACCTGCAT TAATGAATCGGCCAACGCGCGGGGCT
>SIR

AATAAAATGATTATTTTATGAATATATTTCATTGTGCAAGTAGATAGAAATTACATATGTTACATAACACACGAAATAAAC
AAAAAAAGACAATCCAAAAACAAACACCCCAAAAAAAATAATCACTTTAGATAAACTCGTATGAGGAGAGGCACGT
>Rep/RepA (reverse complemented)
TCAGTGACTCGACGATTCCCGAGCAAAAAAAGTCTCCCCGTCACACATGTAGTGGGTGACGCAATTATCTTTAAAGTA
ATCCTTCTGTTGACTTGTCATTGATAACATCCAGTCTTCGTCAGGATTGCAAAGAATTATAGAAGGGATCCCACCTTTT
ATTTTCTTCTTTTTTCCATATTTAGGGTTGACAGTGAAATCAGACTGGCAACCTATTAATTGCTTCCACAATGGGACGA
ACTTGAAGGGGATGTCGTCGATGATATTATAGGTGGCGTGTTCATCGTAGTTGGTGAAATCGATGGTACCGTTCCAAT
AGTTGTGTCGTCCGAGACTTCTAGCCCAGGTGGTCTTTCCGGTACGAGTTGGTCCGCAGATGTAGAGGCTGGGGTG
TCGGATTCCATTCCTTCCATTGTCCTTGTTAAATCGGCCATCCATTCAAGGTCAGATTGAGCTTGTTGGTATGAGACA
GGATGTATGTAAGTATAAGCGTCTATGCTTACATGGTATAGATGGGTTTCCCTCCAGGAGTGTAGATCTTCGTGGCAG
CGAAGATCTGATTCTGTGAAGGGCGACACATACGGTTCAGGTTGTGGAGGGAATAATTTGTTGGCTGAATATTCCAG
CCATTGAAGCTTTGTTGCCCATTCATGAGGGAATTCTTCCTTGATCATGTCAAGATATTCCTCCTTAGACGTTGCAGTC
TGGATAATAGTTCTCCATCGTGCGTCAGATTTGCGAGGAGAAACCTTATGATCTCGGAAATCTCCTCTGGTTTTAATAT
CTCCGTCCTTTGATATGTAATCAAGGACTTGTTTAGAGTTTCTAGCTGGCTGGATATTAGGGTGATTTCCTTCAAAATC
GAAAAAAGAAGGATCCCTAATACAAGGTTTTTTATCAAGCTGGAGAAGAGCATGATAGTGGGTAGTGCCATCTTGATG
AAGCTCAGAAGCAACACCAAGGAAGAAAATAAGAAAAGGTGTGAGTTTCTCCCAGAGAAACTGGAATAAATCATCTCT
TTGAGATGAGCACTTGGGATAGGTAAGGAAAACATATTTAGATTGGAGTCTGAAGTTCTTACTAGCAGAAGGCAT
>Downstream LIR
GTTGTTGTGACTCCGAGGGGTTGCCTCAAACTCTATCTTATAACCGGCGTGGAGGCATGGAGGCAGGGGTATTTTGG
TCATTTTAATAGATAGTGGAAAATGACGTGGAATTTACTTAAAGACGAAGTCTTTGCGACAAGGGGGGGCCCACGCC
GAATTTAATATTACCGGCGTGGCCCCCCCTTATCGCGAGTGCTTTAGCACGAGCGGTCCAGATTTAAAGTAGAAAATT
TCCCGCCCACTAGGGTTAAAGGTGTTCACACTATAAAAGCATATACGATGTGATGGTATTTGATGGAGCGTATATTGT
ATCAGGTATTTCCGTTGGATACGAATTATTCGTACGACCCTC

Figure 2-20. Supplemental Figure; Sequence of Features within the Single-Component Geminivirus Vector pLSLZ.D.R.

CHAPTER 3

TRANSIENT EXPRESSION OF ZINC-FINGER NUCLEASES IN ARABIDOPSIS USING TOBACCO RATTLE VIRUS

Baltes NJ and Voytas DF (2014). Transient expression of zinc-finger nucleases in Arabidopsis using tobacco rattle virus. Unpublished.

ABRIDGEMENT

Editing genomic DNA using sequence-specific nucleases is now becoming routine practice in many organisms. However, a barrier to achieving efficient genome editing in plants is the delivery to plant cells. The two most commonly-used delivery methods for genome engineering reagents are stable integration into the plants genome, and direct delivery to protoplasts. Unfortunately, both methods have weaknesses, including the time-consuming and laborious procedure of integrating DNA into plant genomes, and the limited number of plant species that can be regenerated from protoplasts. Recently, tobacco rattle virus was shown to be an effective vector for the delivery of nucleases in tobacco (*Nicotiana benthamiana*) and petunia (*Petunia hybrida*). Here we expand upon the utility of tobacco rattle virus for genome engineering by delivering zinc-finger nucleases in the model organism *Arabidopsis thaliana*. First, to characterize virus movement and protein expression in *Arabidopsis* plants, we engineered the virus genome to encode the green fluorescent protein. Using these virus vectors, we infected *Arabidopsis* plants and show virus trafficking into rosette leaves, cauline leaves and floral tissues. By modifying TRV to carry a zinc-finger nuclease (Zif268:*Fok*I) and GFP, we demonstrated transient delivery of nucleases to leaf cells by detecting targeted mutations in leaves where the virus has moved (as indicated by cells with GFP expression). Lastly, we investigated whether TRV-mediated delivery of Zif268:*Fok*I could introduce heritable mutations. The reporter that was stably integrated within *Arabidopsis* plants was an integrated and non-functional beta-glucuronidase gene (GU.US) that can be reconstituted by a Zif268:*Fok*I-induced double-strand break, and subsequent repair by the single-strand annealing pathway. Therefore, if tobacco rattle virus was able to deliver Zif268:*Fok*I to germline cells, we would expect to find seedlings harboring the reconstituted GUS gene, and consequently these seedlings would stain blue when incubated with an appropriate substrate. After screening ~ 10,000 seedlings from 16 different infected parent plants, we found GUS activity in the leaves and roots of 5 seedlings from two different parents, suggesting mutations may have occurred in seed-progenitor cells. Results from this study expand the utility of TRV vectors to the delivery of nucleases in the major model organism, *Arabidopsis thaliana*.

INTRODUCTION

One method to efficiently edit DNA at a sequence-of-interest in plant cells is to introduce a targeted double-strand break (Puchta et al., 1996; Lloyd et al., 2005; Zhang et al., 2010). Normally, double-strand breaks are toxic lesions that, if left unrepaired, can result in cell death. To preserve the integrity of their genomes, cells encode two main DNA repair pathways: non-homologous end joining and homologous recombination. Repair mechanics of these two pathways can be exploited to introduce specific DNA changes into plant chromosomes. Repair by non-homologous end joining can result in insertions or deletions at the break site. Therefore, if

breaks are directed to coding sequences, gene knockout can be facilitated (Zhang et al., 2010). Repair of double-strand breaks by homologous recombination requires a repair template, usually the sister chromatid or homologous chromosome. However, if a synthetic donor molecule is supplied, the cell can copy information from within the donor molecule into the genome (Puchta et al., 1996).

In plants, one barrier to achieving efficient genome editing is the delivery of the reagents to cells. Current methods for delivering nucleases to plant cells include direct delivery to protoplasts (Wright et al., 2005; Townsend et al., 2009; Jiang et al., 2013; Li et al., 2013; Shan et al., 2013b; Cermak et al., 2011; Zhang et al., 2013), stable transformation into plant genomes (Lloyd et al., 2005; Zhang et al., 2010; Christian et al., 2013; Feng et al., 2014; Fauser et al., 2014; Haun et al., 2014; Ayar et al., 2012; Osakabe et al., 2010; Petolino et al., 2010; Fauser et al., 2012; Wendt et al., 2013), transient delivery using *Agrobacterium* T-DNA (Puchta et al., 1996; Reiss et al., 2000; de Pater et al., 2009; Li et al., 2013; Jiang et al., 2013), silicon carbide whiskers (Shukla et al., 2009) and biolistics (Shan et al., 2013b; Ainley et al., 2013). Although these methods are effective, there are several challenges preventing their general use, including a required tissue culture expertise because target cells are from cell cultures or calli, and the time consuming process of stably integrating genome engineering reagents into a plants genome (and then possibly back-crossing to remove the reagents after the targeted modification is introduced). There are several benefits to working with whole plants, including the ease of growing and handling the plant material; however, there are few methods for delivering genome engineering reagents to whole plants. Therefore, the development of additional delivery methods in whole plants is critical for advancing the practice of plant genome engineering.

Recently, two studies have demonstrated that plant viruses may be effective delivery vehicles for genome engineering reagents (Marton et al., 2010; Baltes et al., 2014). Advantages to using viruses are their large host range, and extrachromosomal replication of their genomes which enables the transiently delivery reagents to plant cells. Furthermore, viruses may encode proteins that increase the likelihood of creating the desired modification (Baltes et al., 2014). Because there is a need for developing additionally delivery methods, there is value to expanding upon these proof-of-concept experiments to additional hosts and different target sequences within the host's genome. Here, we focus our attention on the RNA virus, tobacco rattle virus (TRV), and describes its utility in *Arabidopsis thaliana.*

TRV is a positive-strand RNA virus that replicates through a double-stranded RNA intermediate. TRV has two genomes (RNA 1 and RNA 2). The RNA1 genome encodes elements needed for replication and cell-to-cell movement. Genes present on the RNA 1 genome include the RNA-dependent RNA polymerase (responsible for replicating the virus genome), 29K (movement protein) and 16K (suppressor of RNA silencing). Genes present on the RNA 2

genome include the coat protein, 29K (2b) and 18K (2c). Both 2b and 2c proteins are required for nematode transmission.

There has been significant interest in converting TRV into plant vectors for virus-induced gene silencing (VIGS) (Ratcliff et al., 2001) and protein expression. There are several advantages for using TRV, including the mild symptoms plants have when infected with TRV, the silencing efficiency of host genes, and the large host range of TRV. Recently, Marton *et al.,* 2010, demonstrated the usefulness of TRV to serve as a vector for zinc-finger nuclease expression in tobacco and petunia. Here, tobacco and petunia plants were modified to harbor a stably integrated zinc-finger nuclease target site. By infecting these plants with TRV encoding the corresponding zinc-finger nuclease, the authors demonstrate targeted mutations in somatic tissue. The authors then go on to regenerate modified plants by culturing regenerative tissues and developing buds.

Results from the Marton *et al.,* 2010 study, which demonstrated that TRV can effectively deliver zinc-finger nucleases to tobacco and petunia plants, combined with the advantages of virus-based delivery systems, prompted us to explore the utility of TRV for the transient delivery of zinc-finger nucleases in *Arabidopsis*. We first engineered virus vectors that were tagged with GFP and non-destructively visualized TRV movement throughout infected plants. We confirmed that TRV moves throughout the entire *Arabidopsis* plant and can be found in rosettes, cauline leaves and floral tissue. We then demonstrate that TRV can be an effective delivery vehicle for zinc-finger nucleases in somatic tissue. Here, we designed TRV vectors that co-express Zif268:*Fok*I and GFP. In somatic tissue expressing GFP (outside the point of inoculation), we detected non-homologous end joining mutations at the nuclease target site. Finally, we assess next generation seedlings for GUS expression (the target for modification was the GU.US transgene) and found 5 seedlings expressing GUS in leaves and stems from a total of ~10,000 seeds from 16 infected parents.

RESULTS

Engineering TRV vectors for protein expression

To engineer TRV for expressing reporter genes and zinc-finger nucleases, we chose to modify the RNA 2 genome. TRV contains two genomes, RNA 1 and RNA 2 (Figure 1A). The RNA 2 genome encodes the coat protein, 2b and 2c. The 2b and 2c genes produce proteins that are involved in insect transmission and are not required for virus replication or movement within a plant (Ratcliff et al., 2001). Similar to other methods that use tobacco rattle virus to deliver heterologous sequence (Ratcliff et al., 2001), we chose to replace the 2b and 2c genes with our

sequence of interest. To this end, we obtained the TRV T-DNA plasmids pYL192 (contains the entire RNA 1 genome downstream of a duplicated 35S promoter) and pYL156 (contains the RNA 2 genome, minus the 2b and 2c proteins, downstream of a duplicated 35S promoter; herein we refer to this vector as pTRV2) designed for VIGS (Liu et al., 2002) (Figure 3-1A). However, unlike these studies, we are inserting genes downstream of the coat protein, which requires a method to enable translation from an RNA molecule with internal coding sequence. Therefore, we cloned the pea early-browning virus subgenomic promoter downstream of the coat protein and upstream our gene of interest (Figure 3-1B). Protein expression using this promoter has previously been demonstrated in tobacco and petunia (Marton et al., 2010).

Tracking TRV infection and movement within *Arabidopsis*

To determine the efficiency of infection (number of inoculated plants divided by the number of infected plants) and to characterize the movement of our recombinant tobacco rattle viruses within *Arabidopsis* plants, we cloned the green fluorescent protein (GFP) sequence downstream of the pea-early browning virus subgenomic promoter (pTRV2-sgP:GFP; Figure 3-1B). By inserting GFP within the RNA 2 genome, we could non-destructively track the virus as it moves throughout the *Arabidopsis* plant. To initiate infection, leaves from 2 - 3 week old *Arabidopsis* plants were syringe infiltrated with a mixture of *Agrobacterium* containing pYL192 and pTRV2-sgP:GFP. Approximately seven days after infiltration, plants were assessed for successful infection. Our criterion for determining if a plant was infected was if GFP expression was found in non-infiltrated and newly-formed upper rosette leaves. From three different experiments, a total of 62 *Arabidopsis* plants were infiltrated, and, of the infiltrated plants, 14 plants expressed GFP, thereby yielding an infection frequency of approximately 22%. Notably, the infection frequency may be higher because mutations within GFP could result in a non-functional protein, while the virus maintains the ability to infect and spread systemically.

After isolating plants expressing GFP, we then tracked TRV movement. In all infected plants, we observed GFP expression in vasculature cells (Figure 3-2A). We also observed GFP expression in rosettes (Figure 3-2B), stems and cauline leaves (Figure 3-2C) and in floral tissue (Figure 3-2D). As expected, the phenotype of plants infected with TRV-GFP was indistinguishable from non-infiltrated controls. These results demonstrate that our TRV vectors were effective vehicles for delivering GFP in many different tissues within *Arabidopsis*, including leaves and growing points.

To engineer TRV for delivery of zinc-finger nucleases, we initially cloned Zif268:*Fok*I downstream of the pea early-browning virus subgenomic promoter. However, after infiltrating leaves of *Arabidopsis* plants with *Agrobacterium* containing these virus vectors, we had challenges finding infected plants, and, within these infected plants, we couldn't easily detect the

tissue that contained virus. Therefore, we sought to optimize our TRV vectors to allow for easier detection and characterization of the virus position within *Arabidopsis* plants.

There are a number of potential inefficiencies to address in our vector design, including the low frequency of infection, possible genome instability, and uneven movement throughout a plant. Results from our experiments using pTRV2-sgP:GFP demonstrated that, in our hands, the infection frequency is relatively low (approximately 20% of plants become infected), and the infected plants have mild symptoms. Without a phenotype, and without a GFP marker, it will be challenging to identify infected plants. Additionally, within plants that are successfully infected, the nuclease coding sequence may be destroyed or lost by a selective pressure to lose Zif268:*Fok*I. And finally, TRV movement throughout an infected plant may not be uniform, making it difficult to discern which tissues to analyze for mutations. To overcome these potential inefficiencies, we chose to 'tag' the TRV genome with GFP, such that GFP function was linked to the function of Zif268:*Fok*I. This was accomplished by engineering RNA 2 to encode a transcriptionally-linked *Zif268:FokI* and *GFP* sequence (*Zif268:FokI:T2A:GFP*). To preserve function of Zif268*:Fok*I protein (Zif268:*Fok*I may not be active when fused to GFP), a T2A self-cleaving sequence was inserted between GFP and *Zif268:FokI,* permitting the disassociation of the fusion protein after translation (pTRV2-sgP:Zif268:FokI:T2A:GFP; Figure 3-1B). Therefore, not only does the 3' GFP tag permit the facile identification and non-destructive monitoring of the nuclease-expressing virus in *Arabidopsis* plants, it also provides an indirect readout of nuclease stability: frameshift or nonsense mutations occurring in Zif268:*Fok*I coding sequence will subsequently inactivate GFP. However, there may be instances where Zif268:*Fok*I is rendered non-function but GFP retains function. These modifications including deletions, insertions or substitutions with Zif268:*Fok*I coding sequence that inactivate gene function but maintain the reading frame.

We first assessed the infection frequency with our optimized vector and a control vector harboring a mutation within *Fok*I (D450A; predicted to abolish the catalytic activity of *Fok*I). From three separate experiments, a total of 168 plants were infiltrated with a mixture of *Agrobacterium* containing pYL192 and pTRV2-sgP:Zif268:FokI:T2A:GFP, and a total of 116 plants were infiltrated with a mixture of *Agrobacterium* containing pYL192 and pTRV2-sgP:Zif268:FokIΔ:T2A:GFP. GFP expression was found in non-infiltrated and newly-formed upper rosette leaves within 18 plants that received the virus vectors containing the functional *Zif268:FokI* sequence and 6 plants that received the control vectors, resulting in an infection frequency of 10.7% and 5.2%, respectively. The decrease in infection frequency compared to pTRV2-sgP:GFP is possibly due to the vector-size increase (~1 kb additional nucleotides).

Targeted mutagenesis in somatic cells

Using these optimized vectors, we next assessed the ability of TRV to transiently deliver nucleases to somatic cells. Our target for modification was a stably integrated GU.US transgene (Figure 3-1C). Between the direct repeats was a target site for Zif268. One outcome of a DNA double-strand break at the Zif268 target site is repair by the error-prone non-homologous end joining pathway, resulting in insertions or deletions (indels) at the breaksite. We assessed the genomic DNA from infected leaf tissue for these indels. Briefly, *Arabidopsis* plants were infiltrated with a mixture of *Agrobacterium* containing pYL192 and pTRV2-sgP:Zif268:FokI:T2A:GFP. From plants displaying GFP expression (approximately seven days post infiltration) leaf tissue from non-infiltrated rosette leaves was isolated and total genomic DNA was purified. The Zif268 target site was amplified by PCR and the resulting amplicons were digested with *Mse*I (an *Mse*I recognition sequence is present within the spacer region of the Zif268 binding sites). Mutations in amplicons at the Zif268 target site will result in the loss of the *Mse*I site and, as a consequence, will not be digested. These cleavage-resistant products can then be separated from the cleaved product by gel electrophoresis. We observed the presence of an undigested product in tissue samples infiltrated with *Agrobacterium* containing pYL192 and pTRV2-sgP:Zif268:FokI:T2A:GFP, (Figure 3C); and we did not detect cleavage-resistant products in our control experiments. Cloning and sequencing of these cleavage-resistant amplicons confirmed the presence of Zif268:*Fok*I-induced non-homologous end joining mutations (Figure 3D). These results indicate that TRV can transiently deliver nucleases to somatic cells in *Arabidopsis*. Furthermore, finding multiple unique mutations at the Zif268 target site indicates that TRV entered and expressed Zif268:*Fok*I in numerous leaf cells.

We used this approach to attempt to detect Zif268:*Fok*I-induced mutations in the leaves of additional GFP positive plants. In total, we isolated 18 GFP-positive plants that were infiltrated with *Agrobacterium* containing pYL192 and pTRV2-sgP:Zif268:FokI:T2A:GFP. Of these 18 plants, 12 were healthy enough to remove tissue and extract genomic DNA. Interestingly, we did not detect mutations in 9 of 12 infected plants. This may suggest that mutations are occurring at a level below the detection limit of our PCR digest assay (~1% mutagenesis). Alternatively, this could suggest that *Zif268:FokI* was selectively mutated, whereas GFP expression was maintained.

Many of the plants infiltrated with *Agrobacterium* containing pYL192 and pTRV2-sgP:Zif268:FokI:T2A:GFP that were expressing GFP had symptoms (Table 3-2). These symptoms included vein chlorosis, and stunted growth (both roots and shoots). These symptoms were not observed in GFP-expressing plants from the control vectors (TRV2-sgP:GFP or TRV2-sgP:Zif268:FokIΔ:T2A:GFP). Therefore, this phenotype was most likely due to the nuclease activity of Zif268:*Fok*I. This could be caused by both on-target and off-target double-strand breaks occurring within cells and throughout the plant, which may slow down growth and causing cells to

die. Nonetheless, majority of our GFP-expressing plants were able to produce seed, permitting us to further characterize mutations within these plants.

Assessing *Arabidopsis* plants for germinally-transmitted mutations

Although TRV is a non-seed transmissible virus, it has been found to enter meristem cells for a brief period of time before it is ultimately excluded (Martín-Hernández and Baulcombe, 2008). This may suggest that, in *Arabidopsis* plants infected with TRV expressing a sequence-specific nuclease, mutations will occur in cells that give rise to seed. We explored this hypothesis by attempting to detect heritable Zif268:*Fok*I-induced mutations in the next generation seedlings. We predicted that, if this hypothesis is true, our transgenic GU.US plants that are infected with TRV (TRV1 and TRV2 RNA2 TRV2-sgP:Zif268:FokI:T2A:GFP will give rise to seedlings that express GUS. Here, we assumed that double-strand breaks occurring in seed-progenitor cells will, at a certain frequency, be repaired by the single-strand annealing pathway using the direct repeats in the GU.US gene. These cells will then give rise to seed, and a subsequent seedlings that express GUS protein.

To test this hypothesis, we isolated seeds from 16 GFP-positive plants infiltrated with *Agrobacterium* containing pYL192 and pTRV2-sgP:Zif268:FokI:T2A:GFP. These seeds were sterilized and plated on MS agar media. Two weeks after plating, the resulting seedlings were collectively stained in a solution with X-Gluc. From a total of ~10,000 seeds, we found evidence of GUS activity in the leaves and roots of five seedlings (Figure 3-4). These seedlings came from two parent plants that we detected non-homologous end joining mutations in leaf cells (1.1 and 3.4; Table 3-2). After determining that plant lines 1.1 and 3.4 harbored non-homologous end joining mutations in somatic cells, and that they gave rise to one or more GUS-expressing next-generation seedlings, we tried to isolate living plants harboring a modification at the GU.US locus. To this end, we grew seedlings from these two lines (~170 in total) in soil and extracted genomic DNA from a rosette leaf. We tested these seedlings for (i) the presence of a non-homologous end joining mutation, (ii) the reconstitution of the GUS allele, or (iii) GUS activity (here, one rosette was removed and stained in a solution containing X-Gluc). In all of the tested seedlings, we did not detect GUS activity. Furthermore, we did not detect mutations or the reconstitution of the GU.US transgene, suggesting that the germinal transmission frequency is < 1%.

Noteworthy, and not including the lines that yielded the five blue seedlings, we found an additional four seedlings with partial GUS activity (e.g., GUS activity in the roots but not shoots). These seedlings either came from a plant that was infected with TRV encoding Zif268*:Fok*I, or they came from a control plant infected with TRV encoding Zif268*:Fok*IΔ (Figure 3-4). This suggests that partial GUS activity in the roots of next-generation seedlings is not due to a Zif268:*Fok*I-induced double-strand break. We did not find any seedlings staining completely blue

with our other controls (pTRV2-sgP:GFP or no-*Agrobacterium*). Together, these results suggest that TRV can facilitate the mutagenesis of seed-progenitor cells; however, further work is needed to confirm this observation (e.g., screen more seed from infected plants and molecular confirmation of the modified target site).

DISCUSSION AND FUTURE DIRECTIONS

TRV has great potential to be an effective delivery vehicle for sequence-specific nucleases. TRV has a host range of over 60 species, including soybean, tomato and potato, which would enable transient editing of DNA in many economically important crops. Furthermore, TRV is a mild virus, causing very little symptoms within infected plants; and TRV is strictly RNA, which would circumvent government regulation regarding plants that have been delivered foreign DNA.

In this study, we demonstrated the utility of tobacco rattle virus for the delivery of nucleases in *Arabidopsis*. We show that indeed tobacco rattle viruses encoding zinc-finger nucleases can spread throughout *Arabidopsis* plants. We further demonstrate that cells containing the virus also contain mutations at the nuclease target site. And finally, we assessed next-generation seedlings for evidence of double-strand breaks at the nuclease target site. We found, from two parent plants that were infected with TRV, a total of five seedlings expressing GUS in roots and shoots. As an alternative approach to obtain modified *Arabidopsis* plants, it may be possible to regenerate plants from leaves harboring somatic mutations. The resulting population of plants could then be screened for heritable mutations.

Whereas our results suggest TRV can deliver nucleases to seed-progenitor cells, there are several caveats with our approach. First, we assume that double-strand breaks (and subsequent repair by single-strand annealing) occurring within seed-progenitor cells will be detectable as a heterozygous or homozygous allele in the next-generation seedlings. However, Zif268:*Fok*I-induced double-strand breaks may result in cell death within these progenitor cells, and consequently, heritable mutations will never be detected. Supporting this theory, a previous report demonstrated that growing apices of plants, including the germline cells, are hypersensitive to DNA damage (Fulcher and Sablowski, 2009). This mechanism is thought to safeguard the integrity of their genome in germline cells. Additionally, we observed what appeared to be cell death along the veins of plants infected with TRV carrying *Zif268:FokI* but not in plants with TRV carrying *Zif268:FokIΔ.* This possible cell-death may be due to the on- and off-target double-strand breaks caused by Zif268*:Fok*I, which would further promote the death of germline cells.

We assessed if TRV can spread into germline cells and deliver zinc-finger nucleases for creating heritable mutations. However, we need to ensure that we are not capturing mutagenic events that result due to the direct delivery of the RNA 2 genome to germline cells. To control for this possibility, we could generate another set of TRV vectors that harbor an inactivating mutation within the movement protein gene encoded on RNA 1. This will prevent the virus from moving systemically throughout the plants, however, we would still capture mutagenic events if RNA 2 was directly delivered to germline cells.

To better determine if TRV can facilitate targeted mutagenesis of germline cells, we could take alternative approaches for assessing heritable genome modifications. For example, to determine if TRV enters *Arabidopsis* germline cells, we could engineer RNA 2 to deliver an effector protein that will permanently turn on, or shut off a reporter gene. This would circumvent the need to generate a possibly-lethal double-strand break. This protein, for example, could be a transposase that removes a negative-selection marker that is within the *Arabidopsis* genome (Nishizawa-Yokoi et al., 2014). Additionally, we could use TRV to deliver a different zinc-finger nuclease. For example, if off-target Zif268:*Fok*I-induced double-strand breaks are responsible for killing germline cells, we could deliver a different nuclease with less off-target activity. There are many well-characterized zinc-finger pairs that could be applied, including nucleases targeting *ADH1* and *TT4* (Zhang et al., 2010). Here, a single zinc-finger monomer from these pairs could be inserted into the RNA 2 genome and the corresponding DNA target site (inverted homodimeric sequence) could then be integrated into the plants genome.

Furthermore, it may be beneficial to develop a more sensitive assay for isolating next-generation seedlings with mutations. In this study, we used a single-strand annealing reporter gene where the reconstitution of gene expression requires repair by homologous recombination. In plant cells, repair of double-strand breaks by single-strand annealing is not as frequent as other repair pathways, including non-homologous end joining (Siebert and Puchta, 2002). By changing our single-stranded annealing reporter gene to one that requires non-homologous end joining for the reconstitution of gene expression, we may capture a higher number of events, thereby improving the likelihood of identifying seedlings with heritable mutations.

Taken together, results from this study expand the utility of TRV vectors to the delivery of zinc-finger nuclease monomers in *Arabidopsis thaliana*.

MATERIALS AND METHODS

Plasmid construction: Binary T-DNA plasmids containing RNA 1 (pYL192) and RNA 2 (pYL156) were obtained as a kind gift from Dr. Dinesh Kumar (Caplan and Dinesh-Kumar, 2006). To modify

pYL156 for expression of heterologous proteins, the pea-early browning subgenomic promoter (accession number X78455; nucleotides 323-509) was synthesized on several oligonucleotides (Integrated DNA Technologies), fused in a PCR reaction, and cloned into TRV2 following the coat protein gene (pTRV2-sgP). To generate pTRV2-sgP:GFP, GFP coding sequence was amplified by PCR from pTC23 (contains mGFP) (Haseloff et al., 1997) using primers: 5'-CTCCATGG**GGATCC**ATGAAGACTAATCTTTTTCTCTTTC and 5'-ACATG**CCCGGG**TTAAAGCTCATCATGTTTGTATAG. The bold letters indicate the *Bam*HI and *Xma*I restriction enzyme sites for cloning into pTRV2-sgP. To generate pTRV2-sgP:Zif268:T2A:GFP, Zif268:FokI coding sequence was amplified by PCR from pDW1345 (Wright et al., 2005), GFP was amplified by PCR from pTRV2-sgP:GFP and T2A was amplified by PCR from pZHY013 (Zhang et al., 2013). These three fragments were fused together in another PCR (the three fragments contained ~20 nt of homology to the neighboring fragment). This fused sequence was cloned into pTRV2-sgP using the *Xba*I and *Xma*I restriction sites.

