Breakthrough Technologies

Targeted Mutagenesis of Duplicated Genes in Soybean with Zinc-Finger Nucleases[1][W][OA]

Shaun J. Curtin, Feng Zhang, Jeffry D. Sander, William J. Haun, Colby Starker, Nicholas J. Baltes, Deepak Reyon, Elizabeth J. Dahlborg, Mathew J. Goodwin, Andrew P. Coffman, Drena Dobbs, J. Keith Joung, Daniel F. Voytas, and Robert M. Stupar*

Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, Minnesota 55108 (S.J.C., W.J.H., A.P.C., R.M.S.); Department of Genetics, Cell Biology, and Development (F.Z., C.S., N.J.B., D.F.V.) and Center for Genome Engineering (F.Z., C.S., N.J.B., D.F.V.), University of Minnesota, Minneapolis, Minnesota 55455; Molecular Pathology Unit and Center for Cancer Research (J.D.S., E.J.D., M.J.G., J.K.J.) and Center for Computational and Integrative Biology (J.D.S., E.J.D., M.J.G., J.K.J.), Massachusetts General Hospital, Charlestown, Massachusetts 02129; Department of Pathology (J.D.S., J.K.J.) and Biological and Biomedical Sciences Program (J.K.J.), Harvard Medical School, Boston, Massachusetts 02115; and Department of Genetics, Development, and Cell Biology, Bioinformatics and Computational Biology Program, Iowa State University, Ames, Iowa 50011 (D.R., D.D.)

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.

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 back-

ground. 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 *Fok*I 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; Wright et al., 2006; Maeder et al., 2009; Joung et al., 2010; Kim et al., 2010; Sander et al., 2011). 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 (Sander et al., 2011). 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 *Bcc*I and then PCR amplified at the *GFP* locus, such that only sequences with new mutations at the *Bcc*I recognition site would be amplified (Fig. 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 (Fig. 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 I; Supplemental Table S1). 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
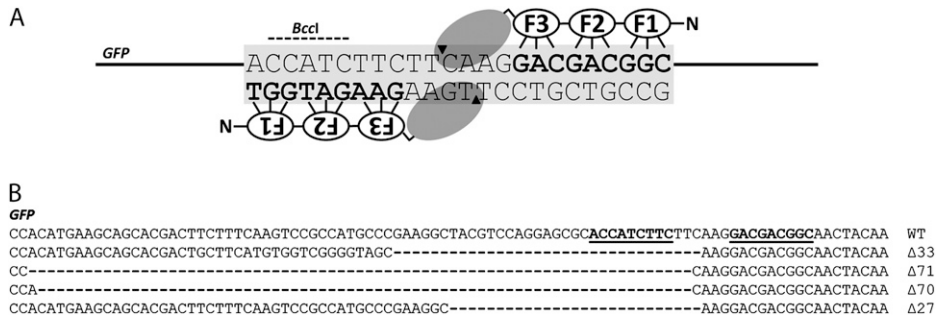
Figure 1. Detection of ZFN-induced mutations at a GFP transgene in soybean. A, The position of the OPEN ZFN target site is represented by a gray rectangle. The target sequence of both left and right ZFAs recognize a 9-bp sequence (indicated in bold). A strategy involving the restriction enzyme *Bcc*I 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 (Supplemental Table S3). 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 *Bcc*I 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.

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 I; Supplemental Table S1).

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 (Fig. 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 (Fig. 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 I; Fig. 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 (Fig. 3). We failed to recover mutations from the remaining three ZFN constructs (Supplemental Table S1).

