Cas9-Guide RNA Directed Genome Editing in Soybean

Zhongsen Li*, Zhan-Bin Liu, Aiqiu Xing, Bryan P. Moon, Jessica P. Koehlhofer, Lingxia Huang, R. Timothy Ward, Elizabeth Clifton, S. Carl Falco, and A. Mark Cigan
DuPont Pioneer Agricultural Biotechnology, Wilmington, Delaware 19803

ORCID IDs: 0000-0001-8697-3388 (Z.L.); 0000-0002-5983-5235 (A.M.C.).

Recently discovered bacteria and archaea adaptive immune system consisting of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) endonuclease has been explored in targeted genome editing in different species. Streptococcus pyogenes Cas9-guide RNA (gRNA) was successfully applied to generate targeted mutagenesis, gene integration, and gene editing in soybean (Glycine max). Two genomic sites, DD20 and DD43 on chromosome 4, were mutagenized with frequencies of 59% and 76%, respectively. Sequencing randomly selected transgenic events confirmed that the genome modifications were specific to the Cas9-gRNA cleavage sites and consisted of small deletions or insertions. Targeted gene integrations through homology-directed recombination were detected by border-specific polymerase chain reaction analysis for both sites at callus stage, and one DD43 homology-directed recombination event was transmitted to T1 generation. T1 progenies of the integration event segregated according to Mendelian laws and clean homozygous T1 plants with the donor gene precisely inserted at the DD43 target site were obtained. The Cas9-gRNA system was also successfully applied to make a directed P178S mutation of acetolactate synthase1 gene through in planta gene editing.

Plant transformation is most commonly achieved by Agrobacterium tumefaciens infection or particle bombardment, both of which have inherent challenges, such as random gene integration, endogenous gene interruption, multiple gene copies, and often unpredictable gene expression. Hundreds of events must be screened to identify a single copy-integrated gene that does not interrupt any endogenous gene. Site-specific integration (SSI) approach has been developed to place genes at previously screened genomic sites through recombinase-mediated cassette exchange (RMCE) using a recombinase, such as Causes Recombination (CRE) or flippase (FLP; Nanto et al., 2005; Chawla et al., 2006; Louwerse et al., 2007; Li et al., 2009). However, the SSI target sites are still generated by random insertions and must be maintained as unique lines to accept new genes by a second round of transformation.

DNA homology-directed recombination (HDR), commonly used to transform yeast (Saccharomyces cerevisiae) and some model animal species, is rarely successful in plant transformation. There are only a few reported attempts to change introduced genes or endogenous genes through HDR in model plants, such as Arabidopsis (Arabidopsis thaliana) and tobacco (Nicotiana tabacum; Halfter et al., 1992; Offringa et al., 1993; Miao and Lam, 1995; Kempin et al., 1997; Hanin et al., 2001). In one example using a positive-negative selection scheme to enrich HDR events, gene targeting was estimated at a frequency below 5.3 × 10⁻³ to endogenous targets of Lotus japonicus, although no HDR events were ever obtained (Thykjaer et al., 1997). Helped by an effective positive-negative selection and efficient A. tumefaciens transformation, Terada et al. (2002, 2007) successfully modified a rice (Oryza sativa) endogenous gene Waxy and later, a gene family member Alcohol dehydrogenase2 by HDR.

DNA double-strand breaks (DSBs) are naturally repaired by nonhomologous end joining (NHEJ), HDR, or microhomology-mediated end joining (Bleuyard et al., 2006). Homing endonucleases, such as I-SceI and I-CreI, have been used to generate artificial DSBs to stimulate HDR. HDR frequency at an artificial I-SceI recognition site previously placed in tobacco was increased by up to 100-fold when an I-SceI expression cassette was introduced together with a donor DNA by A. tumefaciens transformation (Pucht et al., 1996; Siebert and Puchta, 2002). Mutations of artificially introduced I-SceI recognition site in maize (Zea mays) were detected in 1% of analyzed F1 plants when I-SceI was introduced by crossing and activated by gene excision (Yang et al., 2009). Through the codelivery of a donor and an I-SceI expression DNA by either A. tumefaciens or biolistic transformation, the 35S promoter of the donor DNA was precisely inserted at previously introduced I-SceI sites at practical frequencies (D’Halluin et al., 2008).

* Address correspondence to zhongsenli97@gmail.com.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Zhongsen Li (zhongsenli97@gmail.com).

Z.L. designed and supervised the experiments, analyzed the data, and wrote the article; Z.-B.L. designed and supervised the gene editing experiments; A.X., B.P.M., J.P.K., L.H., R.T.W., and E.C. performed most of the experiments; S.C.F. conceived the project; A.M.C. supervised parts of the project and edited the article.