***Agrobacterium* transformation:** RNA 1 (pYL192) and RNA 2 (pTRV2-sgP:GFP; pTRV2-sgP:Zif268:FokI:T2A:GFP; pTRV2-sgP:Zif268:FokIΔ:T2A:GFP) T-DNA vectors were transformed into *Agrobacterium tumefaciens* GV3101 by the freeze-thaw method. Transformed cells were plated on Luria Broth media containing 50 mg/L of kanamycin and 50 mg/L of gentamycin. To prepare *Agrobacterium* for infiltration into *Arabidopsis* plants, single colonies were used to inoculate a 2 mL liquid Luria Broth containing 50 mg/L of kanamycin and 50 mg/L of gentamycin. Cultures were grown for ~16 h in a 28℃ shaker. The next day, approximately 200 μl of culture was transferred to 50 mL of liquid Luria Broth containing 10 mM MES, 20 μM acetosyringone, 50 mg/L kanamycin and 50 mg/L gentamycin. Following an overnight incubation in a 28℃ shaker, Agrobacterium cells were pelleted by centrifugation at 5000 g for 10 minutes and then resuspended in infiltration media (10 mM $MgCl_2$, 10 mM MES, 200 μM acetosyringone) to a final OD_{600} of 1.5. The resuspended culture was then incubated at room temperature for 3-4 hours. Immediately before infiltration, *Agrobacterium* samples containing pYL192 were mixed in a 1:1 ratio with *Agrobacterium* containing the TRV2 plasmid of interest.

Plant transformation: *Arabidopsis* plants, approximately 2-3 weeks old (3-7 leaf stage), were syringe infiltrated with *Agrobacterium* cultures. A needleless 1 mL syringe was used to facilitate infiltration of entire rosette leaves. Two rosette leaves were infiltrated per plant. Immediately following infiltration, plants were covered with a clear dome and placed in a chamber at 22℃ with a 16/8-h photoperiod.

PCR detection of mutations at the Zif268 binding site: To detect Zif268:*Fok*I-induced non-homologous end joining mutations at the GU.US target, we adopted a previously described PCR-digest approach (Zhang et al. 2010). Total DNA was extracted from a single, upper rosette leaf on

infected or non-infected plants. The Zif268 target site was then amplified by PCR using primers 5' TAACTATAACGGTCCTAAGGTAG and 5' AGAAATCATGGAAGTAAGACTGC. Between the Zif268 binding sites is an *Mse*I site. Incubation of wild type amplicons with *Mse*I will result in the cleavage of the 193 bp product to a 129 and 64 bp product. Amplicons containing NHEJ mutations will not be cleaved by MseI, and, therefore, can then be separated from the cleaved amplicons by gel electrophoresis.

To detect the reconstituted GUS gene, genomic DNA from seedlings (rosette leaves) was isolated. The GUS gene was amplified using primers that are homologous to sequence immediately flanking the direct repeats (5' CAAAATTTGTTGATGTGCAGGTA and 5' AATCGGCTGATGCAGTTTCTCCTG). The predicted amplicon sizes were 2,316 bp (GU.US) and 1,134 bp (GUS).

GUS assay: To test for GUS expression, *Arabidopsis* seedlings (~2 weeks old) grown on MS agar plates were stained in a solution of X-Gluc (50 mM phosphate buffer, 10 mM EDTA, 0.1% Triton X-100, 1 mM ferricyanide, 1 mM ferrocyanide, 1mM X-Gluc). Briefly, seedlings were submersed in X-Gluc solution, vacuum infiltrated and then incubated at 37℃ for ~24 h. Following staining, chlorophyll was removed by submerging seedlings in 80% ethanol.

Table 3-1 Infection frequencies of different TRV vectors in *Arabidopsis*

Vector	Total Plants	GFP + Plants	Infection Frequency
pYL192 + TRV2-sgP:GFP	62	14	22.6
pYL192 + TRV2-sgP:Zif268:FokI:T2A:GFP	168	18	10.7
pYL192 + TRV2-sgP:Zif268:FokIΔ:T2A:GFP	116	6	5.2
Total	**346**	**38**	**11.0**

Table 3-2: Summary of somatic and germline mutations in Arabidopsis plants infected with TRV

	GFP+ Plant			Next Generation Seedlings			
TRV2 Vector	**Plant ID**	**Phenotype**	**NHEJ**	**Seeds screened**	**GUS +**	**NHEJ PCR**	**SSA PCR**
TRV2-sgP:Zif268:FokI:T2A:GFP	1.1	++++	**D**	1120	4	0/120	0/120
	1.2	++++	**ND**	600	0		
	1.3	++++	**ND**	300	0		
	1.4	++++	**ND**	800	0		
	2.1	++	**ND**	200	0		
	2.2	++	**D**	200	0	0/47	0/47
	2.3	++	**ND**	200	0		
	3.1	++	N/A	50	0		
	3.2	+++	N/A	1300	0		
	3.3	++++	**ND**	1250	0*		
	3.4	++	**D**	200	1		
	3.5	++	**ND**	100	0		
	3.6	+++	**ND**	1100	0		
	3.7	+++	**ND**	950	0		
	3.8	+	N/A	0	N/A		
	3.9	+	N/A	0	N/A		
	3.10	++++	N/A	930	0		
	3.11	++	N/A	20	0		
TRV2-sgP:Zif268:FokIΔ :T2A:GFP	1.1	++++	N/A	600	0		
	1.2	++++	N/A	800	0		
	1.3	++++	N/A	400	0		
	1.4	++++	N/A	500	0		
	3.1	++++	**ND**	680	0*		
	3.2	++++	**ND**	550	0		
TRV2-sgP:GFP	3.1	++++	ND	800	0		
	3.2	++++	N/A	800	0		
	3.3	++++	N/A	600	0		
	3.4	++++	N/A	610	0		
Non-Infiltrated Control				10500	0		

D, detected a cleavage-resistant product; ND, did not detect a cleavage-resistant product; N/A, did not test. ++++, wild type phenotype; +++, mild phenotype: slightly stunted growth; ++, moderate phenotype: plants were noticeably smaller than controls, produced lower numbers of seeds; +, severe phenotype: stunted growth, no seeds produced., found two seedlings with blue roots. Plants 3.8 and 3.9 did not produce seed.*

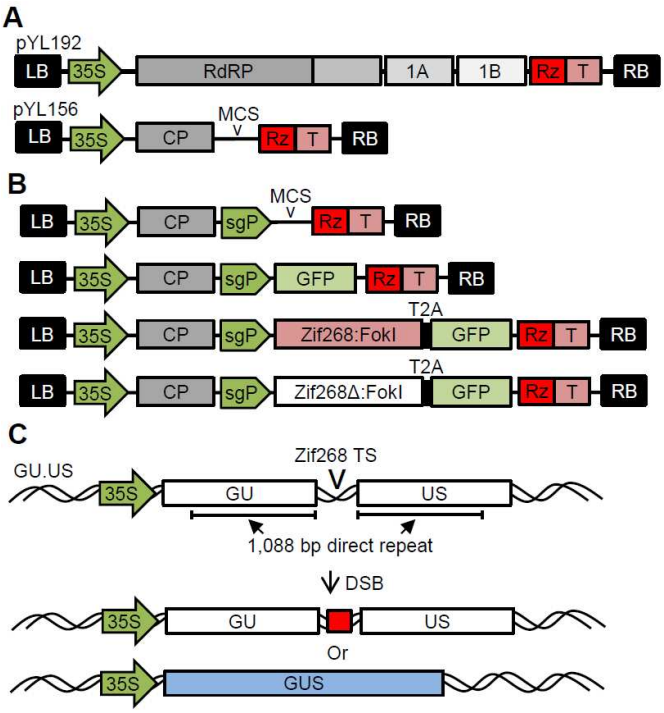

Figure 3-1. TRV T-DNA vectors and the target for modification in *Arabidopsis*. A. Illustration of the pYL192 and pYL156 TRV T-DNA plasmids. 35S, duplicated 35S promoter from cauliflower mosaic virus; LB, left T-DNA border; RB, right T-DNA border; RdRP, RNA-dependent RNA polymerase; CP, coat protein; Rz, rybozyme; T, nopaline synthase terminator. **B.** Illustration of the TRV2 vectors used in this study. Names of the vectors from top to bottom: pTRV2-sgP; pTRV2-sgP:GFP; pTRV2-sgP:Zif268:FokI:T2A:GFP; pTRV2-sgP:Zif268:FokIΔ:T2A:GFP. Δ, mutation in *Fok*I (D450A) that causes loss of nuclease activity; sgP, subgenomic promoter; MCS, multiple cloning site; GFP, green fluorescent protein; T2A, ribosomal skipping sequence. **C**. Illustration of the GU.US reporter that is stably integrated within the *Arabidopsis* genome (top). Targeted double-strand breaks can be repaired imperfectly by the NHEJ pathway (middle) or by the single-strand annealing pathway (bottom). TS, target site; DSB, double-strand break.

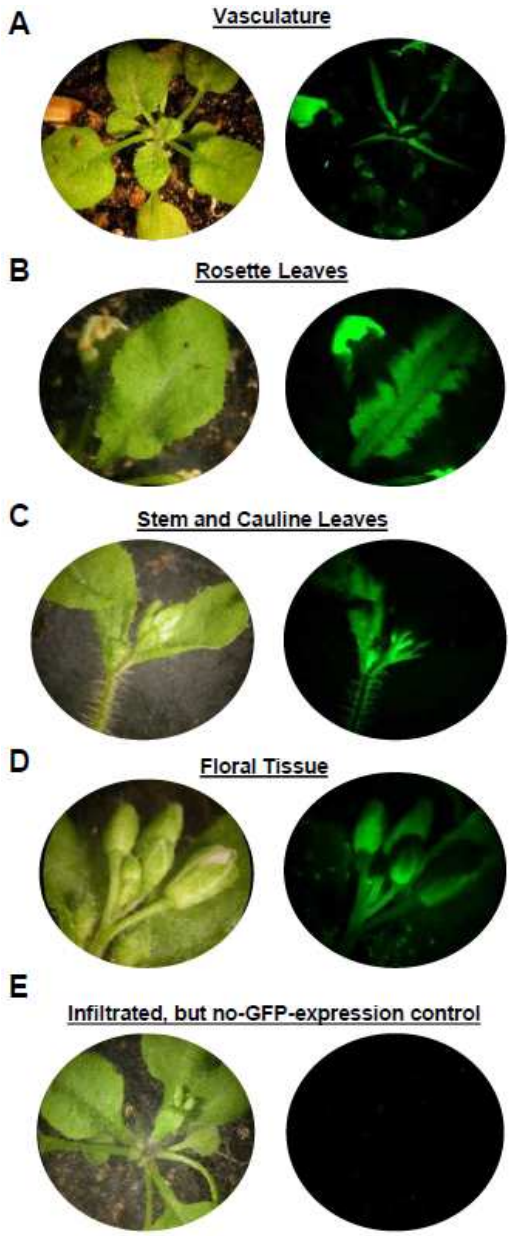

Figure 3-2. Movement of TRV within *Arabidopsis* plants. Plants were infiltrated with *Agrobacterium* containing pYL192 and pTRV2-sgP:GFP. **A.** GFP expression within vasculature cells. GFP image brightened 50%; two second exposure. **B.** GFP expression within vasculature cells and rosette leaves. GFP image brightened 50%; two second exposure. **C.** GFP expression in shoot cells and cauline leaves. GFP image brightened 40%; one second exposure. **D.** GFP expression in floral tissue. GFP image brightened 40%; one second exposure. **E.** Representative image of a plant that was infiltrated with Agrobacterium containing TRV vectors, but did not express GFP. Fluorescent image brightened 50%; one second exposure.

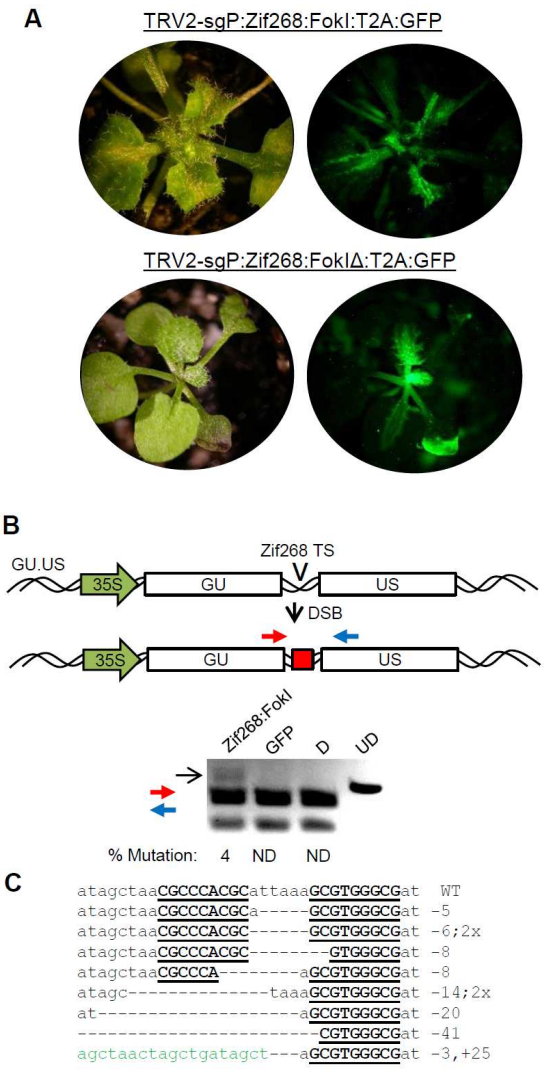

Figure 3-3. Targeted mutagenesis in somatic cells by TRV-mediated delivery of Zif268:FokI. **A**, GFP expression in plants infiltrated with a mixture of Agrobacterium containing pTRV1 and TRV2-sgP:Zif268:FokI:T2A:GFP (top) or pTRV1 and TRV2-sgP:Zif268:FokIΔ:T2A:GFP. GFP images were brightened 50% (top) and 40% (bottom); exposure times were one second (top) and four seconds (bottom). **B**, Illustration of the primers used to PCR-amplify the Zif268 target site (top). The resulting amplicons were digested with *Mse*I (there is an *Mse*I site within the Zif268 spacer) and separated on an agarose gel (bottom) The Zif268:FokI lane represents data from plant 2.2 (Table 3-2). D, digested with MseI; UD, undigested; GFP, refers to a GFP-expressing plant that was infected with TRV1 and TRV2-sgP:GFP; ND, not detected. **C**. Sequences of amplicons containing targeted mutations. Sequences were obtained from cleavage-resistant amplicons from plant 1.1. Upper case, bold and underlined letters indicate the Zif268 binding sequences. Green letters indicate insertion sequence.

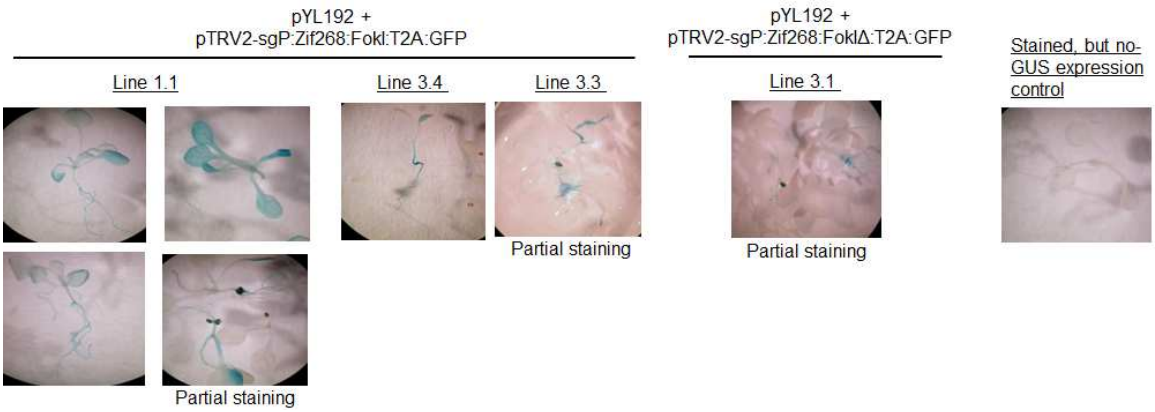

Figure 3-4. Examples of GUS activity in seedlings from TRV-infected parent plants. Control is an image of a seedling from line 1.1 (pYL192 + pTRV2-sgP:Zif268:FokI:T2A:GFP) that did not stain blue after incubation in a solution with the GUS substrate.

CHAPTER 4

TARGETED DELETION AND INVERSION OF TANDEMLY ARRAYED GENES IN *ARABIDOPSIS* THALIANA USING ZINC FINGER NUCLEASES

ABRIDGEMENT

Tandemly arrayed genes (TAGs) or gene clusters are prevalent in higher eukaryotic genomes. For example, approximately 17% of genes are organized in tandem in the model plant *Arabidopsis* thaliana. The genetic redundancy created by TAGs presents a challenge for reverse genetics. As molecular scissors, engineered zinc finger nucleases (ZFNs) make DNA double strand breaks (DSBs) in a sequence specific manner. ZFNs thus provide a means to delete TAGs by creating two DSBs in the gene cluster. Using engineered ZFNs, we successfully targeted seven genes from three TAGs on two *Arabidopsis* chromosomes, including the well-known *RPP4* gene cluster, which contains eight resistance (R) genes. The resulting gene cluster deletions ranged from a few kb to 55 kb with frequencies approximating 1% in somatic cells. We also obtained large chromosomal deletions of ~9 Mb at approximately one tenth the frequency, and gene cluster inversions and duplications were also achieved. This study demonstrates the ability to use sequence specific nucleases in plants to make targeted chromosome rearrangements and create novel chimeric genes for reverse genetics and biotechnology.

INTRODUCTION

Genome sequences of many plant species have been completed, and all contain a significant number of tandemly arrayed genes (TAGs). For example, approximately 14% of rice genes (Matsumoto et al., 2005), 17% of *Arabidopsis* genes (The Arabidopsis Genome Initiative, 2000), and 35% of maize genes (Messing et al., 2004) are organized in tandem. The genetic redundancy resulting from TAGs presents a challenge for reverse genetics (Jander and Barth, 2007). It is difficult, if not impossible, to eliminate TAG expression using methods such as EMS mutagenesis and TILLING (Koornneef et al., 1982; McCallum et al., 2000), T-DNA and transposon insertional mutagenesis (Alonso et al., 2003; Raina et al., 2002; Rosso et al., 2003; Sessions et al., 2002; Woody et al., 2007), or RNA interference and miRNA-based gene silencing (Abbott et al., 2002; Alvarez et al., 2006; Hamilton and Baulcombe, 1999). One promising approach for studying TAGs is to make chromosomal deletions. Ionizing radiation, however, acts randomly (Li et al., 2001), making it difficult to recover the desired deletion. Although the Cre-Lox system has proven effective for making deletions, it relies on large LoxP T-DNA insertion populations (Zhang et al., 2003), which are currently unavailable for most plant species.

An alternative approach to make targeted genome deletions is to use sequence specific nucleases. These proteins – which include zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and meganucleases – make site-specific DNA double strand breaks (DSBs) at a locus of interest (Christian et al., 2010; Smith et al., 2006; Kim et al., 1996). Repair of DSBs occurs by two pathways, namely non-homologous end-joining (NHEJ) and homologous recombination (HR) (Puchta, 2005; Puchta et al., 1996). NHEJ is error-prone and

typically leads to insertions, deletions (indels) and substitutions at the cleavage site. In contrast, repair by HR is typically error-free because it uses a DNA template to correct the break. Of the three nuclease platforms, ZFNs have been most widely used in plants. ZFNs have been successfully used for targeted mutagenesis by NHEJ in *Arabidopsis* (Osakabe et al., 2010; Zhang et al., 2010) and soybean (Curtin et al., 2011), as well as for gene targeting by HR in tobacco (Townsend et al., 2009) and maize (Shukla et al., 2009). In addition, Petolino et al. reported that when a 4.3 kb GUS transgene was flanked by two ZFN sites, it could be efficiently deleted from the tobacco genome (Petolino et al., 2010), thus demonstrating that ZFNs can induce chromosomal deletions of transgenes in plants. Although ZFN-mediated deletion, inversion and duplication of endogenous chromosomal DNAs has been achieved in human cells (Lee et al., 2010; Lei et al., 2011), none of these chromosome rearrangements have yet to be demonstrated in plant cells using sequence specific nucleases.

In *Arabidopsis*, the receptor-like kinase (*RLK*) and the nucleotide-binding and leucine rich repeat (NB-LRR) resistance (R) gene families are large and have ~600 and ~150 gene members, respectively (Meyers et al., 2003; Shiu et al., 2004). Both gene families play important roles in plant development and immunity. For example, many plant hormone receptors and almost all plant immune receptors are members of these two families. Genes in both families are organized in tandem throughout the genome. In this study, we sought to delete endogenous TAGs using ZFNs that target three *RLK* gene clusters and one large R gene cluster. We successfully demonstrated targeted deletions, inversions and duplications of multiple gene clusters as well as large chromosomal deletions exceeding 9 Mb.

RESULTS

Strategy for targeted deletion of TAGs

Chromosomal deletions can be stimulated by two coordinated DSBs (Petolino et al., 2010). In our strategy, we used a single pair of ZFNs, which due to the high sequence similarity among TAGs, created two or more DSBs. Five ZFNs were engineered using the CoDA method to target three *RLK* gene clusters, the *RPP4* R gene cluster and the ASK8 gene cluster (Figure 4-1 and Figure 4-2). These five ZFNs were expected to induce a total of thirteen DNA DSBs in targeted exons of the *Arabidopsis* genome. The resulting size of predicted deletions ranged from a few kb to over 50 kb. In addition to inducing deletions, ligation of broken chromosomes can create novel gene fusions. However, because NHEJ is error-prone we anticipated that indels would be introduced at the cleavage site, thus rendering some of the deleted TAGs non-functional.

Detection of ZFN-induced mutagenesis

For all ZFNs used in this study, both the left and right ZFNs were expressed from an estradiol inducible promoter and separated by a "self-cleaving" T2A peptide, which promotes production of two proteins from one mRNA by a ribosomal skipping mechanism (Szymczak et al., 2004). T1 transgenic plants were screened on MS medium containing both hygromycin (to identify transgenic plants) and β-estradiol, which allowed for induction of ZFN transgenes. Detection of ZFN-induced mutations was carried out using enrichment PCR (Qi et al., 2013b). Indels introduced by ZFNs frequently occur in the "spacer" region where the FokI nuclease domains dimerize and cleave the DNA. Thus, if there is a unique restriction enzyme site in the spacer, this restriction enzyme site will likely be destroyed through ZFN-mediated mutagenesis and thus render the DNA uncuttable by the restriction enzyme. For each locus evaluated, T1 plants in which ZFNs were induced by estradiol were pooled together and genomic DNA was extracted. For the At1g53-ZFN, which targets the At1g53430-At1g53440 gene cluster, ZFN activity was detected at the At1g53430 target site (Figure 4-3A). Similarly, the At1g70-ZFN and At4g16-ZFN were found to be active at the At1g70450-At1g70460 gene cluster and the *RPP4* gene cluster, respectively (Figure 4-3B and C). We did not detect ZFN activity for the At3g21-ZFN and the At5g01-ZFN (Figure 4-2, data not shown). These data suggest that three of the five ZFN pairs are functional, and their activity can be temporally controlled by the estradiol-inducible promoter (Figure 4-3). We thus focused on the three active ZFNs for further study.

Quantification of ZFN activity at seven endogenous loci

We screened multiple T2 populations of *Arabidopsis* plants transformed with each ZFN pair and selected two independent lines showing high ZFN activity. These transgenic plants were used to quantify ZFN activity based on the level of NHEJ mutagenesis in whole seedlings. Note that the frequency of mutation observed by enrichment PCR is influenced by the position of the restriction enzyme site. The AflII and BfaI sites are very close to the middle of the spacer in the At1g70-ZFN and At4g16- ZFN recognition sites, respectively. In these cases, we reasoned the percentage of uncut DNA by these restriction enzymes should approximate the ZFN-mediated mutagenesis frequency. On the other hand, the DdeI site is farther away from the spacer in the At1g53-ZFN target site. Since only a fraction of mutagenesis events are likely assessed by measuring loss of the DdeI site, the T7 endonuclease assay was instead used to measure activity of the At1g53-ZFN. Activity of the three ZFNs at seven endogenous loci ranged from a few percent to 20% (Figure 4-4A-C), with the At4g16-ZFN having the highest activity (13.4% to 20.1%). Differences observed in ZFN activity among different transgenic plants is likely due to different expression levels of the ZFN transgenes, depending on the chromosomal site of T-DNA integration.

To further confirm activity of each ZFN, DNA fragments resistant to restriction enzyme digestion were cloned and sequenced (Figure 4-4A and B). The sequencing results revealed

ZFN-induced mutations at all seven target loci, with small deletions (1-10 bp) being the most prevalent (Figure 4-4D-J). Since the DdeI restriction site is farther away from the spacer, deletions recovered at At1g53430 and At1g53440 were typically larger (24-68 bp) (Figure 4-4I and J).

Deletion of TAGs by ZFNs

The observed mutations at all seven target sites indicate that ZFN- induced DSBs were generated. To examine deletion of TAGs as an outcome to ZFN activity, we conducted PCR using specific primers flanking all ZFN target sites, such that the production of PCR products would indicate the presence of deletions (Figure 4-5A-C). With this strategy, we detected deletions of the At1g53430 gene cluster (Figure 4-5A) and the At1g70450 gene cluster (Figure 4-5B), as well as three different types of deletions at the *RPP4* gene cluster (Figure 4-5C). We further confirmed these deletions by cloning the deletion-specific PCR products and sequencing randomly selected clones (Figure 4-5 A-C, lower panels). Not only did we observe indels at the site of deletions, we also observed perfect ligations of two sticky ends derived from ZFN-cleaved DNA (Figure 4-5 A-C). These events are most likely due to the presence of compatible overhangs. Taken together, these data suggest that deletions of up to 55 kb can be made by ZFNs on different *Arabidopsis* chromosomes.

Large chromosomal deletions by ZFNs

Since we achieved deletions of TAGs spanning up to 55 kb, we next tested whether much larger chromosomal deletions could be generated. To do so, we took advantage of an existing ZFN (ADH1-ZFN), which we previously engineered to target the ADH1 gene (At1g77120) at the end of chromosome 1 (Zhang et al., 2010) (Figure 4-6A and Figure 4-7A). Simultaneous expression of At1g53-ZFN and ADH1-ZFN could potentially result in chromosomal deletions as large as 9 Mb – almost one third the length of *Arabidopsis* chromosome 1 (Figure 4-6A). We first screened independent estradiol-inducible ADH1-ZFN lines and identified the ADH1-ZFN #3 line as showing strong estradiol-inducible mutagenesis activity (Figure 4-7B). We also mapped the ADH-ZFN transgene insertion site in this line to chromosome 2 to aid in genotyping (Figure 4-7C). We then obtained a homozygous T3 ADH1-ZFN #3 line for use in our experiments.

To test for chromosomal deletions, we crossed the homozygous At1g53-ZFN #7 T2 line and the ADH1-ZFN #3 T3 line. F1 plants were obtained, and both ZFNs were induced by growing them on estradiol-containing MS medium. As with the gene cluster deletions, PCR was used to detect the large chromosomal deletions, and deletions were detected for both At1g53430-At1g77120 and At1g53440-At1g77120 (Figure 4-6B). Resulting PCR products were cloned and sequenced (Figure 4-6C and D). Interestingly, the most prevalent product was a ligation of a 7 bp spacer sequence. We predict this product resulted from microhomology-based NHEJ using only 1 bp of microhomology (Figure 4-8). The resulting 7-bp spacer would be expected to be cut

inefficiently by the hybrid ZFN, which contains At1g53-ZFN-left and ADH1-ZFN-right monomers (Figure 4-8). This may explain the PCR product's predominance.

Frequency of ZFN-induced chromosomal deletions

Having demonstrated ZFN-induced chromosomal deletions ranging from a few kb to 9 Mb, we next sought to estimate their frequency of occurrence. We adapted a digital PCR method used to measure deletion frequencies in human cells (Lee et al., 2010). In our case, deletion frequency was detected by PCR using a series of genomic DNA dilutions as templates. Amplification of ADH1 was used as an internal control for DNA copy number. As summarized in Table 4-1, the frequency of gene cluster deletions was approximately 1% -- positively correlated with ZFN activity and negatively correlated with the length of the deletions. As the length of deletion increased to 9 Mb, the frequency dropped to less than 0.1%. Targeted inversions of gene clusters: In addition to deletions, chromosomal DNA released by two DSBs could create inversions (illustrated in Figure 4-9A). Such inversions will have two novel junction sites, which can be detected with PCR using specific primer sets (Figure 4-9A). The predicted novel junctions were, in fact, detected at the At1g53430 gene cluster by PCR; DNA sequencing confirmed that the inversions occurred. (Figure 4-9B). As anticipated, many inversion junctions had small deletions indicative of imprecise NHEJ (Figure 4-9C). We could also detect and confirm inversions at the At1g70450 gene cluster (Figure 4-10), and these occurred at a frequency of about 0.05% as measured by digital PCR. Gene cluster inversions, therefore, appear to occur at a lower frequency than deletions.

Possible targeted duplication of gene clusters

When two DSBs are induced in different TAGs on different chromosomes (either homologous chromosomes or sister chromatids), interchromosomal ligations of the broken DNA ends can result in deletions as well as gene cluster duplications (Figure 4-3A). For TAGs that only contain two genes, such as the At1g70450 gene cluster, the duplication will create a hybrid gene (Figure 4-11A). We indeed detected specific products indicative of ZFN-induced duplications of the At1g70450 gene cluster (Figure 4-11C), and these were also confirmed by DNA sequencing of the cloned PCR products (Figure 4-11D). However, if deleted DNA was circularized and retained in cells, it would give the same PCR products as depicted in Figure 4-11B. Since we cannot distinguish between these two outcomes, and since we do not know how long DNA circles persist in cells, it remains unclear how many of the events detected (Figure S4-11 C and D) truly reflect gene cluster duplications.