To assess whether CoDA ZFNs can discriminate between closely related DNA sequences, we constructed two ZFNs that independently target the *RDR6* homeo-

Table I. *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[b]	RH (F1)	RH (F2)	RH (F3)
DCL1a[c]	Glyma03g42290	cAGCAACCTCTTATAAGAGGGCGTGg	6	TKQILGR	HKSSLTR	RHDQLTR
		gTCGTTGGAGAATATTCTCCCGCACc		SRFTLGR	LKEHLTR	RVDNLPR
DCL1b[c]	Glyma19g45060	cAGCAACCTCTTATAAGAGGGCGTGg	6	TKQILGR	HKSSLTR	RHDQLTR
		gTCGTTGGAGAATATTCTCCCGCACc		SRFTLGR	LKEHLTR	RVDNLPR
DCL4a[c]	Glyma17g11240	tTGCTTCATCACAATGGAGATGATt	5	RGQELRR	QQTNLTR	VGSNLTR
		aACGAAGTAGTGTTACCTCTACTAa		TKQRLVV	VRHNLTR	QTTHLSR
DCL4b[c]	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. [b]Measured in nucleotides. [c]Represents two duplicate copies ("a" and "b") that both perfectly match the ZFN target site.
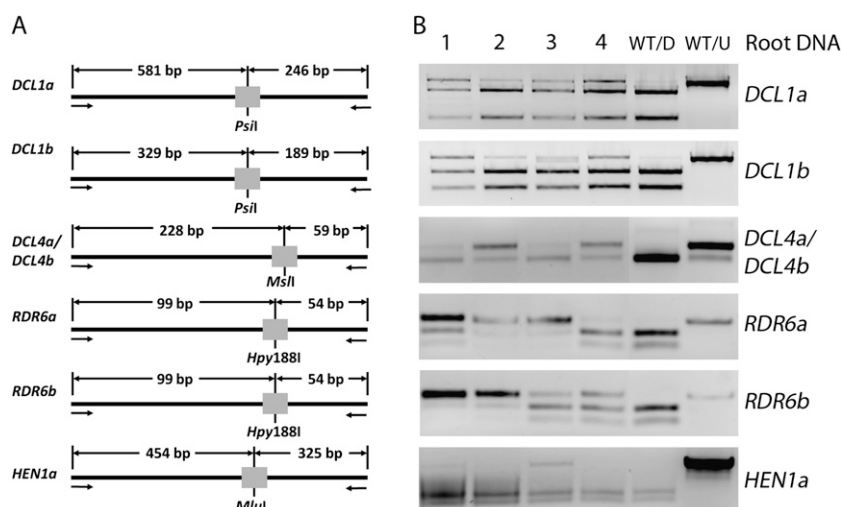
Figure 2. Detection of ZFN-induced mutations in soybean hairy-root tissue. A, A schematic strategy highlighting the restriction endonuclease PCR assays used to enrich mutated DNA sequences from soybean hairy-root tissue. Five ZFNs were designed to target seven genes, two of which targeted duplicate copies 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.

logs *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 I). 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 (Fig. 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 target-ing 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 (Supplemental Table S2).

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 (Supplemental Fig. S1A) and the other T0 seedling with a two-base thymine and adenine insertion at the *DCL4b* locus (Fig. 5A). 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

DCL1a

```
TGATTTACAAGCAGCAACCTCTTATA--AGAGGGCGTGGTGTATCATA    WT
TGATTTACAAGCAGCAACCTCTTATAtaAGAGGGCGTGGTGTATCATA    +2
TGATTTACAAGCAGCAACCTCTTATA---GAGGGTGTGGTGTATCATA    Δ1
TGATTTACAAGCAGCAACCTC--------AGGGCGTGGTGTATCATA     Δ9
TGATTTACA----------------------GGCGTGGTGTATCATA     Δ23
```

DCL1b

```
TGATTTACAGGCAGCAACCTCTTATAAGAGGGCGTGGTGTATCATACT    WT
TGATTTACAGGCAGCAACCTCTA-TAAGAGGGCGTGGTGTATCATACT    Δ1
TGATTTACAGGCAGCAACCTCTT------------------------    Δ34
TGATTTACAGGCAGCAAGAATCT------------------------    Δ25
T------------------------------------------------  Δ57
```