[OPEN] Articles can be viewed without a subscription. www.plantphysiol.org/cgi/doi/10.1104/pp.15.00783
Because homing endonuclease recognition sites do not normally exist in animal or plant genomes, unique agents are developed to specifically recognize a given genomic sequence. Taking advantage of the natural degeneracy of I-CreI recognition sequence, both rational design and experimental screening approaches have been used to create I-CreI derivatives that can recognize various DNA sequences (Seligman et al., 2002; Smith et al., 2006). An engineered I-CreI derivative capable of recognizing a sequence at the maize *liguleless* locus was successfully used to produce mutations with 2- to 220-bp deletions or short insertions at the expected cleavage site (Gao et al., 2010).

Zinc finger nucleases (ZFNs) are a group of engineered endonucleases that uses custom-designed zinc fingers to bind a specified DNA sequence, allowing the linked FokI endonuclease domain to generate a DSB in the recognized sequence (Durai et al., 2005). ZFNs work in pairs, because FokI nuclease subunits have to form dimers to cleave DNA. Mutations, small deletions and insertions, or targeted gene integrations at introduced ZFNs recognition sites were achieved in *Arabidopsis* and tobacco (Lloyd et al., 2005; Wright et al., 2005; de Pater et al., 2009). ZFN-mediated gene targeting was also successfully used to introduce a *Phosphinothricin acetyltransferase* herbicide resistance gene into a tobacco endochitinase gene and a maize inositol-1,3,4,5,6-petakisphosphate2-kinase gene or to confer resistance to sulfo-

**RESULTS**

**Cas9-gRNA-Directed NHEJ and HDR at Chosen Genomic Sites**

To test Cas9-gRNA-directed NHEJ and HDR, we designed Cas9-gRNA and donor DNA constructs to cotransform soybean embryonic callus by particle bombardment. The Cas9-gRNA DNA consists of two linked expression cassettes (Fig. 1A). One cassette contains a soybean *U6* small nuclear RNA gene polymerase III promoter expressing a gRNA that can recognize a 20-bp DNA target site (Mali et al., 2013). Cas9-gRNA DNA QC810 and QC799 each contain a gRNA recognizing the *DD20* and *DD43* targets, respectively (Fig. 2). The other cassette contains a codon-optimized *S. pyogenes* Cas9 gene expressed by a soybean *Elongation factor1 alpha2* (EF1A2) gene constitutive promoter (Li, 2014). The donor DNA construct contains a soybean S-adenosyl methionine synthetase (SAMS) gene promoter expressing the *Hygromycin phosphotransferase* (HPT) gene to confer hygromycin resistance. The donor cassette is flanked by approximately 1-kb *Homologous sequence1* (HS1) and *HS2* derived from the sequences flanking the genomic target site (Fig. 1, B and C). Donor DNA RTW830 and RTW831 each contain homologous sequences derived from the genomic DNA sequences flanking the *DD20* and *DD43* targets, respectively. Yeast FLP recombinase recognition sites *Flipase recognition target1* (FRT1) and *FRT87* are included for future gene targeting by RMCE (Li et al., 2009).

A distal region of chromosome 4 short arm, herein called the DD region, was chosen to test Cas9-gRNA-directed gene targeting. The DD region sequence retrieved from Phytozome (www.phytozome.net) was scanned manually for 23-bp sequences that meet the requirement of an *S. pyogenes* CRISPR target consisting of 20N plus an NGG PAM (Jinek et al., 2012; Cong et al., 2013; Mali et al., 2013). One target site (DD20) was selected from the sense DNA strand, and another site (DD43) was selected from the complementary strand slightly upstream of DD20 (Fig. 2). The 20-bp sequences of *DD20* and *DD43* targets were used to make DD20-CR1 and DD43-CR1 gRNA in Cas9-gRNA DNA constructs QC810 and QC799.
Characterization of Transgenic Events at Callus Stage

The DD20 and DD43 sites were targeted in two separate experiments by cotransforming soybean with target site-specific Cas9-gRNA and donor DNA fragments, respectively (Table I). Target site-specific quantitative PCR (qPCR) assays were designed to detect sequence changes at DD20 and DD43 targets around the expected Cas9-gRNA cleavage sites (Figs. 1B and 2). The assays consistently detected each target in diploid wild-type genomic DNA as one copy (homozygous [homol]). If one target allele was modified, the qPCR assay would detect only the unchanged target allele (i.e., approximately 0.5 copy as heterozygous [het]). If both alleles were modified, the qPCR assay would be negative (null) or detect approximately 0.1 or lower copy, because PCR primers can tolerate 1- to 2-bp mismatches in the DNA templates.

In total, 241 transgenic events from project DD20-CR1 were analyzed by the DD20-specific qPCR. No sequence change was detected in 87 or 36.1% of the events, and therefore, they still contained homo wild-type DD20 alleles. One-allele changes (NHEJ het) were detected in 67 or 27.8% of the events. Biallelic changes (NHEJ null) were detected in 76 or 31.5% of the events. The combined NHEJ mutation frequency was 59.3% for the DD20 target. In total, 263 events from DD43-CR1 project were similarly analyzed with the DD43-specific qPCR, and a combined NHEJ mutation frequency of 76.0% was detected. Similarly, high mutation frequencies and biallelic mutations have been reported in Arabidopsis and tobacco (Fauser et al., 2014; Gao et al., 2015).