DISCUSSION

We previously reported engineering active ZFNs for six endogenous loci in *Arabidopsis* using the CoDA method (Sander et al., 2011). In this study, we successfully targeted seven additional loci with three ZFNs, each of which showed NHEJ mutagenesis activity. Two other ZFNs targeting different endogenous loci failed to show detectable activity in *Arabidopsis*. For these ZFNs, it remains possible that chromatin structure or epigenetic modifications impeded ZFN access to endogenous DNA targets. Our overall 60% success rate is in line with the success rate previously reported for CoDA, and our work further demonstrates the usefulness of the CoDA method for engineering active ZFNs.

We aimed to create deletions of TAGs in the *Arabidopsis* genome using a single ZFN pair to target multiple genes within the cluster. However, not every gene cluster can be deleted with such a strategy. In cases where the TAGs differ significantly in DNA sequence similarity, two different pairs of ZFNs might be required to create the desired deletion (Sollu et al., 2010). In addition, TALENs are good alternatives to ZFNs, since there seems to be less restriction in designing TALENs to target diverse DNA sequences (Bogdanove and Voytas, 2011; Doyle et al., 2012; Reyon et al., 2012). The scale of deletions that we obtained (from a few kb to 9 Mb) suggests ZFNs are capable of creating very large deletions in plants. We recognize, however, that *Arabidopsis* plants are not likely to survive the loss of megabase pairs of chromosomal DNA. Targeted insertions created by HR were previously demonstrated using ZFNs in plants (Cai et al., 2009). Generating deletions (such as in this study) and concomitantly creating insertions by HR may be another genome engineering approach of value for basic research and crop improvement.

Being highly homologous and repetitive, TAGs are particularly prone to change either through unequal crossover or gene conversion. This provides the opportunity to evolve new gene functions, some of which may be adaptive (Hanada et al., 2008; Lehti-Shiu et al., 2009). Our approach for making deletions and duplications in TAGs has demonstrated that we can now create novel chimeric genes which may otherwise not occur naturally. For example, we frequently detected hybrid genes due to perfect ligation of broken chromosomes after loss of the intervening DNA (Figure 3A-C). In addition, duplication of gene clusters increases genetic redundancy and frees some gene members to evolve new functions. Tandem duplication in *Arabidopsis* has provided a means for adaptive evolution of R genes (Meyers et al., 2005). Our approach in creating rearrangements in the complex *RPP4* gene cluster and the two *RLK* gene clusters is thus a promising first step toward generating valuable genetic material for both molecular and evolutionary studies.

Gene cluster inversion is another consequence of our TAG-targeting approach. For TAGs encoding two genes, DNA inversion is likely to destroy the function of both gene targets. However, since inversions preserve DNA sequences (which would otherwise be lost in deletions),

they may have unique applications, such as serving as templates for future genome evolution. Another application of inversions is that they may only knock out two target genes at both ends of the TAG while retaining the function of the genes in between. This is only true if there are more than two members in the TAG, such as the *RPP4* gene cluster evaluated in this study.

To recover plants with germline-transmitted deletions, we screened large T3 populations of At1g70-ZFN #2 (a total of 2539 plants) and At4g16-ZFN #12 (a total of 2322 plants) with no success (data not shown). Clearly, the frequencies of germline-transmitted deletions are much lower than the somatic frequencies. This observation is consistent with other ZFN-mediated mutagenesis studies we have conducted (unpublished), where the frequency of somatic mutation did not directly reflect the frequency of germline mutation. Rather, we believe there is a threshold somatic mutation frequency that must be surpassed to ensure successful germline transmission. In previous work, we found that somatic mutagenesis frequencies in excess of 7% were sufficient to recover germinal mutations at high frequency (Zhang et al., 2010). The observed ~1% somatic deletion frequency observed here appears to be under this threshold.

It has been shown that stem cell niches in *Arabidopsis* are hypersensitive to DNA damage, and an even a low dose of DNA damage can trigger programmed cell death (PCD) selectively in these stem cells (Fulcher and Sablowski, 2009). Interestingly, PCD in the shoot meristem was greatly suppressed when the DNA damage early response gene, ATM, was knocked out. Thus, it might be useful to use an atm mutant background to screen for germline deletions in plants. In addition, we recently reported that *smc6b* mutations promotes NHEJ in *Arabidopsis* (Qi et al., 2013b). As chromosomal breaks that lead to deletions are joined through NHEJ, it is likely that smc6b mutations will enhance deletion frequencies. It is also possible some deletions are deleterious to pollen and egg cells, and this impedes their transmission. As whole plant regeneration is routinely performed from somatic cells for many species (e.g. rice, maize and tobacco), plants with deletions may be easier to achieve by regenerating somatic tissues that have been treated with sequence specific nucleases.

MATERIALS AND METHODS

ZFN assembly: Genomic DNA sequences of target genes were analyzed using the software ZiFiT Targeter (version 3.3) to identify ZFN sites for which ZFNs could be engineered using the CoDA method (Curtin et al., 2011; Sander et al., 2011). DNA sequences encoding ZFNs of choice were assembled by mutagenesis and overlapping PCR using standard molecular cloning procedures. For each ZFN, ZF arrays were first cloned into the yeast expression vectors pCP3 and pCP4 using available XbaI and BamHI sites (Zhang et al., 2010). Then, DNA sequences for the left and right ZF arrays were excised from the yeast expression vectors with XbaI and BamHI, and moved into the pZHY013 entry clone using the XbaI-BamHI and NheI-BglII sites,

respectively. pZHY013 contains an obligate FokI heterodimer architecture (Miller et al., 2007), and the ZFNs are linked by a T2A ribosomal skipping sequence. The plant ZFN expression vectors were made using a Gateway® LR reaction between the above mentioned entry clones and the pFZ19 destination vector (Zhang et al., 2010).

Transgenic plants and expression of ZFNs: *Agrobacterium tumefaciens* GV3101/pMP90 was transformed with pFZ19 plasmids containing the ZFNs. The transformed *A. tumefaciens* strain was then used to transform *Arabidopsis* Col-0 (wild type) plants using the floral dip method (Clough and Bent, 1998). T1 transgenic plants were selected by growing the sterilized seeds on 0.5X MS solid medium (0.8% agar) that contained 100 µg/ml timentin (PlantMedia) and 20 µg/ml hygromycin B (Roche). For inducing ZFN expression, 20 µM β-estradiol (Sigma) was included in the medium.

ZFN activity measurement: One-week old seedlings grown on MS medium with estradiol were harvested for DNA extraction using the CTAB DNA isolation method (Stewart and Via, 1993). Eight T2 transgenic plants from the same T1 parent were bulked to represent each sample, whereas eight wild type plants were bulked as the negative control. To detect ZFN activity, an enrichment PCR procedure was used. Briefly, about 500 ng of genomic DNA from each sample was digested overnight (~16 hrs) with 1 µl of DdeI (for At1g53-ZFN), AflII (for At1g70-ZFN), BfaI (for At4g16-ZFN), BmgBI (for At3g21-ZFN) or EarI (for At5g01-ZFN) in a 20 µl reaction volume. Four microliters of digested genomic DNA was used for PCR amplification of the corresponding ZFN target sites in a 25 µl reaction volume. Ten microliters of unpurified PCR product was then digested with 1 µl of the same restriction enzyme in a 40 µl reaction volume for 12-16 hrs. Digested products were resolved by electrophoresis in 1.5 % agarose gels; mutations created by ZFNs were evidenced by undigested PCR product. An alternative method to detect and measure ZFN activity by restriction digestion is similar to enrichment PCR except that the genomic DNA digestion step is omitted. Four microliters of 50 ng/µl genomic DNA was directly used for PCR and subsequent digestion by the corresponding restriction enzyme. The frequency of ZFN-mediated mutagenesis was measured by quantifying the percentage of undigested PCR product that was resolved by electrophoresis on a 1.5% agarose gel. ZFN activity was also measured by the Surveyor assay in which T7 endonuclease was substituted for Cel-I (Guschin et al., 2010). Briefly, PCR products amplified from genomic DNA templates were purified with a QIAquick® PCR purification kit. For each sample, about 500 ng of purified PCR product was mixed with NEB buffer 2 in a 30 µl volume. To promote heteroduplex formation, PCR amplicons were denatured and re-annealed using the following regime: 95℃ fo r 5 min, 95℃ to 85℃ at -1.5℃/sec, 85℃ to 25℃ at -0.1℃/sec. One microliter of T7 endonuclea se (NEB) was added to each sample for digestion at 37℃ for 1 hr. The digested products w ere resolved by electrophoresis in 1.5% agarose gels, and the frequency of ZFN- mediated mutagenesis was quantified as described previously (Guschin et al., 2010).

Sequence confirmation of mutagenesis and deletions: Both undigested PCR products (for assessing mutagenesis) and PCR products (for detecting deletions, inversions and duplications) were purified using a QIAquick® Gel Extraction Kit. The purified DNA products were then cloned using either the pCR®8/GW/TOPO® TA Cloning Kit or the pCR®2.1 Original TA Cloning Kit (Invitrogen). Multiple clones for each experiment were randomly picked and subjected to DNA sequencing.

Deletion frequency measurements: A method similar to digital PCR analysis (Lee et al., 2010) was used to estimate deletion frequencies. Genomic DNA samples were serially diluted (in a 3X gradient) in distilled water, and they were used for PCR using deletion-specific primer pairs (Table 4-2). The same DNA samples were used to amplify a fragment of the ADH1 gene as a genomic DNA copy number control. In each case, the difference of dilution factors for both deletion PCR and control PCR was used to calculate deletion frequency.

Table 4-1 Frequency of ZFN-induced chromosomal deletions

Transgenic Lines	*At1g53430* to *At1g53440* (deletion of 10.6 kb)	*At1g70450* to *At1g70460* (deletion of 4.5 kb)	*At4g16940* to *At4g16860* (deletion of 41.9 kb)	*At4g16960* to *At4g16860* (deletion of 55.3 kb)	*At1g53440* to *At1g77120* (deletion of 9.027 Mb)	*At1g53430* to *At1g77120* (deletion of 9.037 Mb)
A1g53-ZFN #4	0.3%					
At1g53-ZFN #7	1%					
At1g70-ZFN #1		3%				
At1g70-ZFN #2		3%				
At4g16-ZFN #5			1%	1%		
At4g16-ZFN #12			3%	3%		
ADH1-ZFN #3 X					0.046%	0.137%
At1g53-ZFN #7[a]						

ZFN, zinc finger nucleases.
[a] F1 plants.

Table 4-2 Zinc-finger arrays, recognition sites and recognition helices

Zinc finger arrays	Recognition sites and recognition helix amino acid sequences		
	F1	F2	F3
At1g53-ZF_left	GAG	GCT	GTG
	KHSNLTR	QRSDLTR	RPDALPR
At1g53-ZF_right	GTA	GCT	TAA
	QQSSLLR	QRSDLTR	QRGNLNM
At1g70-ZF_left	GCT	GCT	TAA
	MKNTLTR	QRSDLTR	QRGNLNM
At1g70-ZF_right	GAC	GCG	GTA
	DPSNLIR	RTDTLAR	QGGALQR
At4g16-ZF_left	GAA	GAA	GAA
	QASNLTR	QQTNLTR	QTNNLNR
At4g16-ZF_right	GGA	GCC	GTA
	DNAHLAR	DSSVLRR	QSTSLQR
At3g21-ZF_left	TGT	GCT	GGT
	KRQHLEY	QRSDLTR	HGHRLKT
At3g21-ZF_right	GCT	GCC	GAT
	LRTSLVR	DSSVLRR	LSTNLTR
At5g01-ZF_left	GGA	GGC	GGA
	RPSKLVL	LKEHLTR	QSQHLVR
At5g01-ZF_right	GAA	GAA	GGC
	QASNLTR	QQTNLTR	KNVSLTH

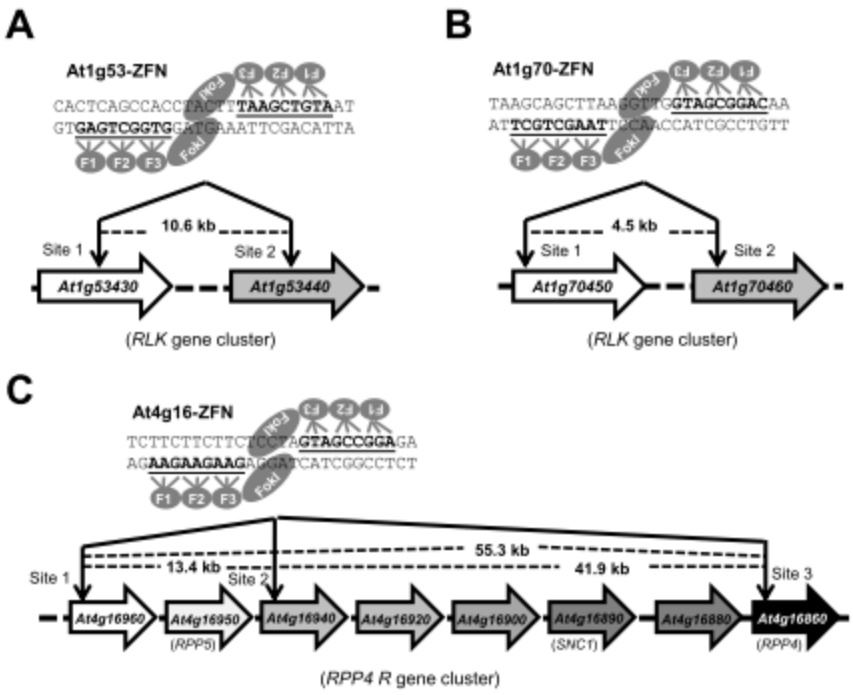

Figure 4-1. Schematic of target genes and ZFN sites. (A) The At1g53-ZFN targets both At1g53430 and At1g53440 in the 14th exon of each gene. (B) The At1g70-ZFN targets At1g70450 in the 1st exon and At70460 in the 2nd exon. (C) The At4g16-ZFN targets three sites in the *RPP4* R gene cluster: the 1st exon of At4g16960, the 5' UTR of At4g16940, and the 1st exon of At4g16860. Cartoons illustrating ZFN pairs depict the DNA recognition triplets for each zinc finger. The zinc finger binding sequences are underlined and the distance between cleavage sites is shown.

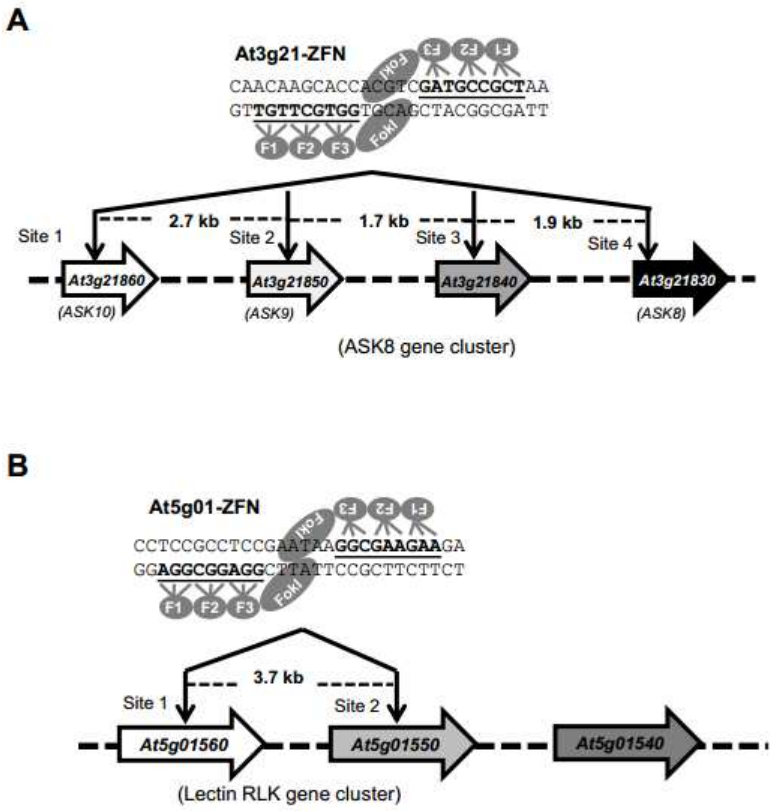

Figure 4-2. Supplemental Figure; ZFNs that target the ASK8 gene cluster and a lectin *RLK* gene cluster. (A) The At3g21-ZFN targets all four members of the ASK8 gene cluster. (B) The At5g01-ZFN targets two genes in a lectin *RLK* gene cluster. Cartoons illustrate the ZFN pairs, and the DNA recognition triplets are indicated. The zinc finger binding sequences are underlined and the distance between cleavage sites is shown.

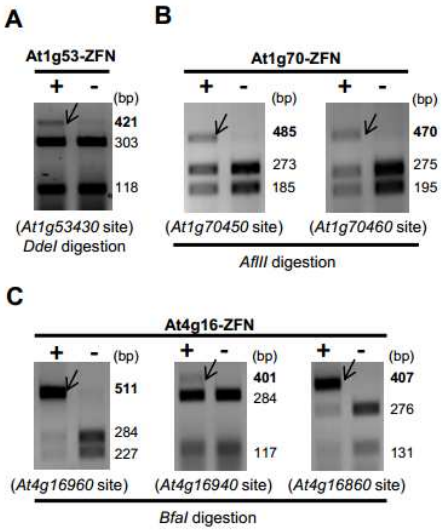

Figure 4-3. Supplemental Figure; CoDA-assembled ZFNs are active in T1 plants. (A) At1g53-ZFN's activity is detected at At1g53430. (B) At1g70-ZFN's activity is detected at both At1g70450and At1g70460. (C) At4g16-ZFN's activity is detected at At4g16960, At4g16940and At4g16860. Activity of ZFNs was measured by enrichment PCR using the restriction enzymes shown in each panel. The uncut bands represent ZFN-induced mutations and are indicated by arrows. Bulked estradiol-treated T1 transgenic plants or wild type plants were compared.

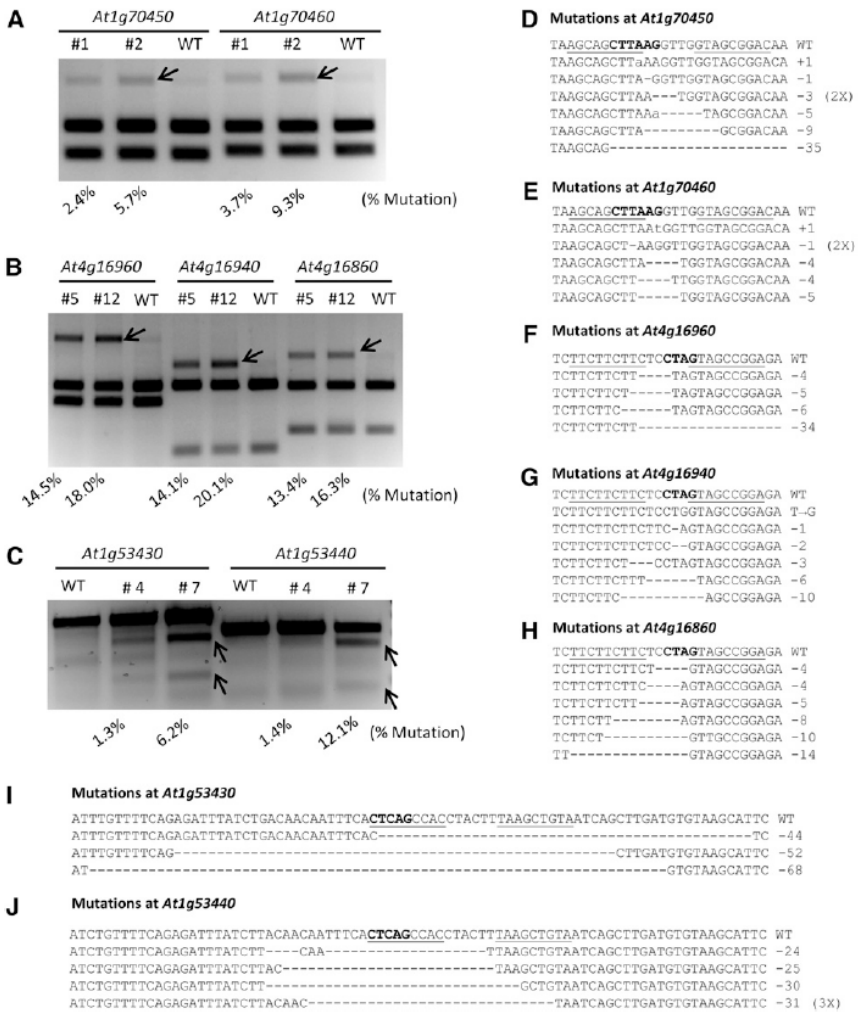

Figure 4-4. ZFN activity at seven endogenous loci. (A) At1g70-ZFN activity at target sites in At1g70450 and At1g70460. PCR products were digested with AflII. (B) At4g16-ZFN activity at target sites in At4g16960, At4g16940 and At4g16860. PCR products were digested with BfaI. (C) At1g53-ZFN activity at target sites in At1g53430 and At1g53440 as measured by the T7 nuclease assay. For both restriction digestion assays (panels A and B) and the T7 nuclease assay (panel C), mutagenesis frequencies (shown at the bottom of the figures) were determined by measuring the signal intensity of each band using the Labworks analysis software. (D-J) ZFN-induced mutations at seven endogenous target sites with different mutation types indicated. The restriction enzyme sites used for activity measurement are marked in bold letters. The uncut PCR products (from panels A and B, and Figure S2) were cloned and sequenced to reveal ZFN-induced mutations at each target site.

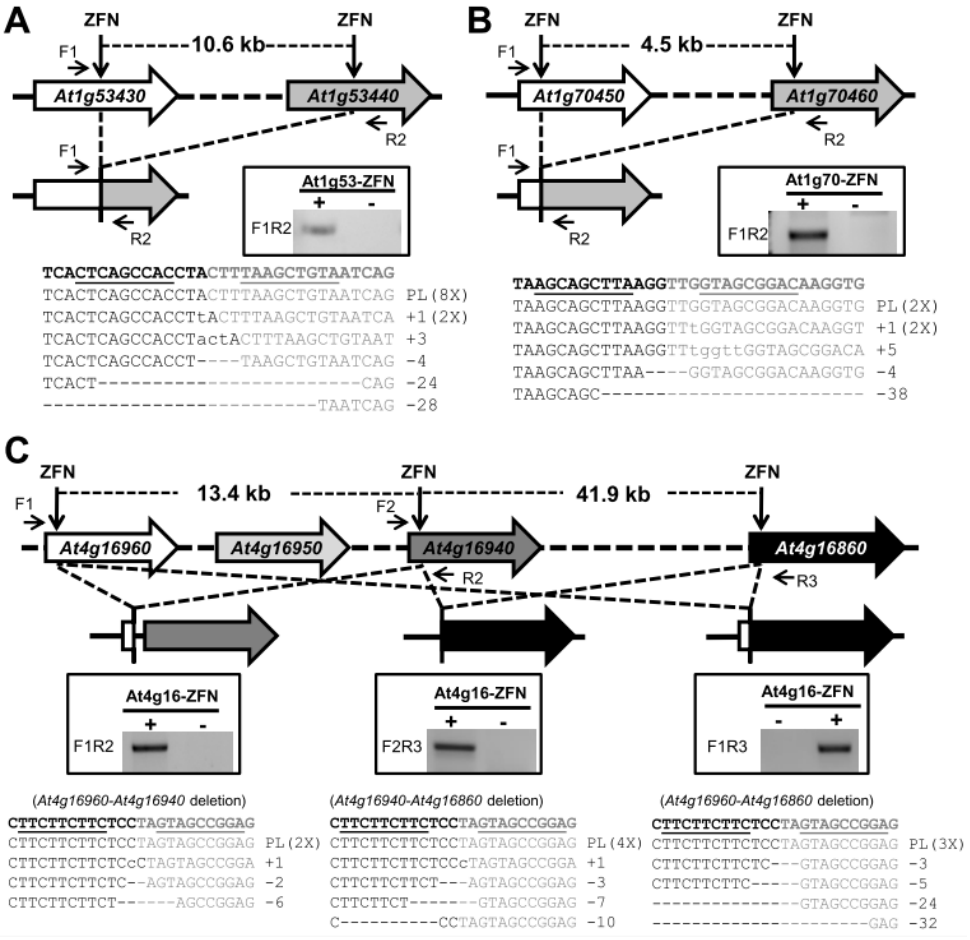

Figure 4-5. Deletion of gene clusters by ZFNs. (A) Deletion of the At1g53430-At1g53440 cluster by the At1g53-ZFN. (B) Deletion of the At1g70450-At1g70460 cluster by the At1g70-ZFN. (C) Three types of deletions at the At4g16960-At4g16860 gene cluster generated by the At4g16-ZFN. The gene clusters and resulting deletions are depicted. Deletion events were confirmed by PCR, as shown in windows in the middle of each panel. The positions of PCR primers are indicated by arrows. PCR products were subsequently cloned and sequenced. The sequencing results shown in the lower panels confirmed perfect ligations after loss of the intervening DNA or ligations with mutations at the ZFN cleavage site.

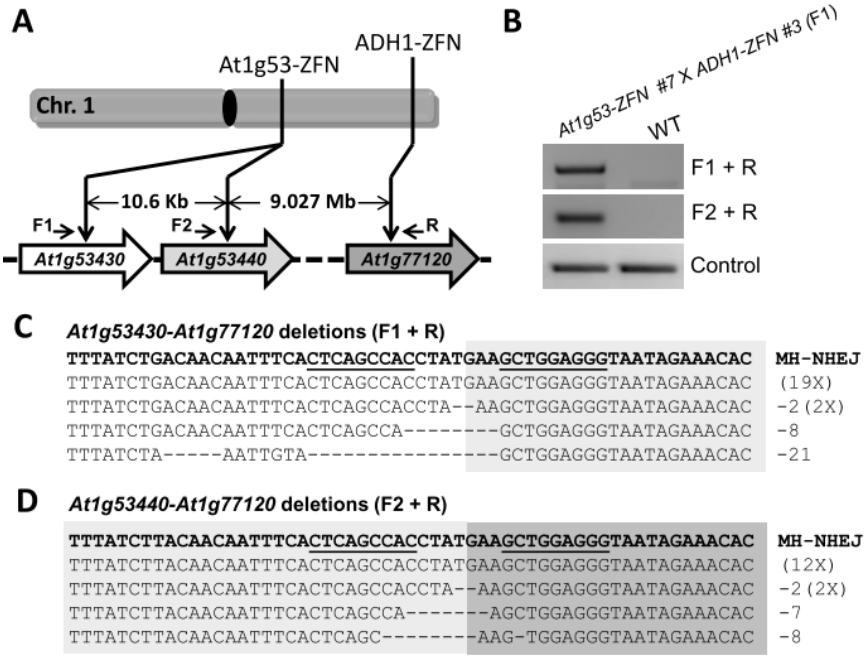

Figure 4-6. Large chromosomal deletions by ZFNs. (A) Schematic of ZFN targets on the right arm of *Arabidopsis* chromosome 1. The distance between the ZFN sites is shown and the positions of primers used to confirm large deletions are indicated. (B) PCR confirmation of large chromosomal deletions. The F1 and R primers amplify the junction fragment of the deletion of 9.037 Mb; primers F2 and R amplify the junction fragment of the deletion of 9.027 Mb (upper panels). PCR amplification of a part of the ADH1 gene was used as a genomic DNA control (lower panel). F1 seedlings generated from the cross between At1g53-ZFN #7 line and Adh1-ZFN #3 line were treated with estradiol, and the wild type plants served as a negative control. (C) Sequenced clones indicative of large chromosomal deletions between At1g53430 and At1g77120. (D) Sequenced clones indicative of large chromosomal deletions between At1g53440 and At1g77120. The DNA sequences resulting from perfect ligation of DNA ends are shown in the first line of the text boxes; deletions with indels are shown below. ZFN binding sequences are underlined. MH-NHEJ, end-joining that appears to have been facilitated by microhomology.

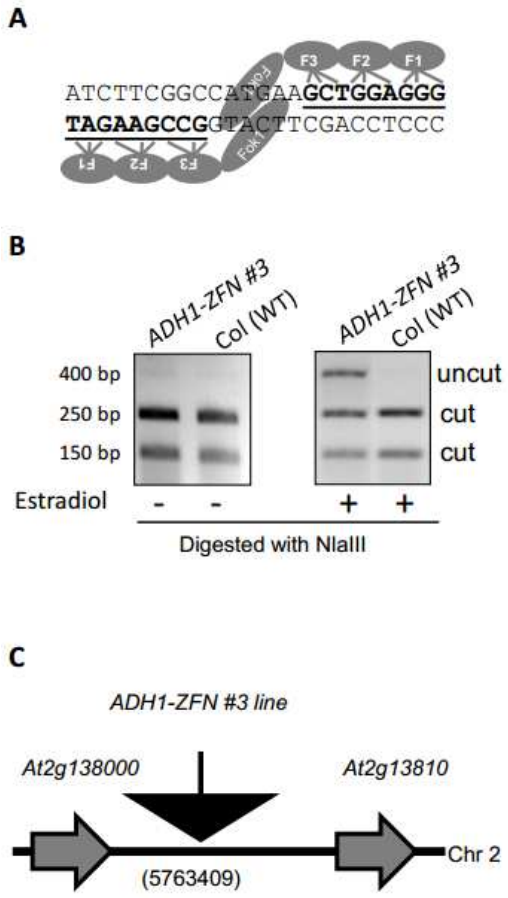

Figure 4-7. Supplemental Figure; An active ADH1-ZFN #3 line. (A) Schematic of the ADH1-ZFN and its target site. (B) ADH1-ZFN activity is highly estradiol-inducible. Mutagenesis activity, as reflected by the uncut band, was detected by PCR and digestion (C) Precise location of the transgene in ADH1-ZFN #3 line as mapped by TAIL-PCR.