DCL4a

```
ATAACTTATCTCATTGCTTCATCACAA-TGGAGATGATTGCACTTTTG    WT
ATAACTTATCTCATTGCTTCATCA--A-TGGAGATGATTGCACTTTTG    Δ2
ATAACTTATCTCATTGCTTCATC-------GAGATGATTGCACTTTTG    Δ6
ATAACTTATCTCATTGCTTCATC---------GATGATTGCACTTTTG    Δ8
ATAACTTATCTCATTGCTTCATCACAAaTGGAGATGATTGCACTTTTG    +1
```

DCL4b

```
ATAACTTATCCCATTGCTTCATCACAATGGAGATGATTGCACTTTTGA   WT
ATAACTTATCCCATTGCTTCATC-----GGAGATGATTGCACTGTTGA   Δ5
ATAACTTATCCCATTGCTTCATCACAA-GGAGATGATTGCACTTTTGA   Δ1
ATAACTTATCCCATTGCTTCATCA----tGAGATGATTGCACTTTTGA   Δ4
```

RDR6a

```
TCGCACAAAGATAAGGCAACGACATCAGAGGAGTGGAAATTTTGACTC    WT
TCGCACAAAGATAAGGCAACGACA-----GGAGTGGAAATTTTGACTC    Δ5
TCGCACAAAGATAAGGCAACGACA---GAGGAGTGGAAATTTTGACTC    Δ3
TCGCACAAAGATAAGGCAACGACATtctAGGAGTGGAAATTTTGACTC    Δ3
```

RDR6b

```
TCGCACAAAGATAAGGCAACGACATCAGAGATGTGGAAATTTTGACTC    WT
TCGCACAAAGATAAGGCcacag-------------aaagcTTTGACTC    Δ13
TCGCACAAAGATAAGGCAACGA-------GATGTGGAAATTTTGACTC    Δ7
TCGCACAAAGATAAGGCAACGA---CAGAGATGTGGAAATTTTGACTC    Δ3
```

HEN1a

```
TAACAAGCAAGTTAGCACGCGACCACGCGTAGACGCATGATTCCGCAC    WT
TAACAAGCAAGTTAGCACGCGAC-----GTAGACGCATGATTCCGCAC    Δ5
TAACAAGCAAGTTAGCACGCGACCA-GCGTAGACGCATGATTCCGCAC    Δ1
TAACAAGCAAGTTAGCACGCGACCA-----AGACGCATGATTCCGCAC    Δ5
```

Figure 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.

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 Supplemental Figure S1.

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 (Fig. 5B). The *dcl4b* mutation segregated exactly 1:2:1 in the T1 prog-

eny, with six seedlings being homozygous for the mutation *dcl4b/dcl4b*, 12 seedlings heterozygous *DCL4b/dcl4b*, and six seedlings homozygous wild-type *DCL4b/DCL4b* (Fig. 5C). 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 (Supplemental Fig. S2) 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
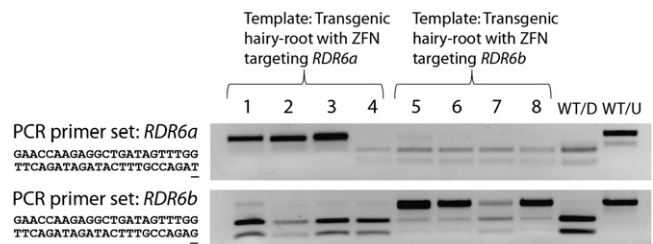