Donor DNA integration was checked by an HPT-specific qPCR assay targeting the SAMS and HPT junction (Fig. 1, C and D). All events contained one or more copies of the donor DNA. Cointegration of the Cas9-gRNA DNA was detected by a Cas9-gRNA-specific qPCR assay (Fig. 1A) in 219 of the DD20-CR1 and 239 of the DD43-CR1 events (Table I).

The het and null NHEJ events from both DD20-CR1 and DD43-CR1 projects were analyzed by 5' and 3' border-specific PCR, with one primer specific to the donor DNA and the other primer specific to a genomic region outside the homologous sequence HS1 or HS2 (Fig. 1D). Events positive for both the 5' and 3' border-specific PCR were considered as putative HDR events (Supplemental Tables S1 and S2); 11 DD20-CR1 and 10 DD43-CR1 putative HDR events were identified (Table I).

Target Site Sequence Analysis of the Putative NHEJ and HDR Events

The target regions of randomly selected 30 DD20 and 30 DD43 NHEJ-null events were amplified by PCR with border-specific primers outside the homologous sequences HS1 and HS2 (Fig. 1B). The PCR fragments were cloned, and three clones (a–c) were sequenced for each PCR band. Examples of uniquely modified DD20 target sequences were aligned (Fig. 3). Deletions ranging from 1 bp in clones B8-1a and A12-13b to the longest 36 bp in clone B4-14c were all near the expected cleavage site (Fig. 3, arrow). Three clones also contained insertions

![Schematics of Cas9-gRNA-directed genome editing](image-url)
with the shortest 2 bp nucleotides TC in clone B7-1a and the longest 155 bp in clone B12-4b. The insertions also occurred around the expected cleavage site (Fig. 3). Examples of uniquely modified DD43 target sequences were similarly aligned (Supplemental Fig. S1). Deletions ranging from 1 bp to the longest 17 bp were all near the expected cleavage site (Supplemental Fig. S1, arrow). Eleven events contained insertions ranging from the shortest 8-bp TTAATTTA to the longest 220 bp. BLAST searches revealed that all of the inserts were derived soybean genome sequences. The border-specific PCR bands of selected DD20 and DD43 putative HDR events were also cloned and sequenced to confirm HDR recombination. Without exception, all of the sequenced 5' border clones contained a part of the 5' flanking sequence upstream of the HS1 region, the HS1 sequence, and a part of the transgenic donor DNA as expected from Figure 1D. Similarly, all of the sequenced 3' border clones contained a part of the transgenic donor DNA, the HS2 sequence, and a part of the 3' flanking sequence downstream of the HS2 region. The confirmed HDR events were kept for T0 plant regeneration.

Characterization of Putative HDR T0 Plants

Up to four T0 plants were regenerated from each putative HDR callus event and screened by the same target site, donor DNA, Cas9-gRNA DNA-specific qPCR, and 5' and 3' border-specific PCR assays used on callus samples (Fig. 1). Surprisingly, most of the T0 plants were no longer positive for the border-specific PCR, indicating that the putative HDR callus events failed to regenerate HDR T0 plants. Only one DD20 event (C5-5; Supplemental Table S1) and five DD43 events (D5-9, E2-4, F3-5, F6-12, and G1-2) regenerated T0 plants that produced both the 5' and 3' border-specific PCR bands (Supplemental Table S2). Many of the T0 plants regenerated from the same callus event contained different numbers of the target site, indicating that the callus event was chimeric. The putative HDR chimeric callus events might contain only a small portion of HDR cells that failed to regenerate HDR T0 plants.

Two DD20 and seven DD43 T0 plants were selected for further PCR analysis using better quality genome DNA (Fig. 4A). Both the DD20 plants produced the expected 5' and 3' HDR border-specific PCR bands, but the bands of plant A12-36 were too faint, suggesting that it was a chimeric event with a small percentage of cells converted to HDR. All seven DD43 plants produced the 5' HDR border-specific band, but only three plants (D5-9, E2-4, and G1-2) also produced the 3' HDR border-specific band. Plants D1-10 and F9-7 failed to produce a specific 3' border band, whereas plants F3-5 and F6-12 produced a larger 3' border band. The plants that contained the 5' border-specific band but not the correct 3' border-specific band were likely target site-specific insertion events containing the donor gene but not in a prefect HDR configuration (Fig. 1D).

The same T0 plant DNA samples were further analyzed by Southern hybridization with two digestions using NdeI and MfeI and an HPT gene probe specific to

![Figure 2](image-url)

**Figure 2.** Sequences of genomic target sites DD20 and DD43. Both DD20 and DD43 targets are at the distal end of chromosome 4 indicated by dashed vertical lines. Target site-specific qPCR amplifies the 64-bp DD20 sequence Gm04:48606498-49065061 or the 87-bp DD43 sequence Gm04:4860622-48860708 according to Phytozome 10.1 soybean genome sequence (phytozome.jgi.doe.gov). DD20-CR1 and DD43-CR1 gRNA targeting sequences and the 3-bp PAM are in bold type. Expected Cas9-gRNA cleavage sites are indicated by vertical arrows. qPCR primers and probes are underlined.