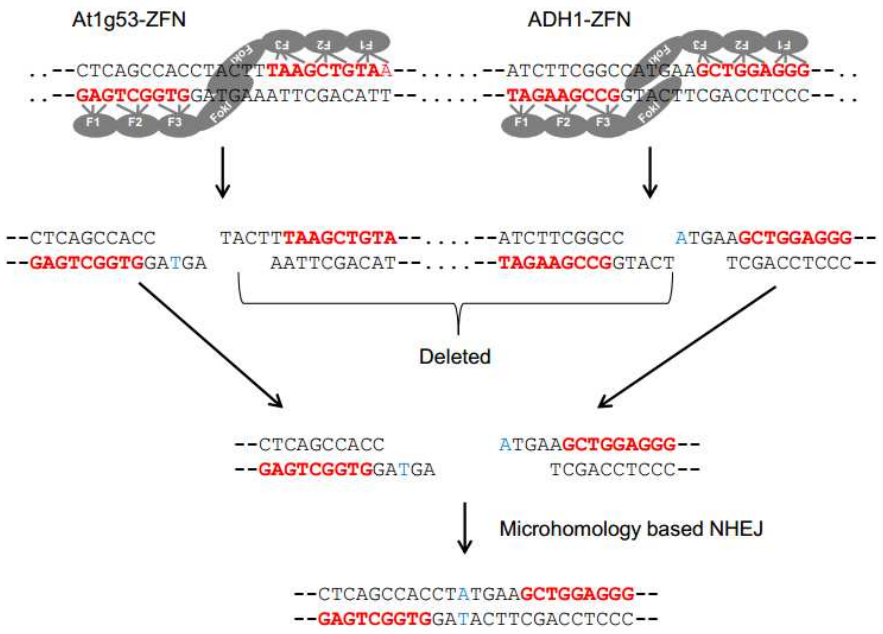

Figure 4-8. Supplemental Figure; A possible NHEJ repair mechanism using 1-bp of microhomology. The process that leads to a common ligation product is depicted. The ZFN binding sites are shown in red and the 1 nt of likely microhomology is marked in blue.

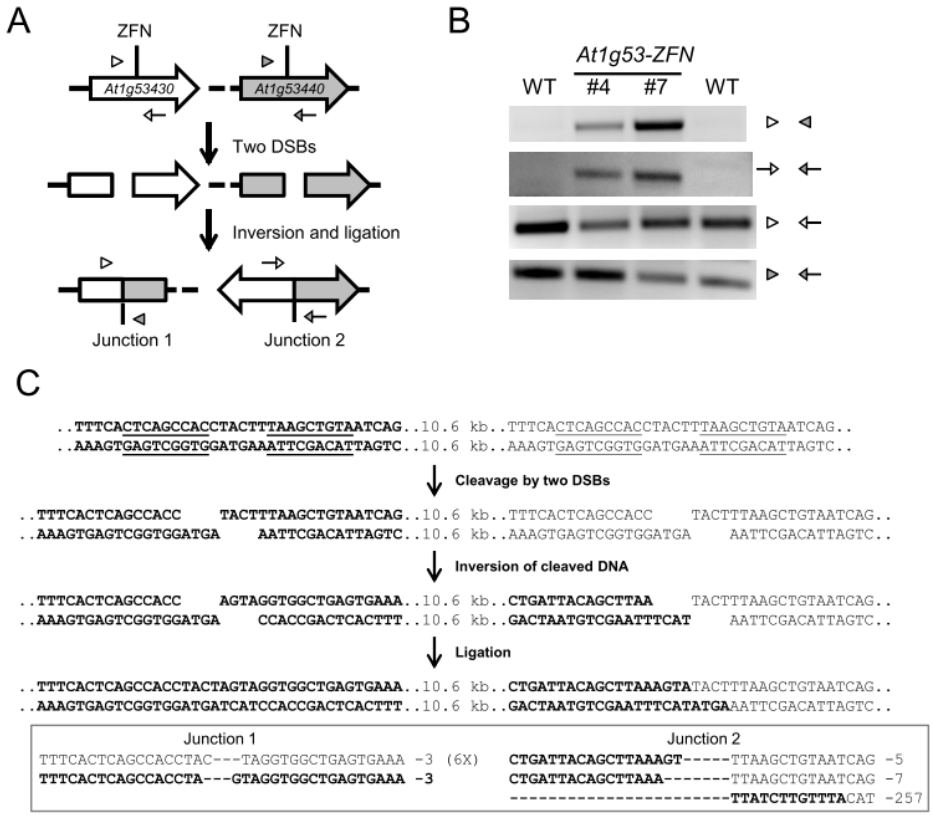

Figure 4-9. Inversion of the At1g53430 gene cluster. (A) Schematic of the At1g53430-At1g53440 gene cluster inversion. Positions of PCR primers for confirmation of inversions are indicated by empty or filled triangles and arrows. (B) PCR confirmation of gene cluster inversions. Two independent T2 lines were used to detect inversions; WT plants were used as the negative control. (C) Detailed depiction of the inversion event and DNA sequence confirmation of the inversions.

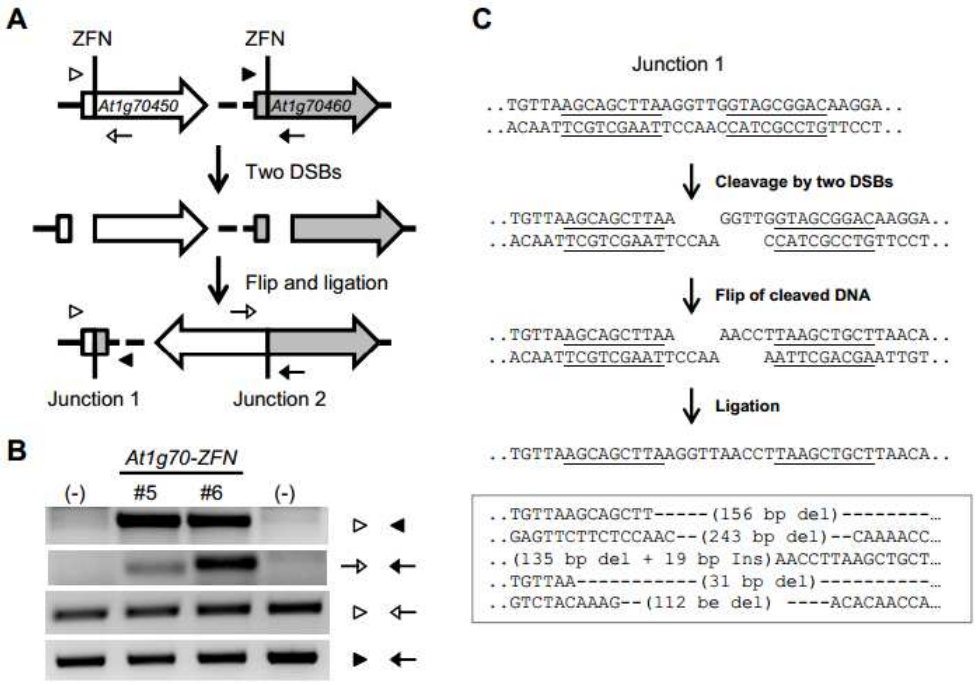

Figure 4-10. Supplemental Figure; Inversion of the At1g70450 gene cluster. (A) Schematic of the At1g70450gene cluster inversion. Positions of PCR primers for confirming inversions are indicated by empty or filled triangles and arrows. (B) PCR confirmation of gene cluster inversions. (C) DNA sequence confirmation of inversions.

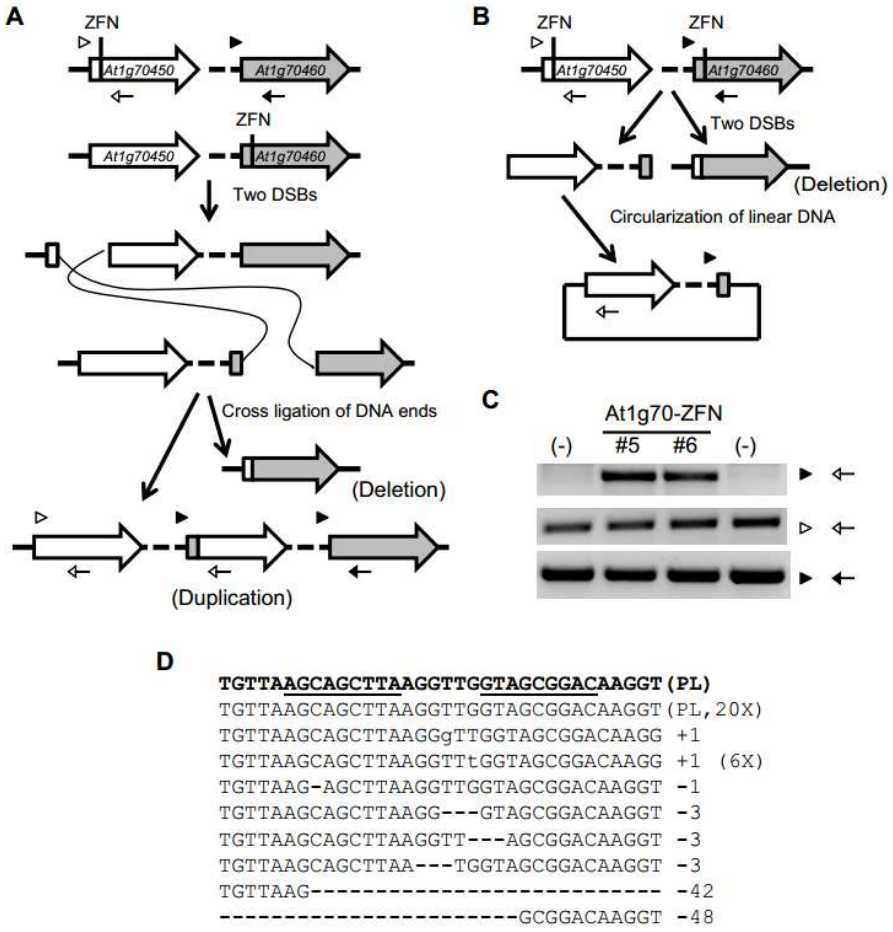

Figure 4-11. Supplemental Figure; Duplication of a gene cluster or circularization of deleted DNA at the At1g70450-At1g70460 locus. (A) Schematic of the At1g70450gene cluster duplication. (B) Schematic of circularization of deleted DNA. (C) PCR confirmation of possible gene cluster duplications. (D) DNA sequence data from clones indicative of possible gene cluster duplications.

Contribution to the work:

At the time we started this study in 2009, zinc-finger nucleases and meganucleases were the only programmable sequence-specific nucleases. Furthermore, designing zinc-finger nucleases to recognize a target sequence remained challenging. In this study, we designed the zinc-finger nuclease pairs At1g53-ZFN, At4g16-ZFN, At1g70-ZFN, At3g21-ZFN and At5g01-ZFN using context-dependent assembly (CoDA) (Sander et al., 2011). Results from this study provided potential amino-acid sequences for zinc-finger recognition helices that will recognize a desired target sequence. To facilitate cloning of zinc-finger nucleases, I developed the method for constructing multi-finger arrays with desired recognition helices (Qi et al., 2014). This method was used to build all the zinc-finger nucleases in this study (Table 4-2).

CHAPTER 5

TARGETED MUTAGENESIS OF DUPLICATED GENES IN SOYBEAN WITH ZINC-FINGER NUCLEASES

ABRIDGEMENT

We performed targeted mutagenesis of a transgene and nine endogenous soybean (Glycine max) genes using zinc-finger nucleases (ZFNs). A suite of ZFNs were engineered by the recently described context-dependent assembly platform—a rapid, open-source method for generating zinc-finger arrays. Specific ZFNs targeting DICER-LIKE (*DCL*) genes and other genes involved in RNA silencing were cloned into a vector under an estrogen-inducible promoter. A hairy-root transformation system was employed to investigate the efficiency of ZFN mutagenesis at each target locus. Transgenic roots exhibited somatic mutations localized at the ZFN target sites for seven out of nine targeted genes. We next introduced a ZFN into soybean via whole-plant transformation and generated independent mutations in the paralogous genes *DCL4a* and *DCL4b*. The *dcl4b* mutation showed efficient heritable transmission of the ZFN-induced mutation in the subsequent generation. These findings indicate that ZFN-based mutagenesis provides an efficient method for making mutations in duplicate genes that are otherwise difficult to study due to redundancy. We also developed a publicly accessible Web-based tool to identify sites suitable for engineering context-dependent assembly ZFNs in the soybean genome.

INTRODUCTION

Soybean (*Glycine max*) is an ancient polyploid and major agricultural legume crop providing nutritional protein and oil that can be processed into a variety of feed and food products. Several genetic bottlenecks throughout its domestication and more recent intensive selection and breeding practices have greatly reduced the genetic variability of soybean germplasm (Hyten et al., 2006). Current efforts to expand genetic tools for breeding and gene discovery include random mutagenesis and RNAi-based approaches. Several published and ongoing studies have utilized chemical mutagens including ethyl methanosulfonate for TILLING (Cooper et al., 2008), radiation mutagens such as fast neutrons (Men et al., 2002), and transposable elements (Mathieu et al., 2009). However, random mutagenesis approaches in a highly duplicated genome such as soybean often result in many lines with no phenotype due to complementation by redundant genes. This can sometimes be circumvented by remutating single-homeolog mutant lines to obtain the required bona fide double-homeolog mutants. Another approach is to identify and combine mutations by genetic crossing (Pham et al., 2010) but this can be time consuming. RNAi-based approaches such as posttranscriptional gene silencing either by hairpin or virus-induced gene silencing vectors suffer from the opposite problem, namely that it is difficult to silence individual gene copies, and rather entire gene families are often silenced (Kachroo et al., 2008; Meyer et al., 2009).

An ideal mutagenesis approach for a highly duplicated genome like soybean would allow for the simultaneous recovery of plants with single or multiple mutations in each member of a gene family of interest without disruption to the rest of the genetic background. Site-directed

mutagenesis using zinc-finger nucleases (ZFNs) provide an attractive method for producing this desired result (Zhang et al., 2010). Engineered ZFNs are a recently developed tool for targeted gene alteration, and their implementation in several model plants and animals suggests they could potentially be of great utility to the soybean research community. Importantly, modification of genes in maize (*Zea mays*) and tobacco (*Nicotiana tabacum*) with ZFNs has been reported (Shukla et al., 2009; Townsend et al., 2009), as well as high-frequency heritable transmission of ZFN-induced mutations in Arabidopsis (*Arabidopsis thaliana*; Zhang et al. 2010).

To bind and cleave a target site, the ZFN forms a heterodimer comprised of left and right monomers. The ZFNs are a fusion of a zinc-finger array (ZFA) consisting of engineered Cys_2His_2 zinc fingers and a nonspecific DNA-cleavage domain of the *Fok*I restriction enzyme (Urnov et al., 2010). The binding sites of a three-finger ZFA are typically 9 bp and are separated by a 5- to 7-bp spacer to allow for the dimerization of the *Fok*I nuclease. This is an important aspect of ZFN design, as the independent binding of both ZFAs separated by the appropriate spacer is critical for correct dimer formation. Upon successful dimerization of the *FokI* monomers, a double-stranded break is generated in the spacer sequence between both ZFA binding sites. This double-stranded break subsequently stimulates the cellular DNA repair pathways, which include the error-prone nonhomologous end-joining and the homology-directed repair. The nonhomologous end-joining pathway ligates double-stranded breaks in DNA, often introducing small nucleotide insertions and deletions (indels) at the target DNA site that can disrupt the gene's reading frame.

Several strategies have been employed for the design and construction of ZFNs including modular assembly, oligomerized pool engineering (OPEN), and context-dependent assembly (CoDA; Maeder et al. 2009; Sander et al. 2011; Wright et al. 2006). CoDA is the most recent platform developed by the Zinc Finger Consortium and has the advantage in that it is rapid and easy to perform because it does not require labor-intensive selection methods. With CoDA, novel ZFNs are created by assembling arrays from a large archive of optimized two-finger units. We have previously demonstrated that CoDA ZFNs can be used to create mutations at the intended target site in transformed soybean roots (Sander et al., 2011). In this article we engineered eight ZFNs using the CoDA platform to target individual genes and duplicate gene pairs in soybean. We demonstrate that in this highly duplicated genome, site-directed mutagenesis with ZFNs can be used to introduce a series of unique allelic combinations in members of a given gene family. Further, we demonstrate that ZFNs can be used to generate heritable mutations in soybean. Our work suggests that ZFNs will be particularly useful for studying plant functional genomics, as the vast majority of plant (and crop) species have experienced polyploidization events in their recent evolutionary history and maintain homeologous and paralogous copies of many genes (Blanc and Wolfe, 2004; Schmutz et al., 2010).

RESULTS AND DISCUSSION

ZFN-Induced Mutagenesis of a GFP Transgene in Soybean Hairy-Root Tissues

To establish the parameters for efficient targeted mutagenesis by ZFNs in soybean, a previously characterized ZFN that targets *GFP*(Maeder et al., 2008) was introduced into soybean by the *Agrobacterium rhizogenes* hairy-root transformation method. In this method, the ZFN is integrated into the soybean chromosome along with genes from *A. rhizogenes* that promote root development. Transgenic hairy roots can be obtained within two weeks of transformation and we reasoned they would be useful for testing the efficacy of ZFN mutagenesis prior to whole-plant transformation and provide an effective means of rapidly screening ZFN function at endogenous targets. The ZFN-targeting *GFP* was driven by an estrogen-inducible promoter (Zuo et al., 2000). The recipient soybean line (cv Jack) harbored a homozygous *GFP* transgene (Hernandez-Garcia et al., 2009) with a *Bcc*I recognition site within the left ZFA target sequence.

Genomic DNA was extracted from the ZFN-transformed hairy-root samples and assayed for mutations in the GFP coding region. To perform this assay, the DNA was first digested with BccI and then PCR amplified at the GFP locus, such that only sequences with new mutations at the BccI recognition site would be amplified (Figure 5-1A). Amplification products were redigested, cloned, and sequenced. Using this PCR enrichment strategy, deletions ranging from 27- to 71-bp biased toward the left ZFA were recovered in five of 13 clones sequenced, indicating that the GFP ZFN effectively introduces mutations in its target sequence (Figure 5-1B). These experiments demonstrated that the hairy-root expression system is a rapid and accurate in vivo screen for ZFN mutagenesis activity in soybean.

ZFN-Induced Mutagenesis of Single and Duplicated Soybean Genes in Hairy-Root Tissues

To test the ZFN-mutagenesis system on endogenous soybean genes, we used the recently described CoDA platform to engineer eight ZFNs that target endogenous soybean genes involved in various aspects of RNA silencing (Table 5-1; Table 5-2). Based on homology to known genes in Arabidopsis, the targeted genes include *DICER-LIKE* (*DCL*), *RNA-DEPENDENT RNA POLYMERASE* (*RDR*), and *HUA ENHANCER1* (*HEN1*) family members (Margis et al., 2006; Wassenegger and Krczal, 2006; Yang et al., 2006). Soybean is a highly duplicated paleopolyploid plant (Schmutz et al., 2010), therefore some of the ZFNs recognize duplicate gene copies. Three ZFN constructs were developed to simultaneously target two paralogous gene copies (two constructs targeting *DCL1a/DCL1b* and one construct targeting *DCL4a/DCL4b*) and five constructs were developed to independently target individual genes (*DCL2a, DCL2b, RDR6a, RDR6b,* and *HEN1a*; Table 5-1; Table 5-2).

The hairy-root transformation method was used to evaluate ZFN mutagenesis of the nine endogenous soybean genes. As with the GFP ZFN, the CoDA ZFNs were driven by an estrogen-responsive promoter, and expression of each ZFN was induced by the introduction of estrogen in the tissue culture media. DNA from transgenic root tissues was screened via the enrichment PCR

method described above to determine whether ZFN activity generated site-specific mutations. Briefly, DNA samples were digested by an appropriate restriction enzyme that recognizes the spacer sequence in the wild-type target site (Figure 5-2A) to enrich for the mutated sequences and then PCR amplified. The resulting PCR products were subsequently redigested and then visualized by agarose gel electrophoresis (Figure 5-2B). If the site was mutated in some cells, then the PCR product would fail to digest a portion of the sample. Undigested DNA fragments were observed for five ZFN-transformed lines, indicating putative mutations in a total of seven gene targets (Table 5-1; Figure 5-2B). The undigested PCR products were cloned and sequenced and several distinct mutated alleles consisting of small insertions or deletions ranging from 1 to 20 bp were recovered (Figure 5-3). We failed to recover mutations from the remaining three ZFN constructs (Table 5-2).

To assess whether CoDA ZFNs can discriminate between closely related DNA sequences, we constructed two ZFNs that independently target the *RDR6* homeologs *RDR6a* and *RDR6b*. These ZFNs differ only by the subsite bound by F3 of the right ZFA, which should enable them to discriminate between the 2 bp distinguishing the *RDR6a* and *RDR6b* target sites (Table 5-1). Enrichment PCR using primers common to both genes recovered ZFN-induced mutations for the predicted ZFN/homeolog combinations. Importantly, no evidence of ZFN activity was observed at noncognate homeologous sites among 16 clones sequenced, indicating specificity of the ZFNs to their respective targets. PCR enrichment assays specific to each homeolog were performed to further validate the specificity of the respective ZFNs (Figure 5-4). The results indicate that the targeted gene copy was mutagenized at a much higher frequency than the off-target copy. We cannot, however, rule out the possibility that some mutations may be occurring at the off-target homeologous gene, albeit at a much lower frequency than the targeted gene.

Taken together, we have shown that CoDA ZFNs are exceptionally potent and selective mutagenic agents. We have also shown that hairy-root transformation is a rapid and reliable method for testing the function of CoDA ZFNs prior to the arduous task of soybean whole-plant transformation.

Mutagenesis and Heritability of a Duplicated Gene Pair in Soybean

Whole-plant transformation was attempted on approximately 100 explants to introduce the ZFN targeting both paralogous copies of the *DCL4* gene (*DCL4a* and *DCL4b*) into soybean. We recovered three ZFN-transformed T0 seedlings from the hormone-treated explants. Additionally, a control group of explants was transformed without hormone induction to gauge potential ZFN toxicity or hormone effects on transgenesis (Table 5-3).

DNA from true and unifoliate leaves from the three hormone-treated plants were screened by the enrichment PCR method to identify ZFN-directed mutations. Restriction enzyme-resistant PCR fragments were recovered from two of the three plants. The undigested PCR

products were subsequently cloned and sequenced, revealing one T0 seedling with an adenine base insertion at the ZFN target site in the *DCL4a* locus (Figure 5-5) and the other T0 seedling with a two-base thymine and adenine insertion at the *DCL4b* locus (Figure 5-6A). The presence of unmutated *DCL4a* and *DCL4b* sequences suggested that only one allele from both*DCL4a* and *DCL4b* had been mutated, thus both plants were likely either heterozygous (*DCL4a*/*dcl4a* and *DCL4b*/*dcl4b*) or chimeric.

Both T0 plants were grown to maturity and seed harvested. While the plant with the *dcl4b* mutation appeared normal, the plant with the *dcl4a* mutation exhibited a severe developmental phenotype with large bulbous internodes and mostly undeveloped and aborted seeds. It is unclear whether this phenotype was due to the *dcl4a* mutation, insertional mutagenesis of the ZFN construct into a dosage-sensitive gene, off-target mutagenesis by the ZFN, or somaclonal variation induced by tissue culture. Only two viable seeds were harvested from the *DCL4a*/*dcl4a* T0 plant, neither of which showed compelling evidence for transmission of the *dcl4a* mutation. The detailed analyses of these plants are shown in Figure 5-5.

The *DCL4b*/*dcl4b* plant produced approximately 500 seeds and these progeny were used to study the heritability of the ZFN-induced mutation. To test the germinal transmission of the *dcl4b* mutation, 24 T1 seedlings were grown and genotyped. A PCR assay using *DCL4b*-specific primers was carried out to determine if the mutation was heritable (Figure 5-6B).

The *dcl4b* mutation segregated exactly 1:2:1 in the T1 progeny, with six seedlings being homozygous for the mutation *dcl4b*/*dcl4b*, 12 seedlings heterozygous *DCL4b*/*dcl4b*, and six seedlings homozygous wild-type *DCL4b*/*DCL4b* (Figure 5-6C). To confirm the genotyping, PCR was performed using *DCL4b*-specific primers on the genomic DNA of a T1 individual putatively homozygous for the mutation. The PCR product was cloned and 16 colonies were sequenced. Sequence data confirmed all colonies had the expected 2-bp insertion for *dcl4b* at the target site (Figure 5-7) and no wild-type allele was recovered from this assay. There was some evidence for increased lateral shoot growth in the *dcl4b*/*dcl4b* individuals, however no striking phenotypic alterations were observed. Further experimental replications with more detailed measurements will need to be performed to confirm and quantify the lateral shoot growth phenotype.

In using ZFNs as mutagens, it may be advantageous to remove the ZFN transgene from subsequent generations to minimize potential toxicity or additional rounds of mutagenesis. Transgene removal could be accomplished by normal genetic segregation. To look for this, the 24 T1 plants were PCR scored for the ZFN transgene. Only one of the T1 plants, a heterozygous *DCL4b*/*dcl4b* individual, lacked the ZFN construct (Figure 5-6C).

Collectively, the results show heritable targeted mutagenesis of a soybean gene. Using the materials generated here, *dcl4* double mutants (*dcl4a*/*dcl4a*/*dcl4b*/*dcl4b*) could be obtained in a variety of ways, including reactivating expression of the ZFNs in a *dcl4b*/*dcl4b* T1 line.

The Identification of ZFN Target Sites in Soybean Using a Web-Based Tool

To aid in the identification of potential sites for ZFN engineering by CoDA, a version of ZFNGenome was implemented for soybean. ZFNGenome is a GBrowse-based (Stein et al., 2002) tool for identifying and visualizing potential target sites for CoDA ZFNs (Reyon et al., 2011). ZFNGenome provides researchers with information about each potential ZFN target site, including its chromosomal location, position relative to transcription initiation site(s), and frequency of occurrence within the genome. Users can query ZFNGenome using several different criteria (e.g. gene ID, transcript ID, or target site sequence). Targets identified using ZFNGenome can be visualized at multiple scales within the flexible GBrowse 1.7 environment and can be imported as annotations into other genome browsers. ZFNGenome is dynamically linked to the Zinc Finger Database (Fu et al., 2009), allowing users access to all available information about zinc-finger reagents, such as the effectiveness of a given ZFN in creating double-stranded breaks.

The ZFNGenome tool for soybean indicates that 36,714 out of 55,582 (approximately 66%) protein-encoding transcripts can be targeted by CoDA ZFNs. There is an average of 2.93 ZFN targets per coding transcript (107,665 target sites among the 36,714 coding transcripts).

ZFNGenome is freely available at http://bindr.gdcb.iastate.edu/ZFNGenome. The interface for the soybean genome is found athttp://bindr.gdcb.iastate.edu/ZFNGenome/Soybean/.

CONCLUSION

In conclusion, we describe a rapid and highly specific method for generating gene mutations in a genetically redundant paleopolyploid crop species. Our data indicate that the CoDA-designed ZFN pairs have a high rate of success as mutagens, and their ease of construction should facilitate the development of additional applications in soybean, for example, to create targeted gene insertions or allelic replacements, both of which have been accomplished in other plant species (Shukla et al., 2009; Townsend et al., 2009). We anticipate that the CoDA platform will be widely adopted as an efficient and powerful functional genomics tool for soybean and other nonmodel plant species with highly duplicated genomes. Similar site-directed approaches, such as the transcription activator-like effector nuclease system (Christian et al., 2010; Li et al., 2011; Miller et al., 2011), may also be successful at simultaneously targeting single and paralogous loci in such genomes.