Figure 4. The mutagenic specificity of the RDR6a and RDR6b ZFN transgenes was assessed by performing PCR enrichment assays with gene-specific primers for each homeolog. The ZFN target sites of this gene pair differ by only two nucleotides, so this experiment was important to measure whether the gene-specific ZFNs could discriminate 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.
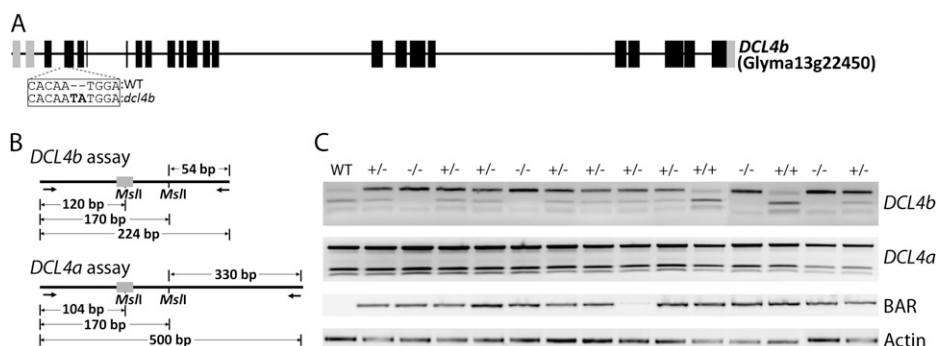
Figure 5. ZFN mutagenesis and heritability in whole-plant soybean. A, Genomic structure of the *DCL4b* gene in soybean. The target site is highlighted by dashed lines with the box indicating the ZFN-induced *dcl4b* mutation (2-bp insertion indicated in bold) relative to the wild type. B, A schematic 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.

DCL4b/dcl4b individual, lacked the ZFN construct (Fig. 5C).

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 at http://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 I). 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 I; Supplemental Table S1) 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 Supplemental Fig. S3 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 ribosome 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 *Xho*I and *Nhe*I 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 Screening for Mutations

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. For a more detailed hairy-root transformation protocol, see Supplemental Materials and Methods S1. 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 (Supplemental Table S3). 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.

A whole-plant transformation method for stable transformation of soybean was carried out using a protocol modified from the Plant Transformation

Facility at Iowa State University (Paz et al., 2006). The disarmed *A. rhizogenes* K599 variant 18r12, which lacks the root-inducing genes (Veena and Taylor, 2007), was used to transform the ZFN construct targeting the *DCL4* paralogs into whole soybean plants. Expression of the ZFN transgene was induced during the cocultivation, shoot induction, and shoot elongation steps of transformation by the application of estrogen to the tissue culture media. A detailed protocol of the whole-plant transformation and ZFN induction is presented in the Supplemental Materials and Methods S1 online. The screening and confirmation of whole-plant mutations followed the same protocol described above for the hairy-root samples. Transgene integration sites were identified by sequencing thermal asymmetric interlaced PCR (Singer and Burke, 2003) products.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Analysis of the ZFN-induced *dcl4a* mutation recovered from whole-plant soybean.

Supplemental Figure S2. ZFN-induced mutations recovered from *DCL4b* T0 and T1 whole-plant soybean.

Supplemental Figure S3. The CoDA method for engineering multifinger arrays.

Supplemental Table S1. The gene target accessions, target sequence, and RHs of CoDA ZFNs that did not generate mutations in the target genes.

Supplemental Table S2. Whole-plant transformation summary of the ZFN transgene targeting *DCL4*.

Supplemental Table S3. Primers used in this study.

Supplemental Materials and Methods S1. A detailed description of the hairy-root and whole-plant transformation methods.

LITERATURE CITED

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

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics **186:** 757–761

Cooper JL, Till BJ, Laport RG, Darlow MC, Kleffner JM, Jamai A, El-Mellouki T, Liu S, Ritchie R, Nielsen N, et al (2008) TILLING to detect induced mutations in soybean. BMC Plant Biol **8:** 9

Curtin SJ, Watson JM, Smith NA, Eamens AL, Blanchard CL, Waterhouse PM (2008) The roles of plant dsRNA-binding proteins in RNAi-like pathways. FEBS Lett **582:** 2753–2760

Fusaro AF, Matthew L, Smith NA, Curtin SJ, Dedic-Hagan J, Ellacott GA, Watson JM, Wang MB, Brosnan C, Carroll BJ, et al (2006) RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. EMBO Rep **7:** 1168–1175