<table>
<thead>
<tr>
<th>Table 1. Cas9-gRNA-directed genome editing at two endogenous genomic sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project</strong></td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>DD20-CR1</td>
</tr>
<tr>
<td>DD43-CR1</td>
</tr>
</tbody>
</table>

Copyright © 2015 American Society of Plant Biologists. All rights reserved.
the donor and HDR DNA. HDR plants would produce specific 5' and 3' border bands as predicted from expected genomic DNA sequences (Supplemental Fig. S2). Partly consistent with the above PCR analysis, specific bands of expected sizes were hybridized only for the putative DD20 HDR event A12-36 and DD43 HDR events D5-9, F9-7, and G1-2 but not for E2-4. Additional HPT bands of nonspecific sizes were detected in most of the T0 plants, indicating that they all contained randomly integrated extra donors. Based on the PCR and Southern analyses, only one DD20 event (C5-5) and one DD43 event (D5-9) were selected for T1 segregation study.

Segregation of Homo T1 HDR Plants

To check if the extra genes contained in the DD20 event C5-5 and the DD43 event D5-9 would segregate away from the HDR insertion, 96 T1 seeds from each event were planted. The emerged 84 C5-5 and 90 D5-9 T1 plants were analyzed by the same target site, donor, and Cas9-gRNA DNA-specific qPCR assays (Fig. 1). The target site qPCR detected the target site segregation. The HPT qPCR detected both randomly integrated donor and the donor integrated at the target site of HDR plants. The Cas9 qPCR detected the segregation of randomly integrated Cas9-gRNA DNA.

The DD20 site segregated as a single locus with 23 homo, 48 het, and 13 null plants (Table II). Multiple copies of donor and Cas9-gRNA DNA were inserted at the DD20 site and segregated as 13 homo, 48 het, and 23 null plants as detected by the HPT- and Cas9-specific qPCR assays. The DD20 site and the transgenes were exclusive to each other, because they would occupy the same DD20 site (Supplemental Table S3). In conclusion, the DD20 event C5-5 contained multiple copies of both the donor and Cas9-gRNA DNA all at the DD20 site and segregated as one locus. No T1 plant with clean DD20 HDR insertion was obtained through segregation.

The DD43 site segregated as a single locus with 18 homo, 50 het, and 22 null plants (Table II). Multiple copies of Cas9-gRNA DNA segregated independent from the DD43 site as 26 homo, 49 het, and 15 null plants. The HPT qPCR assay detected both the HDR insertion at the DD43 site and randomly integrated extra donors that were likely linked with the Cas9-gRNA DNA, but the assay was not sensitive enough to distinguish the two loci (Supplemental Table S4). In conclusion, the DD43 event D5-9 contained a single copy of the donor DNA at the DD43 site and multiple copies of both the donor and Cas9-gRNA DNA at another locus independent from the DD43 site. Three DD43 homo T1 HDR plants (D5-9-12, D5-9-30, and D5-9-53) free of any extra donor or Cas9-gRNA DNA were obtained.

Confirmation of Precise Homology-Directed Gene Insertion by Cas9-gRNA

The integrated genes of homo DD20 T1 plants C5-5-102, C5-5-107, and C5-5-115 and DD43 plants D5-9-12, D5-9-30, and D5-9-53 were amplified by PCR using 5' and 3' border-specific primers (Fig. 1D). Surprisingly, an approximately 6-kb band was amplified from the three DD20 plants, which was not expected considering the multiple copies of donor and Cas9-gRNA DNA.
previously detected by qPCR. An approximately 2-kb band was amplified from DD43 plant D5-9-12 in addition to the expected 5,357-bp HDR fragment detected in all three DD43 plants (Fig. 4C). Some of the PCR fragments were cloned and sequenced.

The DD20 PCR fragment sequence matched most of the expected 5,364-bp DD20 HDR fragment, except for the 5′ border homologous sequence HS1 (Fig. 1D). Instead, the sequence upstream of the DD43 and D5-9-53 precisely matched the expected 5,357-bp DD43 HDR fragment. The approximately 2-kb PCR fragment of plant D5-9-12 turned out to be an NHEJ mutation with 5 bp of gtaca of the DD43 target CCGTAC(gtaca)AGTACAAGGGAC deleted (Fig. 2, sense strand). Because the same approximately 6-kb band was amplified from plant D5-9-12, although weakly because of the presence of the smaller approximately 2-kb band (Fig. 4C), D5-9-12 was likely a chimeraical HDR plant containing cells with the NHEJ DD43 allele. Only D5-9-30 and D5-9-53 were confirmed to be clean homo HDR T1 plants.