MATERIALS AND METHODS

Construction of OPEN and CoDA ZFAs and ZFN Expression Vectors. The GFP ZFN nuclease was engineered by the OPEN platform, and has been previously reported (Maeder et al., 2008). Endogenous soybean (*Glycine max*) genes targeted in this study were named based on homology to previously characterized genes in Arabidopsis (*Arabidopsis thaliana*; names and

Glyma gene identifiers are shown in Table 5-1). ZFN target sites in endogenous soybean gene sequences were identified using the publicly available Web-based program Zinc Finger Targeter (Sander et al., 2007). Target sequences were queried to the soybean genome sequence using the BLAST function (http://www.phytozome.net/soybean) to confirm that ZFN targets were within exons. Two ZFNs that recognize the target sequence of *DCL1* (two copies:*DCL1a*/*DCL1b*) and one ZFN that recognizes the target sequence of *DCL4* (two copies: *DCL4a*/*DCL4b*) were selected to simultaneously target paralogous genes. Five ZFNs that recognize the individual genes copies *DCL2a*, *DCL2b*, *RDR6a*,*RDR6b*, *and HEN1a* genes were also selected. Potential sites were considered in which the ZFAs were separated by 5- or 6-bp spacer sequences (Table 5-1; Table 5-2) since it has been reported that the ZFN linker used in this study had significantly improved activity when designed around target sites with these spacer lengths (Händel et al., 2009). Individual fingers designated F1, F2, and F3 each recognizing 3-bp of the 9-bp target site were identified using the CoDA archive (Sander et al., 2011). The CoDA archive contains 319 F1 and 344 F3 units each of which have been identified in previous arrays to function correctly when arranged with one of the 18 common F2 units (Sander et al., 2011). Selected ZFAs were assembled by combining the three individual fingers using a fusion PCR assay (*Pfu* Turbo, Stratagene) from a collection of individual recognition helices (RHs; F1, F2, and F3 PCR products). For some ZFAs, we rapidly engineered required recognition helix variants using a mutagenic PCR assay. Briefly, new RHs can be generated by converting the amino acid sequence of the desired RH to DNA based on the soybean codon usage and incorporating the required sequence into a primer pair. PCR is carried out using an existing finger plasmid as a template. The PCR product is purified using a standard clean-up kit (Qiagen),*Dpn*I (New England Biolabs) treated to remove the plasmid template, and transformed into DH5α competent cells for in vivo cloning (see Figure 5-8 for a detailed protocol of the mutagenic PCR assay). Upon completion of the fusion PCR, the left and right ZFA half sites were digested with *Bam*HI/*Xba*I and ligated into the *Bam*HI/*Xba*I sites of the pFZ50 expression vector encoding both the *Fok*I nucleases and a T2a ribosomal skipping protein (Zhang et al., 2010). The right ZFA *Bam*HI/*Xba*I half site was ligated into the compatible *Bgl*II/*Nhe*I sites of the L_ZFA/pFZ50 construct. PCR using a proof-reading polymerase was carried out to generate the complete ZFN cassette with forward and reverse primers incorporating an *XhoI* and *NheI* site, respectively. An estrogen-inducible expression vector suitable for soybean transformation was constructed. The inducible cassette from pER8 vector (GenBank: AF309825) was used as a template for PCR with primers incorporating *Not*I sites (Zuo et al., 2000). The inducible cassette was cloned into the *Not*I sites of the binary vector pNB96 (Fusaro et al., 2006). This inducible binary vector was further digested with*Xho*I and *Spe*I to allow for the ligation of the ZFN *Xho*I/*Nhe*I cassette Hairy-Root and Whole-Plant Transformation of ZFNs in Soybeans and

Hairy-Root Transformation. Each ZFN binary construct was independently transformed into *Agrobacterium rhizogenes* strain K599 for hairy-root transformation. Soybean cotyledons were inoculated with the transformed K599 strain using a previously reported protocol (Govindarajulu et al., 2008) to introduce the ZFN transgene into the hairy-root progenitor cells. The following is a detailed protocol for hairy root transformation of soybean cotyledons. Mature soybean seed of [Glycine max (L.) Merr.] cultivar (cv. Bert.) were surface sterilized with Chlorine gas for 16-24 hours. Disinfected seeds were placed (six per plate) on ¼ MS media (2.2g/L MS Basal Medium (Phytotech, USA); dH2O adjusted to pH 5.8; 750mg/L MgCl2; 8g/L granulated agar (Difco, USA)). Seeds were germinated for seven days at 24oC under 18:6 hour photoperiod (~150µmol s-1m-2). An *Agrobacterium* culture was prepared (10mL LB; 200mg/L kanamycin), spiked from a glycerol stock harboring an appropriate ZFN transgene and incubated at 28oC for two days. The bacterial culture was pelleted at 4K for 10 minutes, re-suspended and made up to a volume to 50mL with ¼ MS liquid media. The OD600nm was adjusted to approximately 0.2-0.3 and transferred to an appropriately labeled 90 x 25mm plate. Seven day old cotyledons were harvested and prepared for transformation by cutting between the hypocotyl and the half-way point of the cotyledon. The seed coat was removed if necessary and transferred to the liquid culture. The plate was placed in a vacuum chamber and a vacuum was drawn for five minutes and held under vacuum for a further 20 minutes. The vacuum was released and cotyledons were placed flat side up on sterile 70mm filter paper inserted into a 90 x 25mm plate (the filter paper should be wet from the liquid culture). Plates were wrapped with parafilm and incubated in darkness at 28oC for three days. Cotyledons were next removed from the filter paper and placed in a 90 x 25mm plate and washed with ¼ MS liquid media plus Carbenicillin (500mg/L) for 30 minutes. Cotyledons were removed from the wash liquid and forced into ¼ MS solid media (Carbenicillin 500mg/L; 10µM 17β-estradiol (Sigma-Aldrich)), with the cut surface facing up out of the media (6 cotyledons per plate) and wrapped with parafilm. Approximately fourteen days post-inoculation hairy-roots should appear from calli.

Stable transformation of ZFNs in soybeans. Mature soybean [*Glycine max* (L.) Merr.] seeds of cultivar Bert were surface sterilized with Chlorine gas for 16-24 hours. Approximately 100 disinfected seeds were imbibed with sterile water for 16-20 hours. To protect seeds from light, the plates were covered with aluminum foil. The imbibed seeds were incised along the hilum to separate the cotyledons and to facilitate the removal of the seed coat. The embryonic axis found at the nodal end of the cotyledons was removed in preparation for infection.
An inoculum of *Agrobacterium rhizogenes* 18r12 strain harboring the inducible zinc-finger nuclease was prepared by inoculation of 150ml of YEP broth (Kanamycin (200mg/L), from an appropriate glycerol stock and incubated at 28℃ fo r 2 days on an incubator/shaker. The culture was grown overnight until an OD650=0.7 to 1.0 and centrifuged at 5K for 10 minutes. The supernatant was decanted from the distinctively pink bacterial pellet and subsequently re-

suspended in 30mL of LCCM media (All tissue culture reagents hereafter are products from Phytotech, USA unless otherwise indicated). (LCCM: 0.03% Gamborg's Basal Salt Mixture, 1/10 B5 vitamins, 7.5μM 6-benzylaminopurine (BAP), 0.7μM gibberellic acid (GA3), 20mM MES [2-(N-morpholino)ethanesulfonic acid], 3% sucrose (Fisher Scientific, USA) and 200μM acetosyringone, pH 5.4). The re-suspended inoculum was gently incubated on a shaker incubator at 80rpm for 30 minutes and transferred to a petri-dish containing the half-seed explants for inoculation.

The explants were submerged in the infection media and incubated at room temperature for 30 minutes with occasional agitation. Upon completion of infection, the explants were transferred flat-side down to plates containing co-cultivation media (CCM) lined with filter paper. (CCM: 0.3% Gamborg's Basal Salt Mixture, 1/10 B5 vitamin, 7.5μM BAP, 0.7μM GA3, 0.06% MES, 3% sucrose (Fisher Scientific, USA), 200μM acetosyringone, 50mgL-1 cysteine and 0.5mM dithiothreitol, pH5.4 with 0.425% Noble agar (Fisher Scientific, USA)).

Co-cultivation was carried out for five days at 24^{o}C under 18:6 hour photoperiod (~150μmol s-1m-2). After co-cultivation, the half-seed explants were briefly washed with liquid shoot induction media in a sterile petri-dish (LSIM: 0.3% Gamborg's B5 salts and MSIII iron stock, 3% sucrose (Fisher Scientific, USA), 0.06% MES, 10μM 17β-estradiol (Sigma-Aldrich), 250mgL-1 Ticarcillin, pH5.6). The explants were transferred to tissue culture by embedding the base of the explants (flat side up) into the shoot induction media; this media included the estrogen treatment used to induce the ZFN transgene expression (SIM I: 0.3% Gamborg's Basal Salt Mixture, 5μM BAP, 3% sucrose, 250mgL-1 Ticarcillin, 5μM Kinetin, 10μM 17β-estradiol (Sigma-Aldrich) and 0.7% Noble Agar).

Shoot induction was carried out at 24C with 18:6 hour photoperiod (~150μmol s-1m-2) for fourteen days. The explants were transplanted to shoot induction media II (SIM II), identical preparation to SIM I with herbicide selection (5mgL-1 Glufosinate Ammonium (Sigma-Aldrich)). Prior to SIM II, any large shoots were aseptically cut and discarded. A fresh cut is made at the base of the explants/shoot pad to facilitate uptake of the herbicide. The cut surface was embedded into the SIM II media and incubated at 24C with 18:6 hour photoperiod (~150μmol s-1m-2) for fourteen days. After four weeks on shoot induction media, the cotyledon was removed from the explants and a fresh cut was made at the base of the callus/shoot pad prior to being embedded in the shoot elongation media (SEM) (SEM: 0.44% MS Salts, MSIII iron stock, B5 vitamins, 3% sucrose (Fisher Scientific, USA), 0.06% MES, 0.5 mgL-1 GA3, 50mgL-1, Asparagine, 100mgL-1 L-pyroglutamic acid, 0.1mgL-1 IAA, 0.5mgL-1 zeatin riboside, 250mgL-1 Ticarcillin, 3mgL-1 glufosinate, pH 5.6 and 0.7% Noble Agar).

Callus/shoot pads were transferred to fresh SEM every 2-3 weeks (or as needed) up to 8 weeks or until stems had elongated to greater than 2cm. Shoots surviving selection approximately 2-3cm in length were excised from the callus/shoot pad, dipped into 1mgL-1 sterilised indole-3-butyric acid (IBA) for 1-2 minutes and transferred to rooting media (RM). (RM: 0.3% Gamborg's Basal

Salt Mixture, 2% sucrose (Fisher Scientific, USA), 0.06% MES, 57μM IBA, 250mgL-1 Ticarcillin, pH 5.6).

Shoots were incubated for 1-2 weeks at 24C under 18:6 hour photoperiod (~150μmol s-1m-2) or until two or more roots had developed. Plants were next regenerated into soil planting into jiffy pots with moistened metro mix. Pots in flat without holes covered with a humidome and grow at 24℃ under 18:6 (light:dark) photoperiod for at lea st one week, watering as needed spraying 2-week-old seedlings with 150 mg/L Liberty® herbicide confirm the presence of the bar transgene in T1 progeny.

Screening for mutations. The ZFN transgene was driven by an estrogen-inducible expression system (Zuo et al., 2000). Inoculated cotyledons were incubated on Murashige and Skoog medium plates treated with 10 μM of 17β-estradiol (Sigma-Aldrich). Approximately 1 to 2 weeks after transformation roots were randomly selected for DNA extraction using either a DNeasy (Qiagen) or a hexadecyltrimethylammonium bromide protocol (Curtin et al., 2008). Confirmation of the ZFN transgene in hairy roots was performed using PCR with a transgene-specific primer set. Mutations introduced by the ZFN disrupt the restriction site by insertion or deletion of DNA at the target sequence. PCR using primers designed to span the target site was carried out on predigested genomic DNA. Amplicons were then digested with a restriction enzyme that recognized the wild-type target sequences. Uncleaved products were visualized by agarose gel electrophoresis. Identification of the mutated sequences was accomplished by cloning and sequencing the uncleaved products.

Table 5-1 The gene target accessions, target sequence, and RHs of CoDA ZFNs that generated mutations in the target genes

Gene Name[a]	Accession No.	Target site	Spacer (nt)	RH (F1)	RH (F2)	RH (F3)
DCL1a [b]	Glyma03g42290	cAGCAACCTCTTATAAGAGGGCGTGg	6	TKQILGR	HKSSLTR	RHDQLTR
		gTCGTTGGAGAATATTCTCCCGCACc		SRFTLGR	LKEHLTR	RVDNLPR
DCL1b [b]	Glyma19g45060	cAGCAACCTCTTATAAGAGGGCGTGg	6	TKQILGR	HKSSLTR	RHDQLTR
		gTCGTTGGAGAATATTCTCCCGCACc		SRFTLGR	LKEHLTR	RVDNLPR
DCL4a [b]	Glyma17g11240	tTGCTTCATCACAATGGAGATGATt	5	RGQELRR	QQTNLTR	VGSNLTR
		aACGAAGTAGTGTTACCTCTACTAa		TKQRLVV	VRHNLTR	QTTHLSR
DCL4b [b]	Glyma13g22450	tTGCTTCATCACAATGGAGATGATt	5	RGQELRR	QQTNLTR	VGSNLTR
		aACGAAGTAGTGTTACCTCTACTAa		TKQRLVV	VRHNLTR	QTTHLSR
RDR6a	Glyma04g07150	aGGCAACGACATCAGAGGAGTGGAAa	6	LKKDLLR	HKSSLTR	DRTPLQR
		tCCGTTGCTGTAGTCTCCTCACCTTt		HKPNLHR	RREVLEN	QKPHLSR
RDR6b	Glyma06g07250	aGGCAACGACATCAGAGATGTGGAAa	6	LKKDLLR	HKSSLTR	DRTPLQR
		tCCGTTGCTGTAGTCTCTACACCTTt		HKPNLHR	RREVLEN	ISHNLAR
HEN1a	Glyma08g08650	aGCACGCGACCACGCGTAGACGCAt	5	RSRNLTL	RTDTLAR	ESGALRR
		tCGTGCGCTGGTGCGCATCTGCGTa		EESNLRR	DRGNLTR	QSTSLQR

[a] Gene names are based on homology to previously characterized genes in *Arabidopsis thaliana*.
[b] Represents two duplicate copies ("a" and "b") that both perfectly match the ZFN target site.

Table 5-2. Supplemental Table; The gene target accessions, target sequence and recognition helices (RH) of CoDA zinc-finger nucleases that did not generate mutations in the target genes.

Gene Name[a]	Accession No.	Target site	Spacer (nt)	RH (F1)	RH (F2)	RH (F3)
DCL1a [b]	Glyma03g42290	gGACAGCATCAAGAACGGTGCTGGTc	6	TRAVLRR	QRSDLTR	LTHNLRR
		cCTGTCGTAGTTCTTGCCACGACCAg		RRQKLTI	QRSDLTR	HGHRLKT
DCL1b [b]	Glyma19g45060	gGACAGCATCAAGAACGGTGCTGGTc	6	TRAVLRR	QRSDLTR	LTHNLRR
		cCTGTCGTAGTTCTTGCCACGACCAg		RRQKLTI	QRSDLTR	HGHRLKT
DCL2a	Glyma09g02930	tGCCTTCCTCAGCACTGGTGGTGAAa	6	TRAKLHI	QQTNLTR	RRDNLNR
		aCGGAAGGAGTCGTGACCACCACTTt		RKPNLLR	EAHHLSR	IRHHLKR
DCL2b	Glyma09g02920	tGCCTTCCTCAGTACTGGTGGCGAAa	6	TRAKLHI	QQTNLTR	RRDNLNR
		aCGGAAGGAGTCATGACCACCGCTTt		HRPNLTR	LKEHLTR	MGHHLKR

[a] Gene names are based on homology to previously characterized genes in *Arabidopsis thaliana*.
[b] Represents two duplicate copies ("a" and "b") that both perfectly match the ZFN target site.

Table 5-3. Supplemental Table; Whole plant transformation summary of the ZFN transgene targeting *DCL4.*

ZFN	Treatment during transformation	Ex-plants surviving selection	Transgene positive	Mutant alleles	Frame shift/ in-frame shift
DCL4a/DCL4b	Treatment	22	3	2	2 / 0
DCL4a/DCL4b	None	18	5	0	2 / 0

Figure 5-1. Detection of ZFN-induced mutations at a *GFP* transgene in soybean. A, The position of the OPEN ZFN target site is represented by a gray rectangle. The target sequence of both left and right ZFAs recognize a 9-bp sequence (indicated in bold). A strategy involving the restriction enzyme *Bcc*I and a PCR assay was used to enrich and identify mutated sequences. B, Amplicons from the PCR assay were cloned into pGem T-easy and subsequently amplified by colony PCR using the GFP-specific primers. The sequencing of PCR products revealed large deletions ranging from 27 to 71 bp. The enrichment of mutated *GFP* sequence was biased toward large deletions at the 5' region of the target site since the BccI recognition site CCATC was located on the left ZFA recognition sequence. Typically the restriction site is situated in the middle of the target site, as the majority of obtained indels are minor (1–10 bp) and occur in the spacer region.

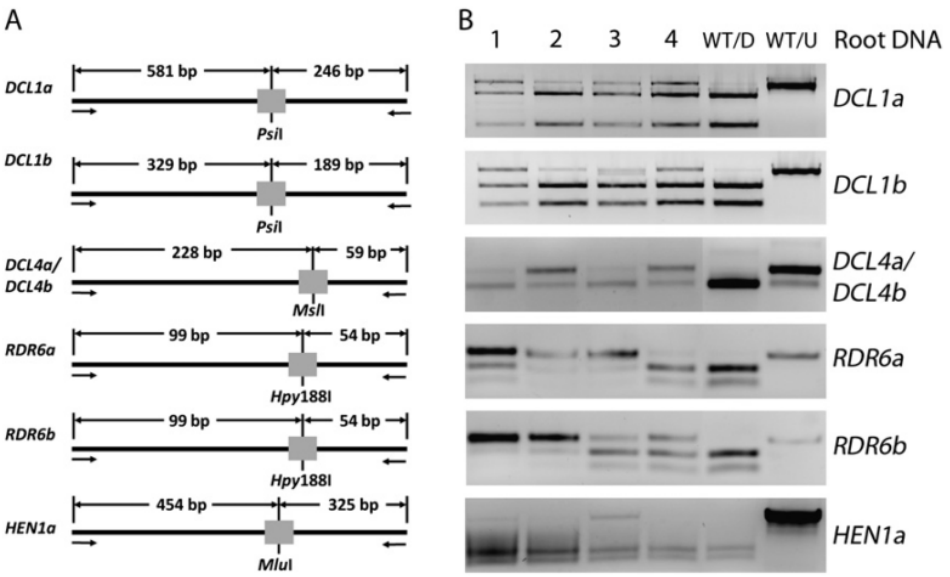

Figure 5-2. Detection of ZFN-induced mutations in soybean hairy-root tissue. A, A schematic strategy highlighting the restriction endonuclease PCR assays used to enrich mutated DNA sequences from soybean hairy-root tissue. Five ZFNs were designed to target seven genes, two ofwhich targeted duplicate copies of *DCL1* and *DCL4* (*DCL1a* and *DCL1b* were successfully targeted by one ZFN construct, but are shown in different sections because they were screened with paralog-specific primers; *DCL4a* and *DCL4b* were successfully targeted by one ZFN construct and are shown together because they were screened with a shared primer set). A single ZFN capable of targeting both copies of *RDR6* could not be identified. The closest match for both *RDR6a* and *RDR6b* was a target site that differed by 2 bp. Two different ZFN constructs were designed to target these respective sites. A ZFN was designed to target one of the duplicate copies of *HEN1*. B, The PCR products from four root samples, including an undigested (WT/U) and digested (WT/D) wild-type control, were separated on a 2% agarose gel. Lanes showing undigested products indicate a portion of the hairy-root cells having novel mutations at the restriction sites.

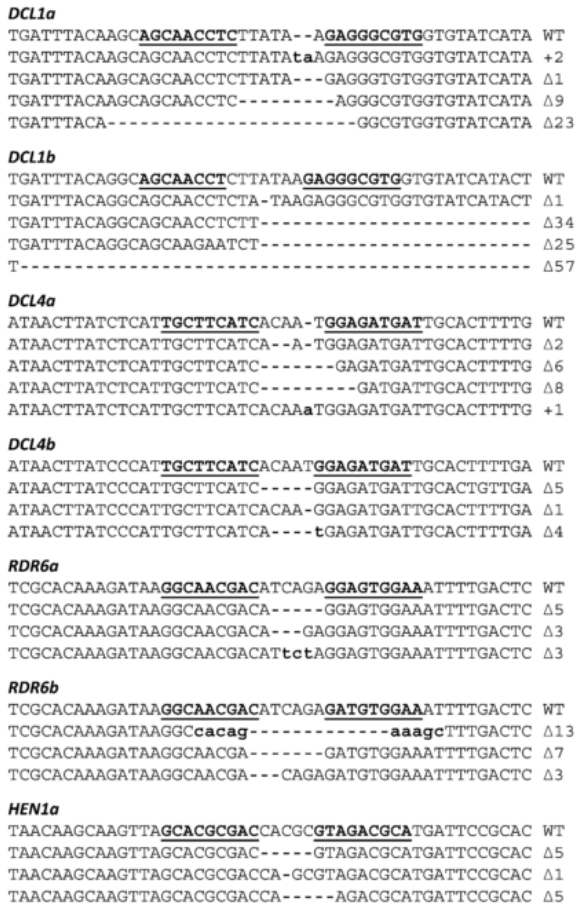

Figure 5-3. Sequences of induced ZFN mutations in soybean hairy-root tissue. The recovered mutated alleles from seven soybean endogenous genes are shown below their respective wild-type sequence. The bold and underlined sequences represent the ZFN target sites of each wild type. Deletions and insertions are indicated by dashes or lowercase letters, respectively. Single roots often produced multiple independent mutations. The numbers to the side indicate the type of mutation and how many nucleotides were involved.

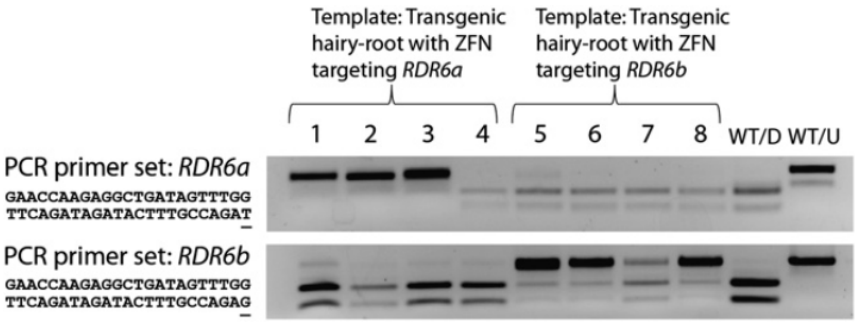

Figure 5-4. The mutagenic specificity of the RDR6a and RDR6b ZFN transgenes was assessed by performing PCR enrichment assays with gene-specific primers for each homeolog. The ZFN target sites of this gene pair differ by only two nucleotides, so this experiment was important to measure whether the gene-specific ZFNs could discrim- inate between the homeologous targets (Table I provides details on the target site differences of the two ZFN transgenes). The PCR enrichment assays are analogous to those shown in Figure 2. A sample of estradiol- induced hairy roots was targeted for mutagenesis using the ZFN transgene targeting RDR6a (roots 1–4) and the ZFN transgene targeting RDR6b (roots 5–8). Primer sets differing at a single nucleotide were designed to allow for homeolog-specific PCR amplification in these samples (the polymorphic nucleotide is underlined in the reverse primer sequences). The top section shows the PCR enrichment results when testing for mutagenesis of gene RDR6a and the bottom section shows the PCR enrichment results when testing for mutagenesis of gene RDR6b. In either case, an undigested top band indicates a mutation within the hairy-root sample. A digested nontransgenic root sample (WT/D) and an undigested nontransgenic root sample (WT/U) serve as controls. Seven out of the eight targeted hairy roots show the presence of the putative mutated top band; root 4 does not show this band, indicating that this sample either failed to transform or the ZFN failed to mutagenize any cells in this root. Faint top band shadows are observed in some samples for the nontargeted gene. Therefore, we cannot rule out the possibility that some mutations may have occurred at the nontargeted homeologous gene, albeit at a much lower frequency than the targeted gene.

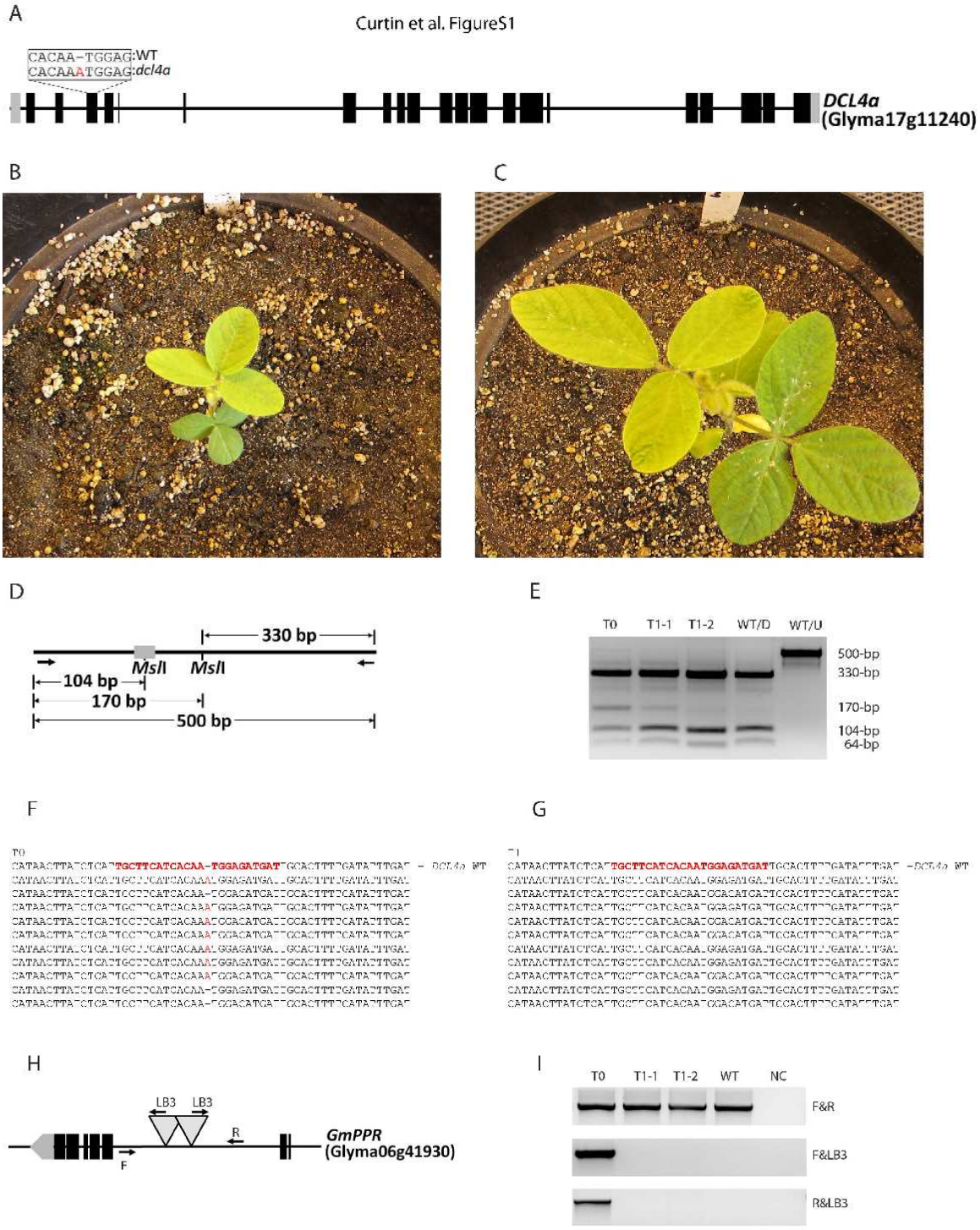

Figure 5-5. Supplemental Figure; Analysis of the ZFN-induced *dcl4a* mutation recovered from whole plant soybean. (A) Genomic structure of the *DCL4a* gene in soybean. The target site is highlighted by slotted lines with the box indication the target sequence of the ZFN-induced *dcl4a* indel and the wild-type (WT) comparison. (B-C) Comparison of the two T1 plants (T1-1 and T1-2) 48 days after germination. (D) A schematic of the strategy used to determine the

segregation frequency of the *dcl4a* mutation in the T1 plants. (E) A gel depicting a putative mutated sequence in the *DCL4a/dcl4a* T0 plant, and gel results from the T1 plants. The presence of 170-bp fragment would indicate a homozygous *dcl4a/dcl4a* plant, which was not observed in either generation. (F) Sequence data generated from the *DCL4a/dcl4a* T0 plant indicated the presence of both the mutated and wild-type allele. (G) Sequence data from the T1 plants were unable to recover the germinal transmission of the 1-bp insertion (data from T1-1 is shown), indicating that they are likely both wild-type *DCL4a/DCL4a.* (H) TAIL-PCR of the *DCL4a/dcl4a* T0 plant located the precise genomic location of the ZFN transgene. Two transgenes were found in the first intron of the gene encoding a *PENTATRICOPEPTIDE REPEAT-CONTAINING PROTEIN* (PPR) (Glyma06g41930). PCR using a primer specific to the PPR-gene (F&R) and a primer specific to the transgene left border confirmed the presence of the ZFN construct in the T0 plant and its absence in both T1 plants (F&LB3/R&LB3). A gel depicting the ZFN transgene status in the T0 and T1 plants is shown in (I), indicating the loss of the transgene in both T1 plants. A no template PCR negative control is also shown (NC).