Fu F, Sander JD, Maeder M, Thibodeau-Beganny S, Joung JK, Dobbs D, Miller L, Voytas DF (2009) Zinc Finger Database (ZiFDB): a repository

for information on C2H2 zinc fingers and engineered zinc-finger arrays. Nucleic Acids Res (Database issue) 37: D279–D283

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

Händel EM, Alwin S, 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

Hernandez-Garcia CM, Martinelli AP, Bouchard RA, Finer JJ (2009) A soybean (Glycine max) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. Plant Cell Rep 28: 837–849

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103: 16666–16671

Joung JK, Voytas DF, Cathomen T (2010) Reply to "Genome editing with modularly assembled zinc-finger nucleases". Nat Methods 7: 91–92

Kachroo A, Fu DQ, Havens W, Navarre D, Kachroo P, Ghabrial SA (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

Kim JS, Lee HJ, Carroll D (2010) Genome editing with modularly assembled zinc-finger nucleases. Nat Methods 7: 91

Li T, Huang S, Jiang WZ, Wright D, Spalding MH, Weeks DP, Yang B (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res 39: 359–372

Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, et al (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31: 294–301

Maeder ML, Thibodeau-Beganny S, Sander JD, Voytas DF, Joung JK (2009) Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. Nat Protoc 4: 1471–1501

Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM (2006) The evolution and diversification of Dicers in plants. FEBS Lett 580: 2442–2450

Mathieu M, Winters EK, Kong F, Wan J, Wang S, Eckert H, Luth D, Paz M, Donovan C, Zhang Z, et al (2009) Establishment of a soybean (Glycine max Merr. L) transposon-based mutagenesis repository. Planta 229: 279–289

Men AE, Laniya TS, Searle IR, Iturbe-Ormaetxe I, Gresshoff I, Jiang Q, Carroll BJ, Gresshoff PM (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

Meyer JD, Silva DC, Yang C, Pedley KF, Zhang C, van de Mortel M, Hill JH, Shoemaker RC, Abdelnoor RV, Whitham SA, et al (2009) Identification and analyses of candidate genes for rpp4-mediated resistance to Asian soybean rust in soybean. Plant Physiol 150: 295–307

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, et al (2011) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29: 143–148

Paz MM, Martinez JC, Kalvig AB, Fonger TM, Wang K (2006) Improved cotyledonary node method using an alternative explant derived from

mature seed for efficient Agrobacterium-mediated soybean transformation. Plant Cell Rep 25: 206–213

Pham AT, Lee JD, Shannon JG, Bilyeu KD (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

Reyon D, Kirkpatrick JR, Sander JD, Zhang F, Voytas DF, Joung JK, Dobbs D, Coffman CR (2011) ZFNGenome: a comprehensive resource for locating zinc finger nuclease target sites in model organisms. BMC Genomics 12: 83

Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y, et al (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8: 67–69

Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D (2007) Zinc finger targeter (ZiFiT): an engineered zinc finger/target site design tool. Nucleic Acids Res (Web Server issue) 35: W599-605

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178–183

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, et al (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature 459: 437–441

Singer T, Burke E (2003) High-throughput TAIL-PCR as a tool to identify DNA flanking insertions. Methods Mol Biol 236: 241–272

Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, et al (2002) The generic genome browser: a building block for a model organism system database. Genome Res 12: 1599–1610

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

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11: 636–646

Veena V, Taylor CG (2007) Agrobacterium rhizogenes: recent developments and promising applications. In Vitro Cell Dev Biol Plant 43: 383–403

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

Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF, et al (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nat Protoc 1: 1637–1652

Yang Z, Ebright YW, Yu B, 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 ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T, et al (2010) High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proc Natl Acad Sci USA 107: 12028–12033

Zuo J, Niu QW, Chua NH (2000) Technical advance: an estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. Plant J 24: 265–273