One het and three homo T1 plants of the DD20 (C5-5-126, C5-5-102, C5-5-107, and C5-5-115) and DD43 (D5-9-86, D5-9-12, D5-9-30, and D5-9-53) events were further evaluated by Southern hybridization to check the integrity of both of the 5′ and 3′ borders and the presence of extra donor and Cas9-gRNA DNA. The HPT probe would hybridize to an expected 5,136-bp 5′ border band with PciI digestion and a 5,278-bp 3′ border band with MfeI digestion for the DD20 plants and an expected 4,258-bp 5′ border band with NsiI digestion and a 4,375-bp 3′ border band with MfeI digestion for the DD43 plants (Fig. 5A). Each randomly inserted donor DNA would be hybridized by the HPT probe as an extra band of unknown size.

The HPT probe detected three bands, including the expected 5,136-bp 5′ border PciI band from the DD20 plants and an approximately 4.2-kb 3′ border band, which was smaller than the expected 5,278-bp MfeI band (Fig. 5B). The results disproved C5-5 as a DD20 HDR event. The additional HPT bands represented extra copies of the donor DNA detected previously by qPCR. Identical bands were detected in the het plant C5-5-126 and homo plants C5-5-102, C5-5-107, and C5-5-115, indicating that all of the genes integrated at one locus. The expected 4,258-bp 5′ border NsiI band and the 4,375-bp 3′ border MfeI band of DD43 HDR plants were detected in all of the DD43 plants. The extra HPT bands detected in the het plant D5-9-86 were absent from the homo plants D5-9-12, D5-9-30, and D5-9-53, indicating that the extra donor DNA had segregated away from the homo T1 plants (Fig. 5B).

The presence of the Cas9-gRNA DNA derived from constructs QC810 and QC799 was checked by a Cas9 probe specific to the 3′ one-half of the Cas9 gene. The Cas9 probe would hybridize a 2,044-bp or larger band for each intact copy of randomly integrated QC810 with Cas9-gRNA DNA. The expected bands were 1,202 bp for DD20 HPT Cas9 and 1,459 bp for DD43 HPT Cas9, indicating that all of the genes integrated at one locus. The expected 4,258-bp 5′ border NsiI band and the 4,375-bp 3′ border MfeI band of DD43 HDR plants were detected in all of the DD43 plants. The extra HPT bands detected in the het plant D5-9-86 were absent from the homo plants D5-9-12, D5-9-30, and D5-9-53, indicating that the extra donor DNA had segregated away from the homo T1 plants (Fig. 5B).

Table II. Segregation of the T1 plants of two putative HDR insertion events

Copies of the target sites and integrated genes in T1 plants were checked by sequence-specific qPCR assays using leaf genomic DNA samples. Sites or genes with both alleles detected positive by the qPCR were considered as homo. Sites or genes with one allele or both alleles detected negative were considered as het or null, respectively. Multiple copies of donor and Cas9-gRNA DNA were inserted at the DD20 site and segregated as a single locus. Thus, the transgenes and DD20 target were exclusive to each other (i.e. the same 23 DD20 homo plants were HPT and Cas9 null, whereas the same 13 DD20-null plants were HPT and Cas9 homo). One copy of donor DNA was inserted at the DD43 target site, and multiple copies of both donor and Cas9-gRNA DNA were also inserted at an unknown site. The two sites segregated independently, but HPT qPCR was not sensitive enough to distinguish them.

<table>
<thead>
<tr>
<th>qPCR</th>
<th>DD20 HDR Event C5-5</th>
<th>DD43 HDR Event D5-9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Homo</td>
<td>HPT</td>
</tr>
<tr>
<td>D20</td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td>D43</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>84</td>
<td>84</td>
</tr>
</tbody>
</table>
Targeted Mutagenesis through Cas9-gRNA-Enabled Gene Editing

ALS is a key metabolic enzyme in branched chain amino acid biosynthesis (Singh, 1999). Seed mutagenesis by ethyl methanesulfonate followed by chlorsulfuron resistance selection yielded a P178S (Pro at position 178 changed to Ser) semidominant mutation tolerant to sulfonylurea herbicides (Sebastian et al., 1989). Recently, another mutant from the same screening was shown to contain two mutations: P178S in ALS1 on chromosome 4 and W560L in ALS2 on chromosome 6 (Walter et al., 2014). Genomic sequence survey revealed two more soybean homologs: ALS3 on chromosome 13 and ALS4 on chromosome 15. Sequences around the P178 position of the four ALS genes are aligned with the nucleotides deviated from ALS1 in bold lowercase letters in Figure 6A.