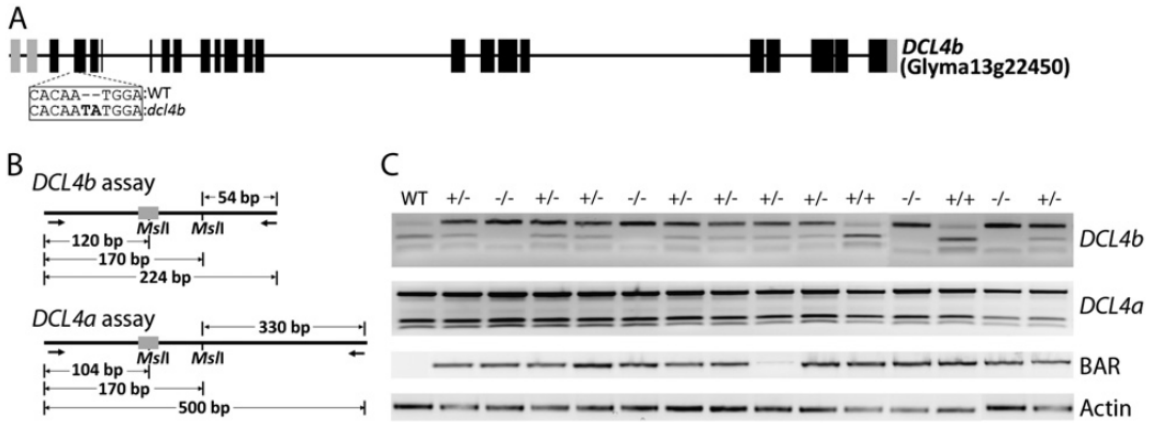

Figure 5-6. ZFN mutagenesis and heritability in whole-plant soybean. A, Genomic structure of the *DCL4b* gene in soybean. The target site is highlighted by dashed lines with the box indicating the ZFN-induced *dcl4b* mutation (2-bp insertion indicated in bold) relative to the wild type. B, A schematic of the strategy used to determine the segregation frequency of the homozygous and heterozygous mutations in the T1 progeny. C, A gel depicting the segregation of the mutation in 14 T1 plants is shown (PCR results for the remaining 10 plants are not shown). The top section shows the *DCL4b* genotype (+/+ indicates *DCL4b*/*DCL4b*, -/- indicates *dcl4b*/*dcl4b*, and +/- indicates heterozygous *DCL4b*/*dcl4b*). The middle section shows the genotyping result for the ZFN transgene (BAR amplicon) and the bottom section shows the PCR positive control (Actin amplicon). The induced *dcl4b* mutation segregated as expected in the 1:2:1 ratio. PCR confirmed that all T1 plants, with the exception of the wild-type control and one heterozygous plant (lane nine), harbored the ZFN transgene.

Figure 5-7. Supplemental Figure; ZFN-induced mutations recovered from *DCL4b* T0 and T1 whole plant soybean. (A) Sequences of the *DCL4b* and *dcl4b* alleles recovered from the T0 plant, indicating a heterozygous genotype. (B) Sequences from a *dcl4b*/*dcl4b* T1 plant. The sequence data indicate that the mutation with the two-base pair insertion was transmitted from the T0 plant. No wild-type allele was recovered from the 15 clones analyzed.

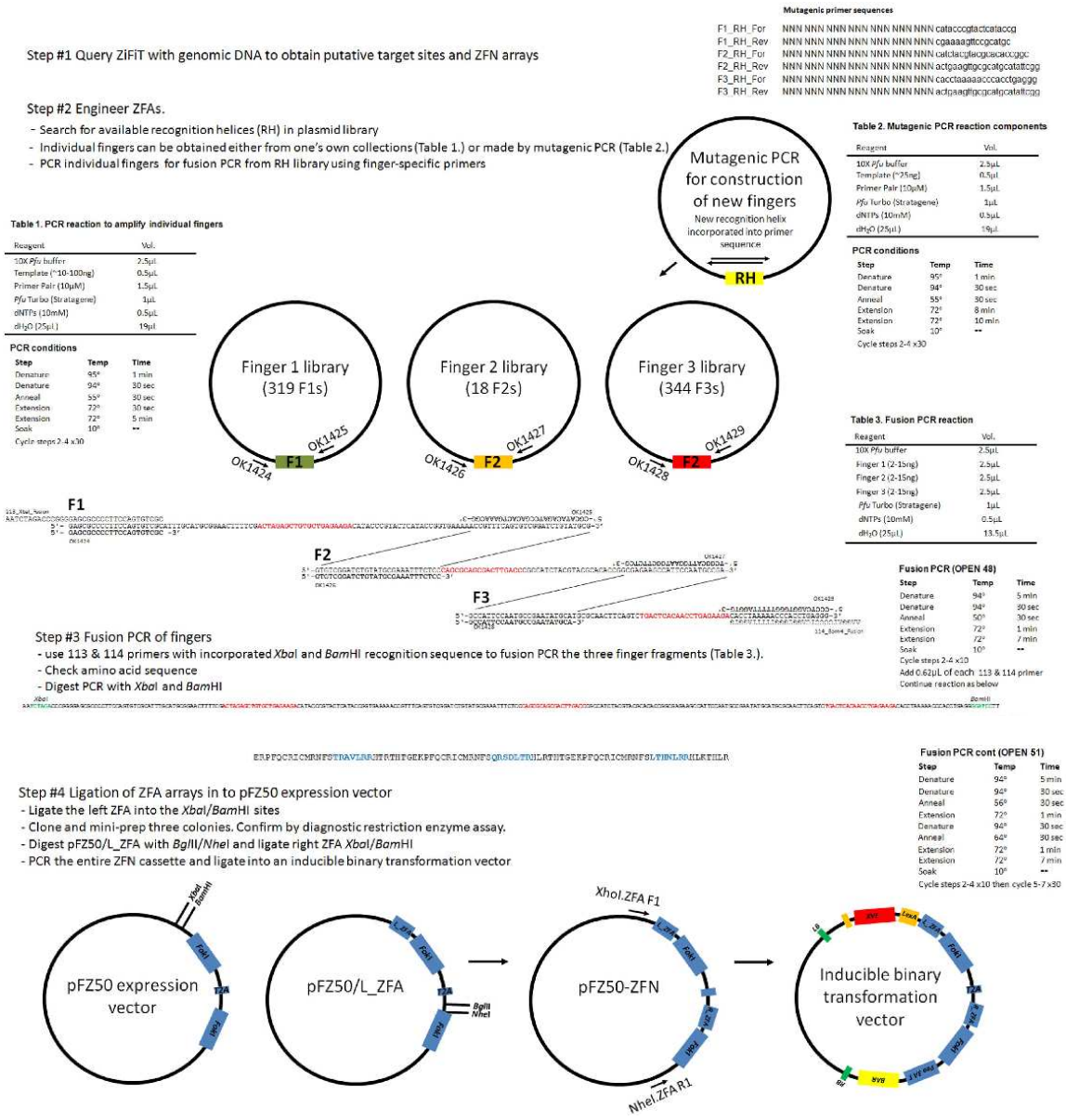

Figure 5-8. Supplemental Figure; The Context-dependent assembly (CoDA) method for engineering multi-finger arrays. Construction of CoDA ZFNs starts with the identification of target sites for a gene-of-interest (GOI). The ZiFIT web-based program (http://zifit.partners.org) is queried with a genomic target sequence. An output file catalogs the target sites of potential ZFNs that can be made by CoDA. Individual fingers can be engineered using a mutagenic PCR assay by incorporating the DNA sequence for each new recognition helix into the primer pair. New fingers are fused together by a fusion PCR assay to construct the zinc-finger arrays (ZFAs). Left and right ZFAs are ligated into an expression vector harboring the left and right *Fok*I nucleases. A final PCR is carried out using primers with incorporated *Xho*I and *Nhe*I sites. The *Xho*I/*Nhe*I ZFN cassette PCR product is then ligated into the *Xho*I/*Spe*I sites of the inducible binary transformation vector.

Contribution to the Work

Similar to the project described in Chapter 4, at the time we began this study, zinc-finger nucleases and meganucleases were the only programmable sequence-specific nucleases, and engineering zinc-finger nucleases to recognize a target sequence remained challenging. In this study, we designed zinc-finger nuclease pairs recognizing sequence within *GFP, DCL1a (*x2*), DCL1b (*x2*), DCL2a, DCL2b, DCL4a, DCL4b, RDR6a, RDR6b,* and *HEN1a* using context-dependent assembly (CoDA) (Sander et al., 2011). Results from the Sander et al., 2011 study provided potential amino-acid sequences for zinc-finger recognition helices that will recognize a desired target sequence. To facilitate cloning of zinc-finger nucleases, I developed a method for constructing multi-finger arrays (Qi et al., 2014). This method was used to build most all of the zinc-finger nuclease pairs used in this study (Table 5-1 and 5-2).

CHAPTER 6

DISSERTATION SUMMARY

Dissertation Summary and Contributions to the Field

In the next 30 years, as the world population increases from 7 to 9 billion people, there will be increasing pressure to efficiently produce food. One potential solution for this concern is to improve the genetics in crop plants such that yields are higher and losses from extreme climate or pathogens are minimized. To realize these solutions, plant researchers require effective tools and methods for modifying plant DNA. Currently, the ability to practice plant genome engineering is mostly limited to research labs that have the resources and expertise to regenerate plants, and to those labs that have invested significant time and effort in understanding the field. To expand the general use of genome engineering reagents, from experts in the field to all researchers interested in editing plant DNA, it is critical that we continue to refine existing technology and methods (or develop new technology) for efficiently editing plant DNA.

The focus of chapter 2 and appendix B was to determine if geminiviruses (plant DNA viruses) could serve as vectors for delivering genome engineering reagents. Here, we were the first to demonstrate that deconstructed geminiviruses can transiently deliver nucleases and repair templates for editing plant DNA. In addition to these proof-of-concept experiments, we also demonstrated that there were virus-specific features that promote repair of double-strand breaks by homologous recombination. Specifically, we found that replication of repair templates and pleiotropic activity of Rep and/or RepA enhance gene targeting. Finally, we show that tobacco leaf cells harboring a precise DNA change can divide and produce calli and plantlets.

Results from chapter 2 and appendix B demonstrate that DNA viruses may be effective tools for delivering genome engineering reagents to plants. However, the scope of this study was relatively limited: experiments were performed solely in tobacco leaf cells, the target for repair was a single locus (a *gus:nptII* transgene), a small zinc-finger nuclease monomer was delivered, and only one type of geminivirus was applied. To further develop geminivirus replicons technology, it will be critical to demonstrate their utility in different plant species and by modifying additional target loci. To demonstrate these points, we are currently using bean yellow dwarf virus replicons to deliver TALENs and repair templates to tomato plants. Preliminary results show that indeed bean yellow dwarf virus replicons replicate in tomato leaf cells, that they can be used to deliver reagents for modifying endogenous DNA, and that the modified cells retain the ability to divide into whole plants (data not shown).

Whereas these results are promising for the continued use of geminivirus replicons, it will important to understand the limitations of this technology. For example, additional studies should explore questions related to the impact replicons have on host cells (Do cells divide slower? Are off-target modifications more frequent? Are gross chromosomal abnormalities more frequent?). Furthermore, additional studies should explore how physical properties of geminivirus replicons impact genome editing (Does increasing the genome size of geminivirus replicons reduce

efficiency of genome editing? Does genome instability occur when replicons harbor large and repetitive TALE sequences?). Results these studies will define the role geminivirus replicons will play in the future for the genome engineering field.

In chapters 3, we characterize tobacco rattle virus (TRV) and its use as a vector for nuclease delivery in *Arabidopsis*. Unlike the geminivirus technology developed in chapter 2, TRV maintains the ability to move from cell to cell and to systemically infect plants. Therefore, we first characterized infection of TRV in *Arabidopsis* by tagging the RNA 2 genome with GFP. We found that TRV moved through the vascular system to systemically infect plants. The virus was also found outside vasculature cells, in rosettes, cauline leaves and floral tissue. We then assessed the ability for TRV to systemically deliver nucleases. Here, *Arabidopsis* plants were infected with TRV harboring a zinc-finger nuclease (Zif268:*Fok*I) and GFP. In systemic leaves expressing GFP, we found Zif268:*Fok*I-induced non-homologous end joining mutations, indicating that TRV can indeed deliver nucleases to *Arabidopsis.* Finally, we assessed next-generation seedlings for heritable mutations. After screening ~10,000 seedlings from 16 different infected parent plants, we found five seedlings with evidence of a recombination event at the GU.US target site.

There are several advantages with using TRV as a vector for delivery of genome engineering reagents, including a wide host range of plants, an ability to systemically infect plants, and a genome that is strictly RNA (circumventing government regulation around foreign DNA in plants). To realize the full potential of TRV for genome engineering, it is critical that the technology is applied in additional plant species, and is used to deliver different sequence-specific nucleases. Whereas the cargo capacity of TRV is too small for delivering a TALEN pair or Cas9 and gRNAs, it may be possible to deliver gRNAs to plants that are expressing Cas9.

The remaining chapters and appendices within this dissertation (chapters 4 and 5, and appendix A) describe our approach to stably integrate sequence-specific nucleases into plant genomes for knocking out gene function or deleting genes within *Arabidopsis* and soybean. At the time these studies began, zinc-finger nucleases were essentially the only class of sequence-specific nucleases that could be ‘rewired’ to cleave a target locus. Therefore, to begin our studies, we needed to develop methods for constructing and cloning functional zinc-finger nucleases. To this end, we developed novel cloning techniques (Qi et al., 2014) that were compatible with the CoDA method for zinc-finger array design (Sander et al., 2011). The resulting customized zinc-finger nucleases were then stably integrated within plant genomes and assessed for activity. Excitingly, majority of the nucleases tested (~75%) had detectable activity against their predicted target site. Furthermore, using these reagents we were able to (i) detect very large chromosomal deletions in *Arabidopsis* DNA (up to 9 Mb; chapter 4), (ii) knockout duplicated and single-copy genes within soybean (chapter 5), and (iii) knockout the *TZP* gene within *Arabidopsis* (appendix A). Results from these studies have expanded the utility of integrating sequence-specific

nucleases to additional plant species, to additional genome modifications, and to additional genes that have yet to be knocked out by conventional methods.

The main goal of most genome engineering projects is to recover a plant with a desired sequence change. With the development TALENs and the CRISPR/Cas system, designing nucleases to bind and cleave a target locus is no longer a bottleneck. Instead, researchers are faced with the challenge of implementing the technology, which includes the delivery of these reagents to plant cells, and the subsequent recovery of modified plants. Currently, researchers only have a handful of delivery options to choose from, and, for some plant species, there may only be one option. Therefore, the bottleneck in implementing genome engineering tools has shifted from the design of sequence-specific nucleases to the delivery of these reagents and subsequent recovery of modified plants.

To further advance the field of genome engineering, it is critical that we continue to develop new and existing delivery methods. Results from this dissertation address these concerns. We have expanded upon the utility of stable integration of genome engineering tools by (i) knocking out genes in additional plant species, including soybean, (ii) deleting large segments of chromosomal DNA in *Arabidopsis* leaf cells, and (iii) knocking out additional endogenous genes within the *Arabidopsis* genome. Furthermore, we have expanded upon the utility of RNA viruses for the delivery of nucleases in the model organism, *Arabidopsis.* And lastly, we have developed a completely novel delivery method that uses elements from DNA viruses to transiently deliver genome engineering reagents. Together, these results have not only advanced the utility of existing delivery methods, but they have introduced a new player in the game of genome engineering: the DNA virus.

References

Abbott, J.C., Barakate, A., Pinçon, G., Legrand, M., Lapierre, C., Mila, I., Schuch, W., and Halpin, C. (2002). Simultaneous suppression of multiple genes by single transgenes. Down-regulation of three unrelated lignin biosynthetic genes in tobacco. Plant Physiol. **128**: 844–853.

Adams, M.J., King, A.M.Q., and Carstens, E.B. (2013). Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2013). Arch. Virol. **158**: 2023–30.

Ainley, W.M. et al. (2013). Trait stacking via targeted genome editing. Plant Biotechnol. J. **11**: 1126–1134.

Akbergenov, R., Si-Ammour, A., Blevins, T., Amin, I., Kutter, C., Vanderschuren, H., Zhang, P., Gruissem, W., Meins, F., Hohn, T., and Pooggin, M.M. (2006). Molecular characterization of geminivirus-derived small RNAs in different plant species. Nucleic Acids Res. **34**: 462–71.

Alberter, B., Ali Rezaian, M., and Jeske, H. (2005). Replicative intermediates of Tomato leaf curl virus and its satellite DNAs. Virology **331**: 441–448.

Alonso, J.M. et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science **301**: 653–657.

Alvarez, J.P., Pekker, I., Goldshmidt, A., Blum, E., Amsellem, Z., and Eshed, Y. (2006). Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species. Plant Cell **18**: 1134–1151.

Ascencio-Ibáñez, J.T., Sozzani, R., Lee, T.J., Chu, T.M., Wolfinger, R.D., Cella, R., and Hanley-Bowdoin, L. (2008). Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol. **148**: 436–54.

Ayar, A., Wehrkamp-Richter, S., Laffaire, J.B., Goff, S., Levy, J., Chaignon, S., Salmi, H., Lepicard, A., Sallaud, C., Gallego, M.E., White, C.I., and Paul, W. (2012). Gene targeting in maize by somatic ectopic recombination. Plant Biotechnol. J. **11**: 305–314.

Bagewadi, B., Chen, S., Lal, S.K., Choudhury, N.R., and Mukherjee, S.K. (2004). PCNA interacts with Indian mung bean yellow mosaic virus rep and downregulates Rep activity. J. Virol. **78**: 11890–11903.

Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P. A., and Voytas, D.F. (2014). DNA replicons for plant genome engineering. Plant Cell **26**: 151–63.

Bar-Ziv, A., Levy, Y., Hak, H., Mett, A., Belausov, E., Citovsky, V., and Gafni, Y. (2012). The Tomato yellow leaf curl virus (TYLCV) V2 protein interacts with the host papain-like cysteine protease CYP1. Plant Signal. Behav. **7**: 983–989.

Bibikova, M., Beumer, K., Trautman, J.K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. Science **300**: 764–764.

Blanc, G. and Wolfe, K.H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell **16**: 1667–1678.

Blevins, T., Rajeswaran, R., Shivaprasad, P. V., Beknazariants, D., Si-Ammour, A., Park, H.S., Vazquez, F., Robertson, D., Meins, F., Hohn, T., and Pooggin, M.M. (2006). Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res. **34**: 6233–6246.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science **326**: 1509–1512.

Bogdanove, A.J. and Voytas, D.F. (2011). TAL Effectors: Customizable Proteins for DNA Targeting. Science **333**: 1843–1846.

Buchmann, R.C., Asad, S., Wolf, J.N., Mohannath, G., and Bisaro, D.M. (2009). Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. J. Virol. **83**: 5005–13.

Cai, C.Q. et al. (2009). Targeted transgene integration in plant cells using designed zinc finger nucleases. Plant Mol. Biol. **69**: 699–709.

Caplan, J. and Dinesh-Kumar, S.P. (2006). Using viral vectors to silence endogenous genes. Curr. Protoc. Microbiol. **Chapter 16**: Unit 16I.6.

Carvalho, C.M., Fontenelle, M.R., Florentino, L.H., Santos, A.A., Zerbini, F.M., and Fontes, E.P.B. (2008a). A novel nucleocytoplasmic traffic GTPase identified as a functional target of the bipartite geminivirus nuclear shuttle protein. Plant J. **55**: 869–880.

Carvalho, C.M., Santos, A.A., Pires, S.R., Rocha, C.S., Saraiva, D.I., Machado, J.P.B., Mattos, E.C., Fietto, L.G., and Fontes, E.P.B. (2008b). Regulated nuclear trafficking of rpL10A mediated by NIK1 represents a defense strategy of plant cells against virus. PLoS Pathog. **4**.

Carvalho, M. and Lazarowitz, S. (2004). Interaction of the movement protein NSP and the Arabidopsis acetyltransferase AtNSI is necessary for cabbage leaf curl geminivirus infection and pathogenicity. J. Virol.

Carvalho, M.F., Turgeon, R., and Lazarowitz, S.G. (2006). The geminivirus nuclear shuttle protein NSP inhibits the activity of AtNSI, a vascular-expressed Arabidopsis acetyltransferase regulated with the sink-to-source transition. Plant Physiol. **140**: 1317–1330.

Castillo, A.G., Kong, L.J., Hanley-Bowdoin, L., and Bejarano, E.R. (2004). Interaction between a geminivirus replication protein and the plant sumoylation system. J. Virol. **78**: 2758–2769.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. a, Somia, N. V, Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. **39**: e82.

Chandran, S.A., Levy, Y., Mett, A., Belausov, E., Ramakrishnan, U., and Gafni, Y. (2012). Mapping of functional region conferring nuclear localization and karyopherin -binding activity of the C2 protein of bhendi yellow vein mosaic virus. J. Gen. Virol. **93**: 1367–1374.

Chen, H., Zhang, Z., Teng, K., Lai, J., Zhang, Y., Huang, Y., Li, Y., Liang, L., Wang, Y., Chu, C., Guo, H., and Xie, Q. (2010). Up-regulation of LSB1/GDU3 affects geminivirus infection by activating the salicylic acid pathway. Plant J. **62**: 12–23.

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

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.-S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. **24**: 132–41.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics **186**: 757–61.

Christian, M., Qi, Y., Zhang, Y., and Voytas, D.F. (2013). Targeted mutagenesis of Arabidopsis thaliana using engineered TAL effector nucleases. G3 (Bethesda). **3**: 1697–705.

Clough, S.J. and Bent, A.F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. **16**: 735–743.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science **339**: 819–23.

Cooper, J.L. et al. (2008). TILLING to detect induced mutations in soybean. BMC Plant Biol. **8**: 9.

Cui, X., Li, G., Wang, D., Hu, D., and Zhou, X. (2005). A Begomovirus DNAbeta-encoded protein binds DNA, functions as a suppressor of RNA silencing, and targets the cell nucleus. J. Virol. **79**: 10764–10775.

Curtin, S.J. et al. (2011). Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiol. **156**: 466–73.

Curtin, S.J., Watson, J.M., Smith, N.A., Eamens, A.L., Blanchard, C.L., and Waterhouse, P.M. (2008). The roles of plant dsRNA-binding proteins in RNAi-like pathways. FEBS Lett. **582**: 2753–2760.

Datla, R.S.S., Hammerlindl, J.K., Pelcher, L.E., Crosby, W.L., and Selvaraj, G. (1991). A bifunctional fusion between beta-glucuronidase and neomycin phosphotransferase: a broad-spectrum marker enzyme for plants. Gene **101**: 239–246.

Delahodde, A., Goguel, V., Becam, A.M., Creusot, F., Perea, J., Banroques, J., and Jacq, C. (1989). Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. Cell **56**: 431–441.

Desbiez, C., David, C., Mettouchi, a, Laufs, J., and Gronenborn, B. (1995). Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. Proc. Natl. Acad. Sci. U. S. A. **92**: 5640–4.

Dogra, S.C., Eini, O., Rezaian, M.A., and Randles, J.W. (2009). A novel shaggy-like kinase interacts with the tomato leaf curl virus pathogenicity determinant C4 protein. Plant Mol. Biol. **71**: 25–38.

Donson, J., Morris-Krsinich, B.A., Mullineaux, P.M., Boulton, M.I., and Davies, J.W. (1984). A putative primer for second-strand DNA synthesis of maize streak virus is virion-associated. EMBO J. **3**: 3069–3073.

Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., Vandyk, J.K., and Bogdanove, A.J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: Tools for TAL effector design and target prediction. Nucleic Acids Res. **40**.

Egelkrout, E.M., Robertson, D., and Hanley-Bowdoin, L. (2001). Proliferating cell nuclear antigen transcription is repressed through an E2F consensus element and activated by geminivirus infection in mature leaves. Plant Cell **13**: 1437–1452.

Ellis, J. and Bernstein, A. (1989). Gene targeting with retroviral vectors: recombination by gene conversion into regions of nonhomology. Mol. Cell. Biol. **9**: 1621–1627.

Elmer, S. and Rogers, S.G. (1990). Selection for wild type size derivatiyes of tomato golden mosaic virus during systemic infection. Nucleic Acids Res. **18**: 2001–6.

Etessami, P., Watts, J., and Stanley, J. (1989). Size reversion of African cassava mosaic virus coat protein gene deletion mutants during infection of Nicotiana benthamiana. J. Gen. Virol. **70**: 277–289.

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

Eybishtz, A., Peretz, Y., Sade, D., Akad, F., and Czosnek, H. (2009). Silencing of a single gene in tomato plants resistant to tomato yellow leaf curl virus renders them susceptible to the virus. Plant Mol. Biol. **71**: 157–171.

Fauser, F., Roth, N., Pacher, M., Ilg, G., Sánchez-Fernández, R., Biesgen, C., and Puchta, H. (2012). In planta gene targeting. Proc. Natl. Acad. Sci. U. S. A. **109**: 7535–40.

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

Feng, Z. et al. (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**: 4632–7.

Florentino, L.H., Santos, A.A., Fontenelle, M.R., Pinheiro, G.L., Zerbini, F.M., Baracat-Pereira, M.C., and Fontes, E.P.B. (2006). A PERK-like receptor kinase interacts with the geminivirus nuclear shuttle protein and potentiates viral infection. J. Virol. **80**: 6648–6656.

Fontes, E.P.B., Santos, A.A., Luz, D.F., Waclawovsky, A.J., and Chory, J. (2004). The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. Genes Dev. **18**: 2545–2556.

Frischmuth, T. and Stanley, J. (1998). Recombination between viral DNA and the transgenic coat protein gene of African cassava mosaic geminivirus. J. Gen. Virol. **79 (Pt 5)**: 1265–71.

Fu, F., Sander, J.D., Maeder, M., Thibodeau-Beganny, S., Joung, J.K., Dobbs, D., Miller, L., and Voytas, D.F. (2009). Zinc Finger Database (ZiFDB): A repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. Nucleic Acids Res. **37**.

Fu, Y., Foden, J. a, Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. **31**: 822–6.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. **32**: 279–84.

Fulcher, N. and Sablowski, R. (2009). Hypersensitivity to DNA damage in plant stem cell niches. Proc. Natl. Acad. Sci. U. S. A. **106**: 20984–20988.

Fusaro, A.F., Matthew, L., Smith, N.A., Curtin, S.J., Dedic-Hagan, J., Ellacott, G.A., Watson, J.M., Wang, M.-B., Brosnan, C., Carroll, B.J., and Waterhouse, P.M. (2006). RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. EMBO Rep. **7**: 1168–1175.

Gardiner, W.E., Sunter, G., Brand, L., Elmer, J.S., Rogers, G., and Bisaro, D.M. (1988). Genetic analysis of tomato golden mosaic virus: the coat protein is not required for systemic spread or symptom development. EMBO J. **7**: 899–904.

Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468**: 67–71.

Gilbertson, R.L., Sudarshana, M., Jiang, H., Rojas, M.R., and Lucas, W.J. (2003). Limitations on geminivirus genome size imposed by plasmodesmata and virus-encoded movement protein: insights into DNA trafficking. Plant Cell **15**: 2578–2591.

Gleba, Y., Klimyuk, V., and Marillonnet, S. (2007). Viral vectors for the expression of proteins in plants. Curr. Opin. Biotechnol. **18**: 134–41.

Gleba, Y., Marillonnet, S., and Klimyuk, V. (2004). Engineering viral expression vectors for plants: the “full virus” and the “deconstructed virus” strategies. Curr. Opin. Plant Biol. **7**: 182–8.

Glick, E., Zrachya, A., Levy, Y., Mett, A., Gidoni, D., Belausov, E., Citovsky, V., and Gafni, Y. (2008). Interaction with host SGS3 is required for suppression of RNA silencing by tomato yellow leaf curl virus V2 protein. Proc. Natl. Acad. Sci. U. S. A. **105**: 157–161.

Gopal, P., Pravin Kumar, P., Sinilal, B., Jose, J., Kasin Yadunandam, A., and Usha, R. (2007). Differential roles of C4 and□C1 in mediating suppression of post-transcriptional gene silencing: Evidence for transactivation by the C2 of Bhendi yellow vein mosaic virus, a monopartite begomovirus. Virus Res. **123**: 9–18.

Govindarajulu, M., Elmore, J.M., Fester, T., and Taylor, C.G. (2008). Evaluation of constitutive viral promoters in transgenic soybean roots and nodules. Mol. Plant. Microbe. Interact. **21**: 1027–1035.

Guilinger, J.P., Thompson, D.B., and Liu, D.R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. **advance on**.

Guschin, D.Y., Waite, A.J., Katibah, G.E., Miller, J.C., Holmes, M.C., and Rebar, E.J. (2010). A rapid and general assay for monitoring endogenous gene modification. Methods Mol. Biol. **649**: 247–256.

Halley-Stott, R.P., Tanzer, F., Martin, D.P., and Rybicki, E.P. (2007). The complete nucleotide sequence of a mild strain of Bean yellow dwarf virus. Arch. Virol. **152**: 1237–40.

Hamilton, A.J. and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **286**: 950–952.

Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K., and Shiu, S.-H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiol. **148**: 993–1003.

Händel, E.-M., Alwin, S., and Cathomen, T. (2009). Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. Mol. Ther. **17**: 104–111.

Hanin, M., Volrath, S., Bogucki, a, Briker, M., Ward, E., and Paszkowski, J. (2001). Gene targeting in Arabidopsis. Plant J. **28**: 671–7.

Haseloff, J., Siemering, K.R., Prasher, D.C., and Hodge, S. (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. Proc. Natl. Acad. Sci. U. S. A. **94**: 2122–7.

Haun, W. et al. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnol. J.: 1–7.

Hayes, R., Petty, I., Coutts, R., and Buck, K. (1988). Gene amplification and expression in plants by a replicating geminivirus vector. Nature **334**: 179–182.

Hayes, R.J., Coutts, R.H.A., and Buck, K.W. (1989). Stability and expression of bacterial genes in replicating geminivirus vectors in plants. Nucleic Acids Res. **17**: 2391–2403.

Hefferon, K.L., Moon, Y.-S., and Fan, Y. (2006). Multi-tasking of nonstructural gene products is required for bean yellow dwarf geminivirus transcriptional regulation. FEBS J. **273**: 4482–94.

Hernandez-Garcia, C.M., Martinelli, A.P., Bouchard, R.A., and Finer, J.J. (2009). A soybean (Glycine max) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. Plant Cell Rep. **28**: 837–849.

Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J., and Gronenborn, B. (1995). Determination of the origin cleavage and joining domain of geminivirus Rep proteins. Nucleic Acids Res. **23**: 910–6.

Horns, T. and Jeske, H. (1991). Localization of abutilon mosaic virus (AbMV) DNA within leaf tissue by in situ hybridization. Virology **181**: 580–8.