A gRNA ALS1-CR1 targeting specifically the complementary strand of the underlined 20-bp ALS1 sequence in Figure 6A was linked to Cas9 gene in construct QC881. The ALS1-CR1 gRNA would not readily recognize any other ALS genes because of the single-nucleotide polymorphisms around the PAM site (Fig. 6A). A donor DNA fragment RTW1026A containing a 1,084-bp ALS1 sequence with five nucleotides AG-T-C-T changes specified by the bold uppercase letters in Figure 6C was used with QC881 to cotransform soybean with chlorsulfuron selection. Replacement of the endogenous ALS1 with the RTW1026A mutant fragment would change the position 178 Pro codon CCC to Ser codon AGC. Only those events containing the P178S mutation would survive chlorsulfuron selection. The other nucleotide T-C-T changes were all silent mutations to prevent the RTW1026A donor DNA from being recognized by the ALS1-CR1 gRNA for cleavage and create a KpnI site by the last T to facilitate analysis (Fig. 6C). The rest of the 1,084-bp RTW1026A fragment served as homologous sequences for HDR after the native ALS1 gene was cleaved by the ALS1-CR1 Cas9-gRNA complex (Fig. 6B).

One chlorsulfuron-resistant event, ALS1-18, was produced and analyzed by two PCR assays. A 1,246-bp region of the ALS1 gene locus was amplified by PCR using primer WOL573 complementary to the 5′ end of the 1,084-bp ALS1 fragment in RTW1026A and primer WOL578 complementary to a genomic region downstream of the 1,084-bp region (Fig. 6B). The 1,246-bp band derived from only the edited als1-18 sample was cut by KpnI into two bands of 695 and 551 bp (Fig. 6B, center) because of the introduction of a KpnI site in the P178S allele. A 730-bp band was amplified using a P178S mutation-specific primer WOL900 and primer WOL578 also from only the modified als1-18 sample (Fig. 6B, right).

The WOL573/WOL578 PCR band of the ALS1-18 event was cloned, and 24 colonies were sequenced. The sequences revealed that the ALS1-18 event contained two mutated als1 alleles. One allele represented by 18 colonies was a perfect P178S conversion containing the

**Figure 5.** Southern hybridization analysis of putative HDR plants. A, Predicted 5′ and 3′ bands (gray lines) that would be hybridized by HPT gene probe (black lines) for DD20 or DD43 HDR plants. B, Genomic DNA of one het (126) and three homo T1 plants of DD20 event C5-5 was digested with PciI or MfI to check the 5′ and 3′ borders using an HPT gene probe. The middle PciI band was consistent with the predicted 5,136 bp (rectangle), but the MfI band of approximately 4.2 kb was smaller than the expected 5,278 bp. One het (86) and three homo T1 plants of DD43 event D5-9 were similarly analyzed with NsiI or MfeI digests. Both the expected HDR-specific 4,258-bp band was consistent with the pre-
exact five nucleotides AG-T-C-T changes as designed. The other allele represented by six colonies lost five nucleotides immediately after the ALS1-CR1 gRNA cleavage site (Fig. 6C).

DISCUSSION

Recent developments in precise genome editing are based on the principle that DNA DSBs can stimulate DNA repair through NHEJ to introduce mutations and HDR to integrate foreign DNA. Although other technologies, such as customized homing endonuclease (Seligman et al., 2002; Smith et al., 2006; Gao et al., 2010), ZFN (Cai et al., 2009; Shukla et al., 2009; Townsend et al., 2009), and TALEN (Cermak et al., 2011; Christian et al., 2013; Zhang et al., 2013; Haun et al., 2014) all are able to cleave specific DNA sequences to induce NHEJ or HDR, Cas9-gRNA seems to be more flexible and effective. The unique target site recognition mechanism of Cas9-gRNA through DNA-RNA base pairing and protein-DNA binding is the simplest way to locate a genomic target (Jinek et al., 2012; Mali et al., 2013; Sternberg et al., 2014). After a Cas9-gRNA system is established, different genomic sites can be targeted by changing the 20-bp sequence in the gRNA gene. This simplicity makes it possible to target multiple targets to be mutated simultaneously by NHEJ, edited by HDR, and potentially regulated by promoter swap or regulatory elements modification.

Cas9-gRNA induced high NHEJ mutation frequencies at both DD20 and DD43 sites compared with customized homing endonucleases, ZFNs, TALENs, and even CRISPR (de Pater et al., 2009; Gao et al., 2010; Christian et al., 2013; Cheng et al., 2013; Jiang et al., 2013; Miao et al., 2013; Nekrasov et al., 2013; Liang et al., 2014; Schiml et al., 2014; Zhang et al., 2014; Jacobs et al., 2015). We evaluated 14 different gRNAs in...
a separate project, and DD20-CR1 and DD43-CR1 were representative of 11 gRNAs with similar or better mutation frequencies. Any unique genomic target site exists in a diploid genome as two alleles with one on each of the two homologous chromosomes. Cas9-gRNA-induced NHEJ was so effective that both alleles were often simultaneously mutated for both DD20 and DD43 targets (Table I). Two mutated target alleles often contained different sequences, such as those found in A12-5a and A12-5c, B3-5a and B3-5c, and B4-14a and B4-14c sequences (Fig. 3). Because none of the NHEJ events were kept for plant regeneration, it was not assessed how the chimeras would be carried to T0 plants or segregated among T1 plants. The putative HDR callus events were similarly chimerical, and many T0 plants regenerated from them lost the HDR insertion, whereas some others contained extra copies of randomly integrated donor and Cas9-gRNA DNA.

The approximately 4% putative HDR frequency at callus stage was artificially high, because all events positive for both the 5′ and 3′ border-specific PCR were counted in the calculation (Table I). Most of the weak positive events failed to regenerate HDR T0 plants, likely because of their chimerical cell composition. The chance of recovering an HDR plant could have been higher if more T0 plants per event were regenerated. Some of the regenerated T0 plants were positive only for the 5′ border-specific PCR, such as D1-10, F3-5, F6-12, and F9-7, of which F3-5 and F6-12 also produced a much larger 3′ border band (Fig. 4A). It was not clear how the 3′ borders of these imperfect but target site-specific events were recombined during the DSB repair.

Excess donor and Cas9-gRNA DNA were used in the first Cas9-gRNA-directed gene targeting attempt to help facilitate HDR. Consequently, most of the transgenic events, including the two selected events C5-5 and D5-9, contained multiple copies of the donor and Cas9-gRNA DNA. Following a similar approach to acquire clean RMCE T1 plants (Li et al., 2009), the extra gene insertions could be segregated away from the target site HDR insertion. Clean homo T1 plants were obtained from DD43 HDR event D5-9, whereas the DD20 HDR event C5-5 unexpectedly contained multiple copies of the donor and Cas9-gRNA DNA also inserted at the DD20 target site. The amounts of DNA used in future transformation need to be optimized to produce simpler HDR events with a reasonable frequency.

If appropriate recombinase recognition sites, such as yeast-originated FRT1 and FRT87, are placed in a genome, the genomic site can be used as a landing site to accept genes through more predictable RMCE (Nanto et al., 2009, 2010). A drawback of current RMCE technology is that the landing sites are created by random integration and often not at preferred loci. Cas9-gRNA-directed HDR can thus be used to place recombinase recognition sites at predesigned genomic sites to create landing sites for future RMCEs (Fig. 1D). Because FRT1 and FRT87 are included in the donor DNA, the DD43 HDR plants D5-9-30 and D5-9-53 can be used as an SSI target line to accept trait genes through RMCE. The advantages of Cas9-gRNA and RMCE technologies can be combined to achieve gene targeting at predesigned genomic sites repeatedly and more effectively.

Traditionally, genes are mutagenized randomly using chemical, physical, or biological agents, such as ethyl methanesulfonate, fast neutron, and transfer DNA, and extensive screening has to be carried out to select mutants with the desired mutations. DNA DSB-induced HDR using ZFNs or TALENs has been applied to selectively edit plant endogenous ALS genes (Townsend et al., 2009; Zhang et al., 2013). We applied the Cas9-gRNA system to specifically target only one of the four soybean ALS homologous genes and conveniently created a P178S ALS1 mutation exactly as designed. The same Cas9-gRNA-enabled gene editing approach can be applied to selectively edit any endogenous genes as intended.

It is now possible to selectively knockout any given endogenous gene using Cas9-gRNA-induced NHEJ. Several genes can also be simultaneously mutated through multiplexing. Because many genes have homologous sequences, one way of multiplexing is to design a single gRNA to target two or more homologous genes sharing a common target site sequence. Another way of multiplexing is using two or more gRNAs in a single transformation to simultaneously modify several unrelated genes (Li et al., 2013; Gao et al., 2015). The two ways of multiplexing were successfully combined in the simultaneous modification of five to nine genes in our hands (data not shown).

MATERIALS AND METHODS

DNA Construction

DNA constructs were made following standard molecular cloning procedures using components from existing DNA constructs (Li et al., 2009; Li, 2014). The Cas9-gRNA constructs QC810 containing U6-9-1-DD20-CR1=EF1A2:Cas9:PINII and QC799 containing U6-9-DD43-CR1=EF1A2:Cas9:PINII differ only in the 20-bp gRNA target sequences DD43-CR1 and DD20-CR2. Soybean (Glycine max) U6 small nuclear RNA promoters U6-9-I and U6-13-I, gRNA, and codon-optimized Streptococcus pyogenes Cas9 genes were all synthesized (GenScript). Unique restriction sites Ascl, XmaII, Ncol, and NcoI were designed in the constructs for subsequent cloning and DNA fragment preparation. Other Cas9-gRNA constructs, such as QC881 U6-9-1:ALS1-CR1=EF1A2:Cas9:PINII, were made by PCR using primers tagged with different 20-bp gRNA target sequences followed by appropriate restriction digestions and T4 ligation cloning. The donor constructs RTW830 containing DD20-HS1-SAMS-FRT1:HTP:NOS-FRT87-DD20-HS2 and RTW831 containing DD43-HS1-SAMS-FRT1:HTP:NOS-FRT87-DD43-HS2 differ only in the DD20 and DD43 homologous sequences HS1 and HS2. The DD20 and DD43 HS1 and HS2 were all synthesized (GenScript) with appropriate restriction sites for linking to the SAMS-FRT1:HTP:NOS-FRT87 cassette. Gene cassettes were released as DNA fragments with Ascl digestion, resolved by agarose gel electrophoresis, and purified using gel extraction kits (Qiagen). The 1,084-bp donor DNA fragment RTW1026A used for ALS1 gene editing was released with HindIII digestion from RTW1026 plasmid that contained a synthesized 1,953-bp fragment of soybean ALS1 gene containing the designed 5-bp modifications (Fig. 6C).