Horváth, G. V, Pettkó-Szandtner, a, Nikovics, K., Bilgin, M., Boulton, M., Davies, J.W., Gutiérrez, C., and Dudits, D. (1998). Prediction of functional regions of the maize streak virus replication-associated proteins by protein-protein interaction analysis. Plant Mol. Biol. **38**: 699–712.

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

Huang, Z., Chen, Q., Hjelm, B., Arntzen, C., and Mason, H. (2009). A DNA replicon system for rapid high-level production of virus-like particles in plants. Biotechnol. Bioeng. **103**: 706–14.

Huang, Z., Phoolcharoen, W., Lai, H., Piensook, K., Cardineau, G., Zeitlin, L., Whaley, K.J., Arntzen, C.J., Mason, H.S., and Chen, Q. (2010). High-level rapid production of full-size monoclonal antibodies in plants by a single-vector DNA replicon system. Biotechnol. Bioeng. **106**: 9–17.

Hyten, D.L., Song, Q., Zhu, Y., Choi, I.-Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C., and Cregan, P.B. (2006). Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. U. S. A. **103**: 16666–16671.

Jacquier, A. and 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.

Jander, G. and Barth, C. (2007). Tandem gene arrays: a challenge for functional genomics. Trends Plant Sci. **12**: 203–210.

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

Kachroo, A., Fu, D.-Q., Havens, W., Navarre, D., Kachroo, P., and Ghabrial, S.A. (2008). An oleic acid-mediated pathway induces constitutive defense signaling and enhanced resistance to multiple pathogens in soybean. Mol. Plant. Microbe. Interact. **21**: 564–575.

Kaliappan, K., Choudhury, N.R., Suyal, G., and Mukherjee, S.K. (2012). A novel role for RAD54: this host protein modulates geminiviral DNA replication. FASEB J. **26**: 1142–60.

Kammann, M., Schalk, H.J., Matzeit, V., Schaefer, S., Schell, J., and Gronenborn, B. (1991). DNA replication of wheat dwarf virus, a geminivirus, requires two cis-acting signals. Virology **184**: 786–790.

Kapila, J., De Rycke, R., Van Montagu, M., and Angenon, G. (1997). An Agrobacterium-mediated transient gene expression system for intact leaves. Plant Sci. **124**: 101–108.

Kim, K. Il, Sunter, G., Bisaro, D.M., and Chung, I.S. (2007). Improved expression of recombinant GFP using a replicating vector based on Beet curly top virus in leaf-disks and infiltrated Nicotiana benthamiana leaves. Plant Mol. Biol. **64**: 103–12.

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

Kittelmann, K. and Jeske, H. (2008). Disassembly of African cassava mosaic virus. J. Gen. Virol. **89**: 2029–2036.

Kjemtrup, S., Sampson, K.S., Peele, C.G., Nguyen, L. V, Conkling, M. a, Thompson, W.F., and Robertson, D. (1998). Gene silencing from plant DNA carried by a Geminivirus. Plant J. **14**: 91–100.

Kong, L.-J. and Hanley-Bowdoin, L. (2002). A geminivirus replication protein interacts with a protein kinase and a motor protein that display different expression patterns during plant development and infection. Plant Cell **14**: 1817–1832.

Kong, L.J., Orozco, B.M., Roe, J.L., Nagar, S., Ou, S., Feiler, H.S., Durfee, T., Miller, A.B., Gruissem, W., Robertson, D., and Hanley-Bowdoin, L. (2000). A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. EMBO J. **19**: 3485–3495.

Koonin, E. V and Ilyina, T. V (1992). Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. J. Gen. Virol. **73 (Pt 10**: 2763–6.

Koornneef, M., Dellaert, L.W., and van der Veen, J.H. (1982). EMS- and radiation-induced mutation frequencies at individual loci in Arabidopsis thaliana (L.) Heynh. Mutat. Res. **93**: 109–123.

Krenz, B., Jeske, H., and Kleinow, T. (2012). The induction of stromule formation by a plant DNA-virus in epidermal leaf tissues suggests a novel intra- and intercellular macromolecular trafficking route. Front. Plant Sci. **3**: 291.

Krenz, B., Windeisen, V., Wege, C., Jeske, H., and Kleinow, T. (2010). A plastid-targeted heat shock cognate 70kDa protein interacts with the Abutilon mosaic virus movement protein. Virology **401**: 6–17.

Krysan, P.J., Young, J.C., and Sussman, M.R. (1999). T-DNA as an insertional mutagen in Arabidopsis. Plant Cell **11**: 2283–90.

Lacatus, G. and Sunter, G. (2009). The Arabidopsis PEAPOD2 transcription factor interacts with geminivirus AL2 protein and the coat protein promoter. Virology **392**: 196–202.

Lai, J., Chen, H., Teng, K., Zhao, Q., Zhang, Z., Li, Y., Liang, L., Xia, R., Wu, Y., Guo, H., and Xie, Q. (2009). RKP, a RING finger E3 ligase induced by BSCTV C4 protein, affects geminivirus infection by regulation of the plant cell cycle. Plant J. **57**: 905–917.

Laufs, J., Schumacher, S., Geisler, N., Jupin, I., and Gronenborn, B. (1995a). Identification of the nicking tyrosine of geminivirus Rep protein. FEBS Lett. **377**: 258–262.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S.G., Schell, J., and Gronenborn, B. (1995b). In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. Proc. Natl. Acad. Sci. U. S. A. **92**: 3879–3883.

Laufs, J., Wirtz, U., Kammann, M., Matzeit, V., Schaefer, S., Schell, J., Czernilofsky, A.P., Baker, B., and Gronenborn, B. (1990). Wheat dwarf virus Ac/Ds vectors: expression and excision of transposable elements introduced into various cereals by a viral replicon. Proc. Natl. Acad. Sci. U. S. A. **87**: 7752–7756.

Lee, H.J., Kim, E., and Kim, J.-S. (2010). Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res. **20**: 81–89.

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

Lehti-Shiu, M.D., Zou, C., Hanada, K., and Shiu, S.-H. (2009). Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. Plant Physiol. **150**: 12–26.

Lei, Y., Lee, C.-L., Joo, K.-I., Zarzar, J., Liu, Y., Dai, B., Fox, V., and Wang, P. (2011). Gene editing of human embryonic stem cells via an engineered baculoviral vector carrying zinc-finger nucleases. Mol. Ther. **19**: 942–50.

Lewis, J.D. and Lazarowitz, S.G. (2010). Arabidopsis synaptotagmin SYTA regulates endocytosis and virus movement protein cell-to-cell transport. Proc. Natl. Acad. Sci. U. S. A. **107**: 2491–2496.

Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M., and 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.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (2011). 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, M.H., Weeks, D.P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat. Biotechnol. **30**: 390–2.

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

Liu, L., Saunders, K., Thomas, C.L., Davies, J.W., and Stanley, J. (1999). Bean yellow dwarf virus RepA, but not rep, binds to maize retinoblastoma protein, and the virus tolerates mutations in the consensus binding motif. Virology **256**: 270–9.

Liu, L., van Tonder, T., Pietersen, G., Davies, J.W., and Stanley, J. (1997). Molecular characterization of a subgroup I geminivirus from a legume in South Africa. J. Gen. Virol. **78**: 2113–2117.

Liu, Y., Schiff, M., Marathe, R., and Dinesh-Kumar, S.P. (2002). Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. Plant J. **30**: 415–29.

Lloyd, A., Plaisier, C.L., Carroll, D., and Drews, G.N. (2005). Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **102**: 2232–7.

Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D., Holmes, M.C., and Naldini, L. (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat. Biotechnol. **25**: 1298–306.

Lozano-Durán, R., Rosas-Díaz, T., Gusmaroli, G., Luna, A.P., Taconnat, L., Deng, X.W., and Bejarano, E.R. (2011). Geminiviruses subvert ubiquitination by altering CSN-mediated derubylation of SCF E3 ligase complexes and inhibit jasmonate signaling in Arabidopsis thaliana. Plant Cell **23**: 1014–1032.

Lozano-Duran, R., Rosas-Diaz, T., Luna, A.P., and Bejarano, E.R. (2011). Identification of host genes involved in geminivirus infection using a reverse genetics approach. PLoS One **6**: e22383.

Lucy, A.P., Boulton, M.I., Davies, J.W., and Maule, A.J. (1996). Tissue specificity of Zea mays infection by maize streak virus. Mol. Plant. Microbe. Interact. **9**: 22–31.

Luna, A.P., Morilla, G., Voinnet, O., and Bejarano, E.R. (2012). Functional analysis of gene-silencing suppressors from tomato yellow leaf curl disease viruses. Mol. Plant. Microbe. Interact. **25**: 1294–306.

Luque, A., Sanz-Burgos, A.P., Ramirez-Parra, E., Castellano, M.M., and Gutierrez, C. (2002). Interaction of geminivirus Rep protein with replication factor C and its potential role during geminivirus DNA replication. Virology **302**: 83–94.

Maeder, M.L. et al. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol. Cell **31**: 294–301.

Maeder, M.L., Thibodeau-Beganny, S., Sander, J.D., Voytas, D.F., and Joung, J.K. (2009). Oligomerized pool engineering (OPEN): an "open-source" protocol for making customized zinc-finger arrays. Nat. Protoc. **4**: 1471–1501.

Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., and Zhu, J.-K. (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.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., van der Oost, J., and Koonin, E. V (2011). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. **9**: 467–477.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. **31**: 833–8.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via Cas9. Science **339**: 823–6.

Margis, R., Fusaro, A.F., Smith, N.A., Curtin, S.J., Watson, J.M., Finnegan, E.J., and Waterhouse, P.M. (2006). The evolution and diversification of Dicers in plants. FEBS Lett. **580**: 2442–2450.

Martín-Hernández, A.M. and Baulcombe, D.C. (2008). Tobacco rattle virus 16-kilodalton protein encodes a suppressor of RNA silencing that allows transient viral entry in meristems. J. Virol. **82**: 4064–71.

Marton, I., Zuker, A., Shklarman, E., Zeevi, V., Tovkach, A., Roffe, S., Ovadis, M., Tzfira, T., and Vainstein, A. (2010). Nontransgenic genome modification in plant cells. Plant Physiol. **154**: 1079–87.

Mathieu, M. et al. (2009). Establishment of a soybean (Glycine max Merr. L) transposon-based mutagenesis repository. Planta **229**: 279–289.

Matsumoto, T. et al. (2005). The map-based sequence of the rice genome. Nature **436**: 793–800.

Matzeit, V., Schaefer, S., Kammann, M., Schalk, H.J., Schell, J., and Gronenborn, B. (1991). Wheat dwarf virus vectors replicate and express foreign genes in cells of monocotyledonous plants. Plant Cell **3**: 247–58.

McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeted screening for induced mutations. Nat. Biotechnol. **18**: 455–457.

McGarry, R.C., Barron, Y.D., Carvalho, M.F., Hill, J.E., Gold, D., Cheung, E., Kraus, W.L., and Lazarowitz, S.G. (2003). A novel Arabidopsis acetyltransferase interacts with the geminivirus movement protein NSP. Plant Cell **15**: 1605–1618.

McGivern, D.R., Findlay, K.C., Montague, N.P., and Boulton, M.I. (2005). An intact RBR-binding motif is not required for infectivity of Maize streak virus in cereals, but is required for invasion of mesophyll cells. J. Gen. Virol. **86**: 797–801.

Men, A.E., Laniya, T.S., Searle, I.R., Iturbe-Ormaetxe, I., Gresshoff, I., Jiang, Q., Carroll, B.J., and Gresshoff, P.M. (2002). Fast Neutron Mutagenesis of Soybean (Glycine soja L.) Produces a Supernodulating Mutant Containing a Large Deletion in Linkage Group H. Genome Lett. **1**: 147–155.

Messing, J., Bharti, A.K., Karlowski, W.M., Gundlach, H., Kim, H.R., Yu, Y., Wei, F., Fuks, G., Soderlund, C.A., Mayer, K.F.X., and Wing, R.A. (2004). Sequence composition and genome organization of maize. Proc. Natl. Acad. Sci. U. S. A. **101**: 14349–14354.

Meyer, J.D.F., Silva, D.C.G., Yang, C., Pedley, K.F., Zhang, C., van de Mortel, M., Hill, J.H., Shoemaker, R.C., Abdelnoor, R. V, Whitham, S.A., and Graham, M.A. (2009). Identification and analyses of candidate genes for rpp4-mediated resistance to Asian soybean rust in soybean. Plant Physiol. **150**: 295–307.

Meyers, B.C., Kaushik, S., and Nandety, R.S. (2005). Evolving disease resistance genes. Curr. Opin. Plant Biol. **8**: 129–134.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell **15**: 809–834.

Miller, J.C. et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. **29**: 143–8.

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

Mor, T.S., Moon, Y.S., Palmer, K.E., and Mason, H.S. (2003). Geminivirus vectors for high-level expression of foreign proteins in plant cells. Biotechnol. Bioeng. **81**: 430–7.

Morin, S., Ghanim, M., Zeidan, M., Czosnek, H., Verbeek, M., and van den Heuvel, J.F. (1999). A GroEL homologue from endosymbiotic bacteria of the whitefly Bemisia tabaci is implicated in the circulative transmission of tomato yellow leaf curl virus. Virology **256**: 75–84.

Moscou, M.J. and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science **326**: 1501.

Muangsan, N., Beclin, C., Vaucheret, H., and Robertson, D. (2004). Geminivirus VIGS of endogenous genes requires SGS2/SDE1 and SGS3 and defines a new branch in the genetic pathway for silencing in plants. Plant J. **38**: 1004–1014.

Muangsan, N. and Robertson, D. (2004). Geminivirus vectors for transient gene silencing in plants. Methods Mol. Biol. **265**: 101–15.

Muller, P.Y., Janovjak, H., Miserez, A.R., and Dobbie, Z. (2002). Processing of gene expression data generated by quantitative real-time RT-PCR. Biotechniques **32**: 2–7.

Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., and Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. **39**: 9283–9293.

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

Noueiry, A.O., Lucas, W.J., and Gilbertson, R.L. (1994). Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. Cell **76**: 925–932.

Offringa, R., Franke-van Dijk, M.E., De Groot, M.J., van den Elzen, P.J., and Hooykaas, P.J. (1993). Nonreciprocal homologous recombination between Agrobacterium transferred DNA and a plant chromosomal locus. Proc. Natl. Acad. Sci. U. S. A. **90**: 7346–7350.

Offringa, R., Groot, M. De, Haagsman, H.J., Does, M.P., van den Elzen, P.J.M., and Hooykaas, P.J.J. (1990). Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation. EMBO J. **9**: 3077–3084.

Ohnesorge, S. and Bejarano, E.R. (2009). Begomovirus coat protein interacts with a small heat-shock protein of its transmission vector (Bemisia tabaci). Insect Mol. Biol. **18**: 693–703.

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

Padidam, M., Sawyer, S., and Fauquet, C.M. (1999). Possible emergence of new geminiviruses by frequent recombination. Virology **265**: 218–225.

Palmer, K.E., Thomson, J.A., and Rybicki, E.P. (1999). Generation of maize cell lines containing autonomously replicating maize streak virus-based gene vectors. Arch. Virol. **144**: 1345–60.

Park, J., Lee, H.J., Cheon, C.I., Kim, S.H., Hur, Y.S., Auh, C.K., Im, K.H., Yun, D.J., Lee, S., and Davis, K.R. (2011). The arabidopsis thaliana homeobox gene ATHB12 is involved in symptom development caused by geminivirus infection. PLoS One **6**.

Paszkowski, J., Baur, M., Bogucki, a, and Potrykus, I. (1988). Gene targeting in plants. EMBO J. **7**: 4021–6.

De Pater, S., Neuteboom, L.W., Pinas, J.E., Hooykaas, P.J.J., and van der Zaal, B.J. (2009). ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. Plant Biotechnol. J. **7**: 821–35.

Peele, C., Jordan, C. V., Muangsan, N., Turnage, M., Egelkrout, E., Eagle, P., Hanley-Bowdoin, L., and Robertson, D. (2001). Silencing of a meristematic gene using geminivirus-derived vectors. Plant J. **27**: 357–366.

Petolino, J.F. and Arnold, N.L. (2009). Whiskers-mediated maize transformation. Methods Mol. Biol. **526**: 59–67.

Petolino, J.F., Worden, A., Curlee, K., Connell, J., Strange Moynahan, T.L., Larsen, C., and Russell, S. (2010). Zinc finger nuclease-mediated transgene deletion. Plant Mol. Biol. **73**: 617–28.

Pham, A.-T., Lee, J.-D., Shannon, J.G., and Bilyeu, K.D. (2010). Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biol. **10**: 195.

Piganeau, M. et al. (2013). Cancer translocations in human cells induced by zinc finger and TALE nucleases. Genome Res. **23**: 1182–1193.

Pilartz, M. and Jeske, H. (1992). Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. Virology **189**: 800–802.

Piroux, N., Saunders, K., Page, A., and Stanley, J. (2007). Geminivirus pathogenicity protein C4 interacts with Arabidopsis thaliana shaggy-related protein kinase AtSKeta, a component of the brassinosteroid signalling pathway. Virology **362**: 428–440.

Potts, P.R., Porteus, M.H., and Yu, H. (2006). Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. EMBO J. **25**: 3377–3388.

Preiss, W. and Jeske, H. (2003). Multitasking in replication is common among geminiviruses. J. Virol. **77**: 2972–2980.

Prieto, J., Redondo, P., Padro, D., Arnould, S., Epinat, J.C., Paques, F., Blanco, F.J., and Montoya, G. (2007). The C-terminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage. Nucleic Acids Res. **35**: 3262–3271.

Puchta, H. (1999). Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. Genetics **152**: 1173–81.

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., Dujon, B., and 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**: 5055–60.

Qi, Y., Li, X., Zhang, Y., Starker, C.G., Baltes, N.J., Zhang, F., Sander, J.D., Reyon, D., Joung, J.K., and Voytas, D.F. (2013a). Targeted Deletion and Inversion of Tandemly Arrayed Genes in Arabidopsis thaliana Using Zinc Finger Nucleases. G3 Genes| Genomes| Genet.

Qi, Y., Starker, C.G., Zhang, F., Baltes, N.J., and Voytas, D.F. (2014). Tailor-made mutations in Arabidopsis using zinc-finger nucleases. Arab. Protoc. **1062**: 193–209.

Qi, Y., Zhang, Y., Zhang, F., Baller, J. A., Cleland, S.C., Ryu, Y., Starker, C.G., and Voytas, D.F. (2013b). Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. Genome Res. **23**: 547–54.

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

Ramirez, C.L. et al. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. Nat. Methods **5**: 374–5.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. Cell **154**: 1380–1389.

Ratcliff, F., Martin-Hernandez, A.M., and Baulcombe, D.C. (2001). Tobacco rattle virus as a vector for analysis of gene function by silencing. Plant J. **25**: 237–245.

Regnard, G.L., Halley-Stott, R.P., Tanzer, F.L., Hitzeroth, I.I., and Rybicki, E.P. (2010). High level protein expression in plants through the use of a novel autonomously replicating geminivirus shuttle vector. Plant Biotechnol. J. **8**: 38–46.

Reiss, B., Schubert, I., Köpchen, K., Wendeler, E., Schell, J., and Puchta, H. (2000). RecA stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not gene targeting, in plants transformed by Agrobacterium. Proc. Natl. Acad. Sci. U. S. A. **97**: 3358–63.

Reyon, D., Kirkpatrick, J.R., Sander, J.D., Zhang, F., Voytas, D.F., Joung, J.K., Dobbs, D., and Coffman, C.R. (2011). ZFNGenome: a comprehensive resource for locating zinc finger nuclease target sites in model organisms. BMC Genomics **12**: 83.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., and Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol. **30**: 460–465.

Richter, K.S., Kleinow, T., and Jeske, H. (2014). Somatic homologous recombination in plants is promoted by a geminivirus in a tissue-selective manner. Virology **452-453**: 287–96.

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

Rocha, C.S., Santos, A.A., Machado, J.P.B., and Fontes, E.P.B. (2008). The ribosomal protein L10/QM-like protein is a component of the NIK-mediated antiviral signaling. Virology **380**: 165–169.

Ronald, P. (2011). Plant genetics, sustainable agriculture and global food security. Genetics **188**: 11–20.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol. Biol. **53**: 247–259.

Rothkamm, K., Krüger, I., Thompson, L.H., and Lobrich, M. (2003). Pathways of DNA double-strand break repair during the mammalian cell cycle. Mol. Cell. Biol. **23**: 5706–5715.

Russell, D.W. and Hirata, R.K. (1998). Human gene targeting by viral vectors. Nat. Genet. **18**: 325–330.

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

San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of eukaryotic homologous recombination. Annu. Rev. Biochem. **77**: 229–257.

Sánchez-Durán, M.A., Dallas, M.B., Ascencio-Ibañez, J.T., Reyes, M.I., Arroyo-Mateos, M., Ruiz-Albert, J., Hanley-Bowdoin, L., and Bejarano, E.R. (2011). Interaction between Geminivirus Replication Protein and the SUMO-Conjugating Enzyme Is Required for Viral Infection. J. Virol. **85**: 9789–9800.

Sander, J.D. et al. (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat. Methods **8**: 67–69.

Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F., and Dobbs, D. (2007). Zinc Finger Targeter (ZiFiT): An engineered zinc finger/target site design tool. Nucleic Acids Res. **35**.

Saunders, K., Lucy, A., and Stanley, J. (1992). RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. Nucleic Acids Res. **20**: 6311–6315.

Schmutz, J. et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature **463**: 178–183.

Selth, L.A., Dogra, S.C., Rasheed, M.S., Healy, H., Randles, J.W., and Rezaian, M.A. (2005). A NAC domain protein interacts with tomato leaf curl virus replication accessory protein and enhances viral replication. Plant Cell **17**: 311–325.

Sessions, A. et al. (2002). A high-throughput Arabidopsis reverse genetics system. Plant Cell **14**: 2985–2994.

Settlage, S.B., Miller, A.B., Gruissem, W., and Hanley-Bowdoin, L. (2001). Dual interaction of a geminivirus replication accessory factor with a viral replication protein and a plant cell cycle regulator. Virology **279**: 570–576.

Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., Zhang, K., Liu, J., Voytas, D.F., Zheng, X., Zhang, Y., and Gao, C. (2013a). Rapid and efficient gene modification in rice and Brachypodium using TALENs. Mol. Plant **6**: 1365–8.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Lian, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L., and Gao, C. (2013b). Targeted genome modification of crop plants using the CRISPR-Cas system. Nat. Biotechnol. **31**: 686–8.

Sharma, P. and Ikegami, M. (2010). Tomato leaf curl Java virus V2 protein is a determinant of virulence, hypersensitive response and suppression of posttranscriptional gene silencing. Virology **396**: 85–93.

Shen, Q., Liu, Z., Song, F., Xie, Q., Hanley-Bowdoin, L., and Zhou, X. (2011). Tomato SlSnRK1 protein interacts with and phosphorylates βC1, a pathogenesis protein encoded by a geminivirus β-satellite. PLANT Physiol. **157**: 1394–1406.

Shen, W. and Hanley-Bowdoin, L. (2006). Geminivirus infection up-regulates the expression of two Arabidopsis protein kinases related to yeast SNF1- and mammalian AMPK-activating kinases. Plant Physiol. **142**: 1642–1655.

Shen, W.H. and Hohn, B. (1994). Amplification and expression of the β-glucuronidase gene in maize plants by vectors based on maize streak virus. Plant J. **5**: 227–236.

Shen, W.H. and Hohn, B. (1991). Mutational Analysis of the Small lntergenic Region of Maize Streak Virus. Virology **183**: 721–30.

Shiu, S.-H., Karlowski, W.M., Pan, R., Tzeng, Y.-H., Mayer, K.F.X., and Li, W.-H. (2004). Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. Plant Cell **16**: 1220–1234.

Shukla, V.K. et al. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature **459**: 437–41.

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

Singh, D.K., Malik, P.S., Choudhury, N.R., and Mukherjee, S.K. (2008). MYMIV replication initiator protein (Rep): Roles at the initiation and elongation steps of MYMIV DNA replication. Virology **380**: 75–83.

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

Sollu, C., Pars, K., Cornu, T.I., Thibodeau-Beganny, S., Maeder, M.L., Joung, J.K., Heilbronn, R., and Cathomen, T. (2010). Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. Nucleic Acids Res. **38**: 8269–8276.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A., and Lewis, S. (2002). The generic genome browser: a building block for a model organism system database. Genome Res. **12**: 1599–1610.

Stenger, D.C., Revington, G.N., Stevenson, M.C., and Bisaro, D.M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. Proc. Natl. Acad. Sci. U. S. A. **88**: 8029–33.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J. a (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature **507**: 62–67.

Stewart, C.N. and Via, L.E. (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques **14**: 748–750.

Suárez-López, P. and Gutiérrez, C. (1997). DNA replication of wheat dwarf geminivirus vectors: effects of origin structure and size. Virology **227**: 389–99.

Suyal, G., Mukherjee, S.K., and Choudhury, N.R. (2013a). The host factor RAD51 is involved in mungbean yellow mosaic India virus (MYMIV) DNA replication. Arch. Virol. **158**: 1931–1941.

Suyal, G., Mukherjee, S.K., Srivastava, P.S., and Choudhury, N.R. (2013b). Arabidopsis thaliana MCM2 plays role(s) in mungbean yellow mosaic India virus (MYMIV) DNA replication. Arch. Virol. **158**: 981–992.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell **33**: 25–35.

Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F., and Vignali, D.A.A. (2004). Correction of multi-gene deficiency in vivo using a single "self-cleaving" 2A peptide-based retroviral vector. Nat. Biotechnol. **22**: 589–594.

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., and Takeda, S. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. EMBO J. **17**: 5497–5508.

Takeuchi, R., Choi, M., and Stoddard, B.L. (2014). Redesign of extensive protein-DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization. Proc. Natl. Acad. Sci. U. S. A. **111**: 4061–6.

Taylor, G.K., Petrucci, L.H., Lambert, A.R., Baxter, S.K., Jarjour, J., and Stoddard, B.L. (2012). LAHEDES: The LAGLIDADG homing endonuclease database and engineering server. Nucleic Acids Res. **40**.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408**: 796–815.

Timmermans, M.C., Das, O.P., and Messing, J. (1992). Trans replication and high copy numbers of wheat dwarf virus vectors in maize cells. Nucleic Acids Res. **20**: 4047–54.

Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K., and Voytas, D.F. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature **459**: 442–5.

Trejo-Saavedra, D.L., Vielle-Calzada, J.P., and Rivera-Bustamante, R.F. (2009). The infective cycle of Cabbage leaf curl virus (CaLCuV) is affected by CRUMPLED LEAF (CRL) gene in Arabidopsis thaliana. Virol. J. **6**: 169.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J. a, Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat. Biotechnol.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat. Rev. Genet. **11**: 636–46.

Vanderschuren, H., Stupak, M., Fütterer, J., Gruissem, W., and Zhang, P. (2007). Engineering resistance to geminiviruses--review and perspectives. Plant Biotechnol. J. **5**: 207–20.

Vanitharani, R., Chellappan, P., Pita, J.S., and Fauquet, C.M. (2004). Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. J. Virol. **78**: 9487–9498.

Varma, A. and Malathi, V.G. (2003). Emerging geminivirus problems: A serious threat to crop production. Ann. Appl. Biol. **142**: 145–164.

Waigmann, E., Ueki, S., Trutnyeva, K., and Citovsky, V. (2004). The Ins and Outs of Nondestructive Cell-to-Cell and Systemic Movement of Plant Viruses. CRC. Crit. Rev. Plant Sci. **23**: 195–250.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell **153**: 910–8.

Wang, Q. and Taylor, M.W. (1993). Correction of a deletion mutant by gene targeting with an adenovirus vector. Mol. Cell. Biol. **13**: 918–927.

Wang, Y., Dang, M., Hou, H., Mei, Y., Qian, Y., and Zhou, X. (2014). Identification of an RNA Silencing Suppressor encoded by a Mastrevirus. J. Gen. Virol.

Ward, A., Etessami, P., and Stanley, J. (1988). Expression of a bacterial gene in plants mediated by infectious geminivirus DNA. EMBO J. **7**: 1583–1587.

Wassenegger, M. and Krczal, G. (2006). Nomenclature and functions of RNA-directed RNA polymerases. Trends Plant Sci. **11**: 142–151.

Wendt, T., Holm, P.B., Starker, C.G., Christian, M., Voytas, D.F., Brinch-Pedersen, H., and Holme, I.B. (2013). TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. Plant Mol. Biol. **83**: 279–85.

West, C.E., Waterworth, W.M., Jiang, Q., and Bray, C.M. (2000). Arabidopsis DNA ligase IV is induced by gamma-irradiation and interacts with an Arabidopsis homologue of the double strand break repair protein XRCC4. Plant J. **24**: 67–78.

Woody, S.T., Austin-Phillips, S., Amasino, R.M., and Krysan, P.J. (2007). The WiscDsLox T-DNA collection: An arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. J. Plant Res. **120**: 157–165.

Wright, D.A., Thibodeau-Beganny, S., Sander, J.D., Winfrey, R.J., Hirsh, A.S., Eichtinger, M., Fu, F., Porteus, M.H., Dobbs, D., Voytas, D.F., and Joung, J.K. (2006). Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nat. Protoc. **1**: 1637–1652.

Wright, D.A., Townsend, J.A., Winfrey, R.J., Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall, B.D., Jondle, M.D., and Voytas, D.F. (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J. **44**: 693–705.

Wright, E.A., Heckel, T., Groenendijk, J., Davies, J.W., and Boulton, M.I. (1997). Splicing features in maize streak virus virion- and complementary-sense gene expression. Plant J. **12**: 1285–97.

Wu, X. et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol.: 1–9.