Plant Transformation

Soybean ‘93B86’ was transformed with QC810 and RTW830 DNA fragments for targeting DD20 site and QC799 and RTW831 fragments for targeting DD43 site following the particle bombardment transformation protocol using
9 pg bp⁻¹ donor DNA and 3 pg bp⁻¹ Cas9-gRNA DNA for each gold particle preparation that was then divided for 12 bombardments. Transgenic events were selected with 30 μg mL⁻¹ hygromycin (Li et al., 2009; Cigan et al., 2015). RTWI1026A and QC81 DNA fragments were cotransformed also at 9.3 pg bp⁻¹ concentrations for ALS1 gene editing using 100 ng mg⁻¹ chlorosulfuron for selection. Transgenic events were sampled at callus stage for PCR analysis to select events for T0 plants regeneration.

qPCR Analysis

qPCR analyses were performed as previously described (Li et al., 2009) using the soybean heat shock protein gene as the endogenous control. All qPCR probes were labeled with FAM-MGB (Applied Biosystems). Oligonucleotides sequences are listed in Supplemental Table S5. The copy numbers of donor DNA RTW180 and RTW181 were checked by qPCR using primers RTW180-F and RTW181-R and probe RTW43-T. The copy numbers of Cas9-gRNA DNA QC799 and QC810 were checked by qPCR using primers Cas9-F and Cas9-R and probe Cas9-T.

PCR and Sequence Analysis

PCR was performed as previously described (Li et al., 2009). The 5’ border of DD20 HDR events was checked with primers DD20-LB and Sams-A1 for a 1,204-bp band, whereas the 3’ border was checked with primers QC498A-S1 and DD20-RB for a 1,454-bp band. The 5’ border of DD43 HDR events was checked with primers DD43-LB and Sams-A1 for a 1,202-bp band, whereas the 3’ border was checked with primers QC498A-S1 and DD43-RB for a 1,454 bp-band. The DD20 target region was amplified by PCR with primers DD20-LB and DD20-RB as a 2,105-bp band, whereas the DD43 target region was amplified with primers DD43-LB and DD43-RB as a 2,108-bp-band. The ALS1 target region was amplified by PCR with primers WOL573 and WOL578 as a 1,246-bp band. The P178S mutant als1 target region was amplified by PCR with primers WOL90 and WOL578 as a 730-bp band.

PCR fragments were cloned in pcRB2-1:TOPO vector with TOPO TA Cloning Kits (Invitrogen). Plasmid DNA was prepared with plasmid DNA Mini-Prep Kits (Qiagen) and sequenced using a capillary DNA analyzer and dye terminator cycle sequencing kits (Applied Biosystems). Sequence assembly and alignment were done with Vector NTI Programs (Invitrogen). Sequence searches were done using BLAST algorithm against the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) database and the Phytozome soybean genome sequence (phytozome.jgi.doe.gov).

Southern Hybridization Analysis

Soybean genomic DNA was prepared from leaf samples and analyzed by Southern hybridization with digestoxigenin-labeled probes (Li et al., 2009). The DNA was digested with restriction enzymes MfeI, NsiI, or PciI and hybridized sequentially with a 794-bp HPT probe labeled by PCR with primers HPT1 and Hpygro2 and a 776-bp Cas9 probe with primers with primers Cas9-58 and Pn2 using PCR DIG Probe Synthesis Kits (Roche).

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Target site alterations of selected DD43 NHEJ events.

Supplemental Figure S2. Southern hybridization analysis of putative HDR T0 plants.

Supplemental Table S1. qPCR and border-specific PCR analyses of selected DD20 events.

Supplemental Table S2. qPCR and border-specific PCR analyses of selected DD43 events.

Supplemental Table S3. qPCR analyses of DD20 C5-5 T1 plants.

Supplemental Table S4. qPCR analyses of DD43 D5-9 T1 plants.

Supplemental Table S5. Oligonucleotides used in the study.

Acknowledgments

We thank DuPont Pioneer colleagues of the Soybean Transformation Group for producing transgenic events, the Controlled Environment Group for managing transgenic plants, and the Genomics Group for DNA sequencing.

LITERATURE CITED


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


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


Schimel S, Fauser F, Puchta H (2014) The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. Plant J 80: 1139–1150