Xie, Q., Sanz-Burgos, A.P., Guo, H., Garcia, J.A., and Gutierrez, C. (1999). GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. Plant Mol. Biol. **39**: 647–56.

Yang, J.-Y., Iwasaki, M., Machida, C., Machida, Y., Zhou, X., and Chua, N.-H. (2008). betaC1, the pathogenicity factor of TYLCCNV, interacts with AS1 to alter leaf development and suppress selective jasmonic acid responses. Genes Dev. **22**: 2564–2577.

Yang, L.P., Fang, Y.Y., An, C.P., Dong, L., Zhang, Z.H., Chen, H., Xie, Q., and Guo, H.S. (2013). C2-mediated decrease in DNA methylation, accumulation of siRNAs, and increase in expression for genes involved in defense pathways in plants infected with beet severe curly top virus. Plant J. **73**: 910–917.

Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. (2006). HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2’ OH of the 3' terminal nucleotide. Nucleic Acids Res. **34**: 667–675.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X., Pierick, C.J., Dobbs, D., Peterson, T., Joung, J.K., and Voytas, D.F. (2010). High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proc. Natl. Acad. Sci. U. S. A. **107**: 12028–33.

Zhang, J., Dong, J., Xu, Y., and Wu, J. (2012). V2 protein encoded by Tomato yellow leaf curl China virus is an RNA silencing suppressor. Virus Res. **163**: 51–58.

Zhang, S., Raina, S., Li, H., Li, J., Dec, E., Ma, H., Huang, H., and Fedoroff, N. V. (2003). Resources for targeted insertional and deletional mutagenesis in Arabidopsis. Plant Mol. Biol. **53**: 133–150.

Zhang, X. and Mason, H. (2006). Bean Yellow Dwarf Virus replicons for high-level transgene expression in transgenic plants and cell cultures. Biotechnol. Bioeng. **93**: 271–9.

Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker, C.G., Bogdanove, A.J., and Voytas, D.F. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol. **161**: 20–7.

Zhang, Z. et al. (2011). BSCTV C2 attenuates the degradation of SAMDC1 to suppress DNA methylation-mediated gene silencing in Arabidopsis. Plant Cell **23**: 273–288.

Zhou, Y., Rojas, M.R., Park, M.-R., Seo, Y.-S., Lucas, W.J., and Gilbertson, R.L. (2011). Histone H3 Interacts and Colocalizes with the Nuclear Shuttle Protein and the Movement Protein of a Geminivirus. J. Virol. **85**: 11821–11832.

Zrachya, A., Glick, E., Levy, Y., Arazi, T., Citovsky, V., and Gafni, Y. (2007). Suppressor of RNA silencing encoded by Tomato yellow leaf curl virus-Israel. Virology **358**: 159–165.

Zuo, J., Niu, Q.W., and Chua, N.H. (2000). Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. Plant J. **24**: 265–73.

APPENDIX A

TARGETED KNOCKOUT OF THE *ARABIDOPSIS THALIANA* GENES *TSPO, JMJC* AND *TZP* BY STABLE INTEGRATION OF ZINC-FINGER NUCLEASES

Summary of Preliminary Data

The focus of this dissertation is on the development and application of methods for precise genome engineering in whole plants. One of the more commonly-used methods for delivering genome engineering reagents is by stable integration into the plants genome. This section aims to apply and expand upon this method by disrupting the function of several *Arabidopsis* genes, including tandem zinc knuckle/PLU3 domain encoding (*TZP;* thought to be involved in the growth and development); jumonji C domain-containing protein (*JMJC*; a histone demethylase thought to be involved in control of the circadian clock); and translocator protein (*TSPO*; thought to be involved in stress regulation). Notably, these genes have yet to be knocked out using conventional methods such as T-DNA insertion (Krysan et al., 1999; Alonso et al., 2003) or chemical mutagenesis (McCallum et al., 2000). Here, we describe preliminary data for this study which includes the design and testing of zinc-finger nucleases targeting *TZP, JMJC*, and *TSPO* the generation of transgenic plants harboring the most active zinc-finger nuclease pair, and the isolation of a *tzp* mutant plant.

In addition to knocking out these genes, we integrated our reagents into different mutant backgrounds (*ku70* and *lig4*) to explore how DNA repair-pathway mutants affect the mutation profile and the probability of finding a mutant plant. The likelihood of creating a desired genetic change is not only due to the intrinsic properties of sequence-specific nucleases, but also from the conditions within the target cells (e.g., cell cycle phase and availability of DNA repair proteins). One method to increase the likelihood of creating a desired modification is to manipulate DNA repair pathways. For example, deleting the host plant's *SMC6* gene, which codes for proteins involved in sister chromatid recombination (Potts et al., 2006), resulted in an increased frequency of targeted mutagenesis by ~16 fold (Qi et al., 2013b). Furthermore, deleting host genes involved in classical non-homologous end joining (*KU70* and *LIG4*) shifted DNA break repair to either homologous recombination or the alternative non-homologous end joining pathway (Qi et al., 2013b) (characterized by repair that uses short stretches of microhomology, usually 2-25 bp, on either side of the break).

Results and Discussion

We began our study by engineering zinc-finger nucleases to recognize and cleave DNA within the coding sequence of three genes, *TZP, TSPO* and *JMJC*. Potential target sites within *TZP, JMJC* and *TSPO* (containing a high frequency of tandem GNN sequences), were identified using ZiFiT software (Sander et al., 2011). After finding several potential zinc-finger nuclease target sites, we determined the amino acid composition of the corresponding recognition zinc-finger arrays by using the context-dependent assembly method (CoDA) (Sander et al., 2011)

(Table A-1). Using these methods, we were able to design, *in silico*, four zinc-finger nuclease pairs recognizing sites within each of our target genes (two pairs were designed for *JMJC*).

To engineer plasmids containing our *in-silico*-designed three-finger arrays, we developed a method for cloning zinc-fingers and their corresponding recognition helices (Qi et al., 2014). The completed three-finger zinc-finger arrays were subsequently cloned into a Gateway® entry vector, and then recombined into an estradiol-inducible (XVE) destination T-DNA vector. The final T-DNA vectors contained a hygromycin gene (for plant selection), an XVE promoter system, and transcriptionally-linked zinc-finger nuclease pairs. To produce two proteins from a single transcript, we separated the two zinc-finger nucleases by a self-cleaving T2A sequence. These T-DNA sequences were integrated into the *Arabidopsis* genome using the floral dip method (Clough and Bent, 1998). In addition to WT plants, the T-DNA was also integrated into *ku70* and *lig4* backgrounds (Qi et al., 2013b).

We next assessed the activity of our zinc-finger nucleases in T1 seedlings. Briefly, seeds from plants that were dipped in *Agrobacterium* were plated onto MS agar with hygromycin and estradiol. Two weeks after plating, the healthy (and most likely transgenic) plantlets were pooled and total genomic DNA was extracted. We first predigested the genomic DNA with a restriction enzyme that recognizes a site near or within the zinc-finger nuclease spacer. We then used this predigested genomic DNA as a template in a PCR designed to amplify the appropriate zinc-finger nuclease target sequences. The resulting amplicons were digested with the restriction enzyme that cleaves near the spacer region, and then separated by gel electrophoresis (Figure A-1). If the zinc-finger nuclease was active, we would predict to see amplicons that were resistant to cleavage due to non-homologous end joining mutations that destroyed the restriction enzyme site. Three of our four zinc-finger nuclease pairs had detectable levels of cleavage-resistant amplicons. Cloning and sequencing of these bands revealed non-homologous end joining mutations at the zinc-finger nuclease target sites (Figure A-1). These results validated that three of our nucleases had activity at their predicted target locus, and furthermore, these data demonstrated that we generated active zinc-finger nucleases targeting each of our three genes of interest.

After determining that our nucleases are active, we next sought to isolate plants with heritable knockout mutations within our genes of interest. Several T1 plants harboring our three active zinc-finger nucleases pairs were germinated on media containing hygromycin and estradiol. Two weeks after plating, transgenic seedlings were transferred to soil and allowed to set seed. To find plants containing a heterozygous or homozygous mutation within the gene-of-interest, we collected T2 seed and planted this seed in soil. We then isolated genomic DNA from leaf tissue, amplified the zinc-finger nuclease target site, and digested the resulting amplicons with the restriction enzyme that cleaves within the zinc-finger nuclease spacer region. After

screening the first 32 seedlings from *lig4* plants containing the *TZP*-targeted zinc-finger nuclease pair, we identified five seedlings with mutations. Specifically, we isolated plant #18 which harbored a heterozygous 44 base pair deletion at the nuclease target sequence (the other *TZP* allele had a wild type sequence at the target site; Figure A-2C). This mutation was predicted to cause a frameshift mutation, resulting in an early stop codon that precedes three introns. Interestingly, the 44 base pair deletion was repaired using 5 bp of microhomology (TGAAG). This is consistent with repair by the alternative non-homologous end joining pathway. Notably, after screening seedlings from parents containing the TSPO651 and JMJC2086 zinc-finger nuclease pairs, we did not find any seedlings with mutations.

To further characterize the 44 base pair deletion at the TZP locus, we genotyped ten T3 seedlings using PCR (Figure A-3). After amplifying the zinc-finger nuclease target site, we were able to distinguish the wild type and mutant *TZP* alleles by gel electrophoresis. We determined that the mutation was faithfully passed down to the T3 seedlings. Of the ten plants we screened, one was wild type at both alleles, three were homozygous mutant, and six were heterozygous.

Taking the study as a whole, these results clearly demonstrate that stable integration of engineered zinc-finger nucleases in the *Arabidopsis* genome can result in targeted mutations at the corresponding target locus. Furthermore, our results demonstrate that DNA repair pathway choice can be manipulated by integrating genome engineering reagents into different mutant backgrounds. Here, this was demonstrated by recovering a mutant *tzp* plant that was generated by alternative non-homologous end joining repair of the double-strand break (as determined by the 5 bp of microhomology used to repair the break).

MATERIALS AND METHODS

Plasmid construction. Individual zinc-fingers and their recognition helices were assembled in pCP3 and pCP4 (Zhang et al., 2010). The completed zinc-finger arrays (containing three zinc-fingers) were then removed from these vectors by digesting with *Xba*I and *Bam*HI and cloned into a Gateway entry vector (AttL1 and AttL2) with heterodimeric *Fok*I coding sequences separated by a T2A sequence and cloning sites for two zinc-finger arrays (pZHY013) (Zhang et al., 2013). Gateway entry vectors containing zinc-finger nuclease pairs were recombined into pFZ19 (XVE-inducible T-DNA destination vector) (Zhang et al., 2010).

Floral dip transformation. *Arabidopsis* plants were transformed with T-DNA encoding zinc-finger nuclease pairs by floral dip transformation (Clough and Bent, 1998). Briefly, plants ~ 20 cm tall, were dipped in a solution containing *Agrobacterium,* 5% sucrose and 0.03% Silwet L-77*.* Plants were covered with a plastic dome for 24 hours.

Selecting T1 plants and inducing zinc-finger nuclease expression. Seeds from plants dipped in *Agrobacterium* were sterilized using a solution of 50% bleach (7 min with vortexing). The seeds were washed three times in sterile water, and plated on MS agar containing 20 μM 17β–estradiol and 20 μg/mL of hygromycin.

PCR to detect mutations. The zinc-finger nuclease target sequence within *TZP* (2033) was amplified by PCR using primers 5'-TGCCCGAATGCTCCCCTGTAT and 5'-TCCTTGAGCTTTGTCCCTCTGTGT. The restriction enzyme that was used to digest unmodified alleles was *Nla*III. The zinc-finger nuclease target sequence within *TSPO* (651) was amplified by PCR using primers 5'- TACGTATTTCCTCGGCACAAGCGA and 5'-ACGGCTTTACCAGATTACCAGCGA. The restriction enzyme that was used to digest unmodified alleles was *Hpy*188III. The zinc-finger nuclease target sequence within *JMJC* (2086) was amplified by PCR using primers 5'- GGAGCATCCTTAGGCATTCAAGGT and 5'-TGAACCACAGAAAGAGCACTCA. The restriction enzyme that was used to digest unmodified alleles was *Mnl*I. The zinc-finger nuclease target sequence within *JMJC* (2179) was amplified by PCR using primers 5'- GGAGCATCCTTAGGCATTCAAGGT and 5'-TGAACCACAGAAAGAGCACTCA. The restriction enzyme that was used to digest unmodified alleles was *Hpy*188I.

Proposed future studies:

1) Continue screening T2 seedlings from parent plants expressing the JMJC2086 and TSPO651 zinc-finger nucleases.
2) As an alternative approach to disrupt gene function design Cas9:gRNAs and TALENs targeting *TSPO* and *JMJC.*

Table A-1 Amino acid sequences of zinc-finger array recognition helices and their corresponding targeting sequences within *TZP, JMJC* and *TSPO* genes.

	DNA target sequence			Recognition helix amino acid sequence		
	F1	F2	F3	F3	F2	F1
TZP2033_Left_array	*GAA*	*GTA*	*GAG*	REDNLGR	QRSSLVR	HRPNLTR
TZP2033 Right array	*GCC*	*GCT*	*GAA*	QRNNLGR	QRSDLTR	DRRTLDR
JMJC2086_Left_array	*GTC*	*GGA*	*GAG*	VHWNLMR	QSAHLKR	TSTLLKR
JMJC2086_Right_array	*GAA*	*GTT*	*GAA*	QTNNLGR	HKSSLTR	QRSNLAR
JMJC2179_Left_array	*GCT*	*GTT*	*GCA*	QGNTLTR	HKSSLTR	TKQILGR
JMJC2179_Right_array	*GTT*	*GCT*	*GAA*	QRNNLGR	QRSDLTR	QATLLRR
TSPO651_Left_array	*GAA*	*GCG*	*GAC*	DPSNLRR	RTDTLAR	QQSNLSR
TSPO651_Right_array	*GCG*	*GTA*	*GGA*	QTTHLSR	QRSSLVR	KADTLVR

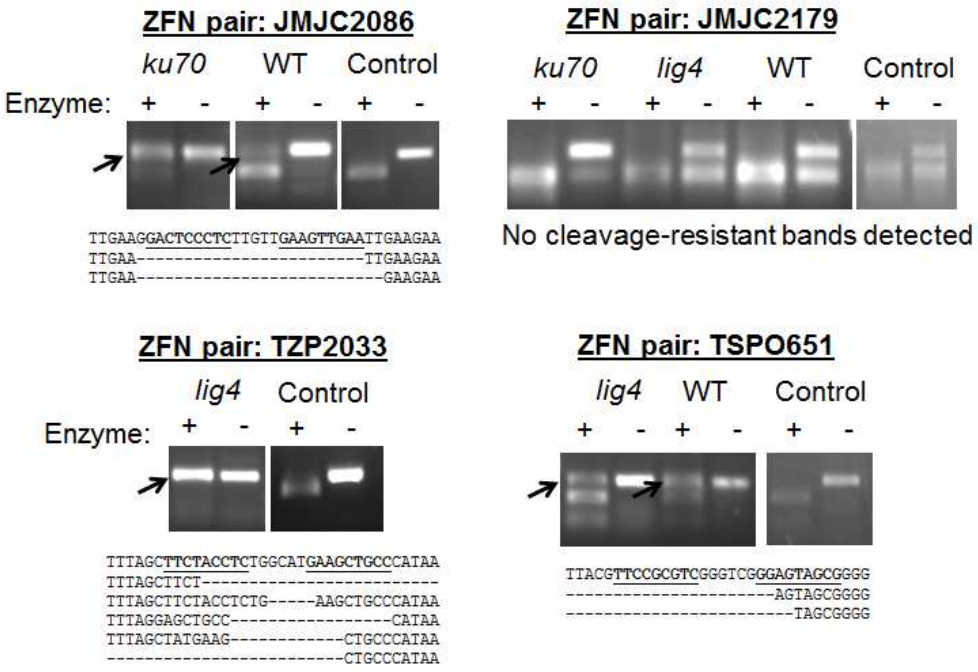

Figure A-1. Assessing zinc-finger nuclease activity in *Arabidopsis* T1 seedlings. T-DNA harboring zinc-finger nuclease pairs (Table A-1) was integrated into the *Arabidopsis* genome by floral dip. Transgenic T1 seedlings were grown on media and transcription of nuclease coding sequence was induced by estradiol. Gels represent data from a pool (~5) of seedlings that were grown on media containing hygromycin (to select for transgenic plants) and estradiol (to induce nuclease expression). Genomic DNA was extracted from the pool of seedlings and predigested with a restriction enzyme whose site is near the target site spacer. Following digestion, the nuclease target site was then amplified by PCR and the resulting amplicons were digested with the restriction enzyme. Cleavage-resistant amplicons were cloned and sequenced. Sequences containing mutations are shown below the corresponding gel.

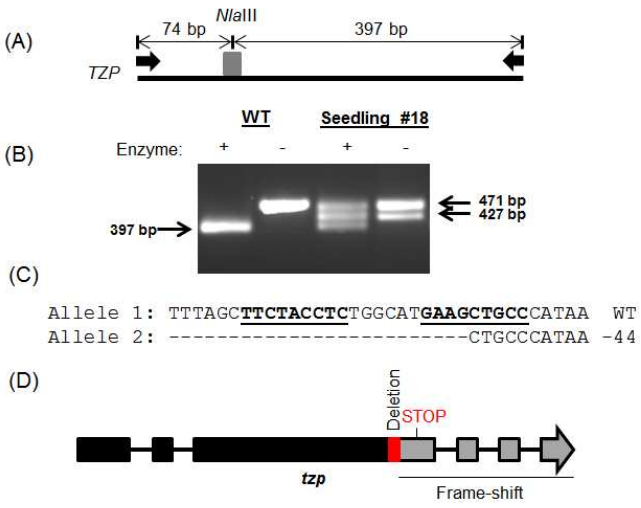

Figure A-2. Isolating an *Arabidopsis* seedling harboring a mutation within *TZP.* (A) Illustration of PCR amplicons containing the *TZP*-targeted zinc-finger nuclease target site. The black arrows represent PCR primers. The gray box represents the zinc-finger nuclease target site. *Nla*III is a restriction enzyme that recognizes a site (CATG) that is within the spacer of the nuclease target site. (B) Seeds from a parent plant containing the zinc-finger nuclease pair were planted in soil and screened for mutations at the *TZP* gene. After PCR amplification the nuclease target site, seedling #18 produced two products, one at the expected 471 bp, and the other at 427 bp. Digestion of these two products by *Nla*III resulted in three bands: one at the WT size (this was later found to be undigested product), one at 427 bp and the last at 397 bp. (C) PCR products were cloned and sequenced from seedling #18. One *TZP* allele contained a WT sequence, while the other contained a 44 bp deletion. (D) The 44 bp deletion is predicted to result in a frameshift mutation and a premature stop codon.

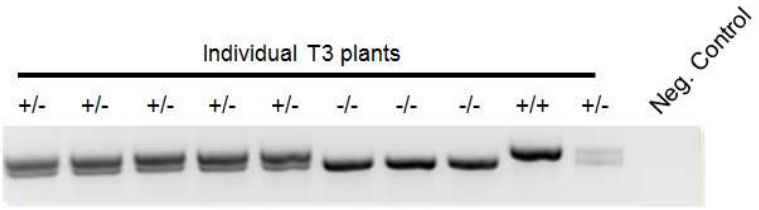

Figure A-3. Genotyping T3 seedlings. Seedlings generated from plant #18 (harboring a heterozygous mutation at *TZP*) were genotyped by PCR. Primers were designed to amplify the TZP2033 zinc-finger nuclease pair target site. Amplicons containing the 44 bp deletion are distinct from the WT counterpart after gel electrophoresis. +/-, heterozygous; +/+, wild type; -/- homozygous mutant.

APPENDIX B

DNA REPLICONS FOR PLANT GENOME ENGINEERING: DETERMINING GENE TARGETING FREQUENCIES IN TOBACCO

In chapter 2, we described the development, characterization and application of geminivirus replicons for editing DNA in *Nicotiana tabacum.* There were several key experiments that helped conclude that replicons enabled efficient homologous recombination-based genome editing. However, these experiments were all performed in leaf cells, which have exited the cell cycle, or they were performed in a non-quantitative manner (e.g., to determine if modified leaf cells maintain the ability to divide, we simply regenerated calli and plantlets). To further characterize the utility of geminivirus replicons for plant genome engineering, it would be beneficial to determine the efficiency that modified plants can be regenerated, and to compare this efficiency in different cell types, including callus, protoplasts, and germline cells.

Results

Here, we attempt to further characterize geminivirus replicons by calculating the gene targeting frequency (the total number of true gene targeting events divided by the total number of true events plus the number of illegitimate events) of geminivirus replicons and conventional T-DNA in tobacco calli. The gold standard for calculating the number of illegitimate events is to use a T-DNA vector with similar components (or the same components) as the experimental vector, and to determine the number of times this T-DNA integrates illegitimately. For these studies, we chose to use pCAMBIA2301 to determine the illegitimate recombination frequency; pCAMBIA2301 contains the same T-DNA backbone as our replicon vectors, and also contains 35S:*NPTII* and 35S:*GUS* between the T-DNA borders. We chose to use pCAMBIA2301 for two reasons: the phenotype resulting from illegitimate recombination of pCAMBIA2301 T-DNA will be phenotypically similar to that of the true recombination events using our replicon vectors–both will result cells staining blue due to a single (or few) GUS genes being expressed, and the T-DNA backbone is the same as our experimental vectors, therefore the T-DNA transfer and integration efficiencies should be comparable.

To determine the gene targeting frequency of our geminivirus replicons and conventional T-DNA, we regenerated modified callus from tobacco leaf tissue. Here, leaves from ~4-6 week old *Nicotiana tabacum* plants harboring the *gus:nptII* transgene were syringe infiltrated with (i) a mixture of *Agrobacterium* containing pLSLZ.D and pREP, (ii) *Agrobacterium* containing p35SZ.D, and (iii) *Agrobacterium* containing pCAMBIA2301. Five days after infiltration, the infiltrated leaves were removed from the plant and surface sterilized. The sterile leaves were cut into ~1 cm by 1 cm discs and transferred to regeneration media (4.4 g MS with vitamins, 1 mg/L 6-benzylaminopurine, 0.1 mg/L 1-naphthaleneacetic acid, 30 g sucrose, 8 g agar pH 5.7). Notably, no kanamycin was added to the media: we assumed that both unmodified and modified cells would divide at the similar rates, such that, after staining the tissue in a solution containing X-Gluc, the modified callus cells would be observable in a population of unmodified callus cells. After approximately 2 weeks on regeneration media, leaf discs were stained in X-Gluc and blue

calli were counted (Figure B-1, images on the left). To determine the frequency of recombination events (both illegitimate and true), we calculated the total number of blue spots and divided this number by the total circumference of the leaf discs (Figure B-2). We chose this method of calculation because callus tissue grows out from the outer edges of leaf discs, and not from the center of the discs. Using this approach, we calculated the average frequency of blue calli per centimeter of callus to be 0.136 for p35SZ.D (conventional T-DNA), 0.802 for pLSLZ.D and pREP (replicon), and 3.148 for pCAMBIA2301. We then calculated gene targeting frequencies by dividing the frequency of true gene targeting events (0.802 for our replicon sample and 0.136 for our conventional T-DNA sample) by the combined frequency of true gene targeting events and illegitimate events. The gene targeting frequency of our replicon was 2.0×10^{-1} (0.802/(0.802+3.148)) and conventional T-DNA was 0.4×10^{-1} (0.136/(0.136+3.148)). These results demonstrate that geminivirus replicons enhance the likelihood of capturing a modified callus by 5 fold compared to conventional T-DNA.

Discussion

In chapter 2, we found that delivering our genome engineering reagents to leaf cells using geminivirus replicons resulted in a 25-fold increase in the relative gene targeting frequencies as compared to conventional T-DNA (Figure 2-11C). Here, using the same vectors, we demonstrate a five-fold increase in gene targeting frequencies. Why is there a five-fold difference between these two results? One reason could be the difference in cell types between the two experiments. In chapter 2, we delivered our reagents to non-dividing, resting leaf cells and, here, whereas we initially delivered our vectors to leaf cells, the cells then began to divide and grow into callus. Therefore, the reason for the five-fold difference may be from the initiation of cell division by the growth hormones in the regeneration media. To further support this hypothesis, we observed in chapter 2 that expression of Rep/RepA results in a five-fold enhancement in gene targeting frequencies. We hypothesized that this enhancement is due to the pleiotropic activity of RepA. From previous work, RepA from mastreviruses was demonstrated to bind to the plant retinoblastoma related protein (RBR), thereby pushing the cell into S-phase. Furthermore, homologous recombination is more likely to repair DNA double-strand breaks as leaf cells move from a resting phase of the cell cycle to S-phase. Therefore, one explanation for the difference in these results is that cellular division caused by the hormones in the regeneration media is enhancing gene targeting frequencies of conventional T-DNA vectors to levels comparable to those seen in non-dividing cells after expressing Rep and RepA.

Future experiments

1) Explore replicon size and how it affects replication and genome stability by southern blot analysis.
2) Explore the frequencies of gene targeting when delivering additional sequence-specific nucleases.
3) Optimize the delivery of Cas9 and gRNAs within geminivirus replicons. Test different locations of Cas9 coding sequence, including within the replicon borders, between the T-DNA borders but not within the replicon, and stably integrated into the host's genome.
4) Explore the capacity for geminivirus replicons to carry out multiplex gene targeting where replicons carry two different donor molecules for two different targets.
5) Test bean yellow dwarf virus replicons in different species (e.g., *Arabidopsis,* potato, tomato); and test different replicon backbones (e.g., tomato yellow leaf curl virus, wheat dwarf virus, maize streak virus) in a wide range of plants.
6) Assess the genome of modified plants for off-target effects.

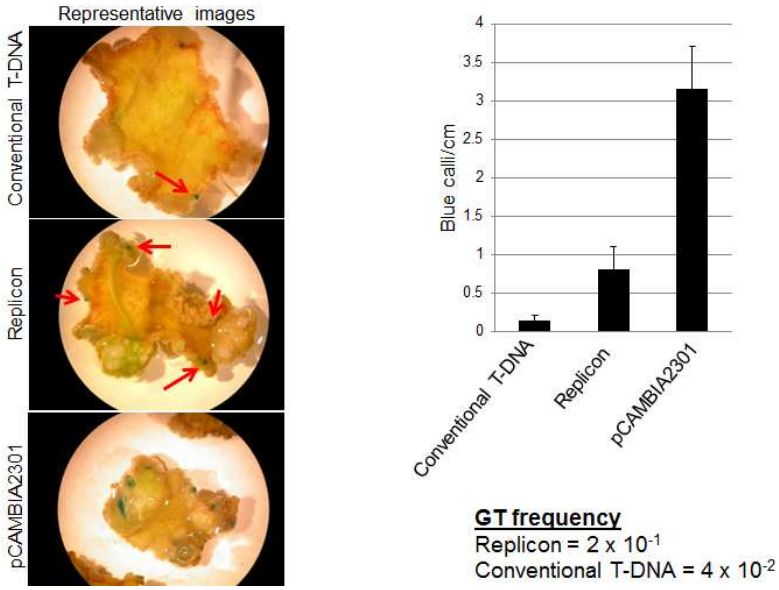

Figure B-1. Calculating the gene targeting frequency of geminivirus replicons and conventional T-DNA. Leaves growing on transgenic *Nicotiana tabacum* plants were syringe infiltrated with a mixture of *Agrobacterium* containing pLSLZ.D and pREP (replicon), *Agrobacterium* containing p35SZ.D (conventional T-DNA) and *Agrobacterium* containing pCAMBIA2301 (35S:NPTII and 35S:GUS between the T-DNA borders). Leaves were removed from the plant five days post infiltration and were cut into discs and placed on regeneration media. The images on the left are representative images of leaf discs stained in X-Gluc two weeks after plating. The arrows point to GUS-expressing calli. The graph on the right represents the frequency of GUS-expressing calli (calculated by dividing the total number of blue calli by the total circumference of the leaf discs). The gene targeting frequency (GT frequency) was calculated by dividing the frequency of GUS-expressing calli in the experimental samples by the total frequency of GUS-expressing calli from the experimental samples and pCAMBIA2301.

References

Clough, S.J. and Bent, A.F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. **16**: 735–743.

Krysan, P.J., Young, J.C., and Sussman, M.R. (1999). T-DNA as an insertional mutagen in Arabidopsis. Plant Cell **11**: 2283–90.

McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeted screening for induced mutations. Nat. Biotechnol. **18**: 455–457.

Potts, P.R., Porteus, M.H., and Yu, H. (2006). Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. EMBO J. **25**: 3377–3388.

Qi, Y., Starker, C.G., Zhang, F., Baltes, N.J., and Voytas, D.F. (2014). Tailor-made mutations in Arabidopsis using zinc-finger nucleases. Arab. Protoc. **1062**: 193–209.

Qi, Y., Zhang, Y., Zhang, F., Baller, J.A., Cleland, S.C., Ryu, Y., Starker, C.G., and Voytas, D.F. (2013). Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. Genome Res. **23**: 547–54.

Sander, J.D. et al. (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat. Methods **8**: 67–69.

Sander, J.D., Maeder, M.L., Reyon, D., Voytas, D.F., Joung, J.K., and Dobbs, D. (2010). ZiFiT (Zinc Finger Targeter): An updated zinc finger engineering tool. Nucleic Acids Res. **38**.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X., Pierick, C.J., Dobbs, D., Peterson, T., Joung, J.K., and Voytas, D.F. (2010). High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proc. Natl. Acad. Sci. U. S. A. **107**: 12028–33.

Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker, C.G., Bogdanove, A.J., and Voytas, D.F. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol. **161**: 20–7.