CRISPR/Cas9 and TALENs generate heritable mutations for genes involved in small RNA processing of *Glycine max* and *Medicago truncatula*

Shaun J. Curtin, Yer Xiong, Jean-Michel Michno, Benjamin W. Campbell, Adrian O. Stec, Tomas Čermák, Colby Starke, Daniel F. Voytas, Andrew L. Eamens and Robert M. Stupar

Received 15 May 2017; revised 17 October 2017; accepted 21 October 2017.

Abstract

Processing of double-stranded RNA precursors into small RNAs is an essential regulator of gene expression in plant development and stress response. Small RNA processing requires the combined activity of a functionally diverse group of molecular components. However, in most of the plant species, there are insufficient mutant resources to functionally characterize each encoding gene. Here, mutations in loci encoding machinery involved in small RNA processing in soybean and *Medicago truncatula* were generated using the CRISPR/Cas9 and TALEN mutagenesis platforms. An efficient CRISPR/Cas9 reagent was used to create a bi-allelic double mutant for the two soybean paralogous *Double-stranded RNA-binding* (*GmDrb*) and *Tobacco etiolated-rodlet* (*GmHen1*) genes. These mutations, along with a CRISPR/Cas9-generated mutation of the *M. truncatula Hua enhancer 1* (*MtHen1*) gene, were determined to be germ-line transmissible. Furthermore, TALENs were used to generate a mutation within the *soybean* *Dicer-like 2* gene. CRISPR/Cas9 mutagenesis of the soybean *Dicer-like 3* gene and the *GmHen1a* gene was observed in the T2 generation, but these mutations failed to transmit to the T1 generation. The irregular transmission of induced mutations and the corresponding transgenes was investigated by whole-genome sequencing to reveal a spectrum of non-germ-line-targeted mutations and multiple transgene insertion events. Finally, a suite of combinatorial mutant plants were generated by combining the previously reported *GmDc1a*, *GmDc1b* and *GmDc1b* mutants with the *GmDrb2a* double mutant. Altogether, this study demonstrates the synergistic use of different genome engineering platforms to generate a collection of useful mutant plant lines for future study of small RNA processing in legume crops.

Keywords: CRISPR/Cas9, mutagenesis, soybean, Medicago, small RNA, Drb2.

**Introduction**

Genome engineering technologies encompass the construction and introduction of site-specific nucleases (SSNs) into host genomes to create double-stranded breaks (DSBs) at targeted loci. The repair of the DSBs by the error-prone nonhomologous end-joining (NHEJ) pathway of the host organism often leads to nucleotide insertion and/or deletion (indel) events that can result in gene disruption. By taking advantage of this endogenous repair mechanism, plant biologists can generate modifications at targeted loci ranging in size from a single to several hundreds of base pairs in length, and in some cases much larger megabase pair deletions (Cermák et al., 2017; Qi et al., 2013; Voytas, 2013).

Site-specific nucleases are synthetic enzymes with programmable DNA target specificities and include the (i) zinc-finger nuclease (ZFN; Kim et al., 1996; Sander et al., 2011), (ii) transcription activator–like effector nuclease (TALEN; Christian et al., 2010; Miller et al., 2011) and (iii) clustered, regularly interspersed, short palindromic repeat (CRISPR)-associated protein (*Cas9*) (CRISPR/Cas9) system (Cong et al., 2013; Mali et al., 2013). ZFN and TALEN-mediated site-directed mutagenesis has been reported in several plant species including maize (*Zea mays*), tobacco (*Nicotiana tabacum*), barley (*Hordeum vulgare*), soybean (*Glycine max*) and Arabidopsis (*Arabidopsis thaliana*) (Cermák et al., 2011; Curtin et al., 2011; Li et al., 2012; Mahfouz et al., 2011; Wendt et al., 2013). CRISPR/Cas9 is the most recent platform and is comprised of the endonuclease (Cas9) and the spacer gRNA, which when combined, generate a site-specific DSB. Following transformation, the regenerated population of plants, the T0 generation, can be screened directly for lines harbouring mutations at the targeted locus and the heritable transmission of the introduced mutation(s) confirmed in the subsequent T1 generation of plants. Once a mutation has been heritably transmitted, it is considered stable and can be readily and reliably transmitted to subsequent generations.

Multiple reports have described the generation of somatic mutations using CRISPR/Cas9 in soybean (Cai et al., 2015; Du...
et al., 2016; Jacobs et al., 2015; Lowder et al., 2015; Michno et al., 2015; Tang et al., 2016). To date however, few studies have demonstrated whole-plant soya bean mutagenesis (Cai et al., 2017; Li et al., 2015). Here, we describe a set of plant expression vectors that can accommodate either the CRISPR/Cas9 or TALEN entry vectors to target the genes encoding the machinery proteins involved in small RNA processing in soya bean and M. truncatula (Baltes et al., 2014; Curtin et al., 2017; Zhang et al., 2013). Small RNAs are central regulators of gene expression associated with legume reproductive development, symbiotic nitrogen fixation, pathogen defence and the response to either biotic or abiotic stress (Ankit et al., 2014; Bustos-Sanmamed et al., 2013; Li et al., 2010; Yan et al., 2015; Zhai et al., 2011). We have previously demonstrated that the use of site-specific nucleases to generate mutations in the genes encoding the core machinery proteins of RNA silencing pathways is a highly successful tool for the study of small RNA biology in nonmodel plant species (Curtin et al., 2016). The newly generated mutant lines reported here greatly advance such resources and pave the way for future advances in legume small RNA biology research.

**Results**

**Assessment of CRISPR/Cas9 functionality in soya bean hairy-root tissue and TAL-effector nuclease activity in yeast**

Three binary plant expression vectors, each with a unique promoter, were converted into Gateway™ destination vectors. The promoters were selected for their constitutive and/or germ-line expression and included the Glycine max ubiquitin (Gmubi), Arabidopsis ubiqutin10 (UBQ10) and Agrobacterium rhizogenes Root-lociD (rolD) gene promoters (Figure 1a). To assess the functionality of each vector in soya bean, we used a previously reported set of CRISPR/Cas9 entry vectors that allow for the incorporation of the Cas9, along with a single (1x-plex), or double 2x-plex gRNA cassette, via a multiplex LR clonase reaction (Life Technologies, CA) (Baltes et al., 2014; Curtin et al., 2017; Figure 1b). We next constructed two single gRNAs entry cassettes that targeted the soya bean GmDcl3a gene (Glyma.04g057400; herein referred to as GmDcl3a) and carried out the cloning reaction to combine the Cas9 entry cassettes into each destination vector (Figure S1a). All six combinations of the CRISPR/Cas9 plant expression vectors were next transformed into A. rhizogenes strain K599 for the generation of transgenic plant tissue using a soya bean hairy-root ex vitro transformation assay (Figure S2a; Taylor et al., 2006). DNA extracted from 2-week-old hairy-root tissue was enriched for the mutated sequence by a restriction enzyme assay to remove nonmutated DNA sequences, and this preparation was subsequently used as PCR template (Figure S2b). A target amplicon clone and sequence (TACAS) assay confirmed deletions at both GmDcl3a loci (GmDcl3a-1 and GmDcl3a-2) from all three assessed expression vectors (Figure S2c).

To test whether the three assessed plant expression vectors could also be used in conjunction with a previously reported TALEN platform, we selected three soya bean genes to be targeted for TALEN-directed mutagenesis, namely GmDcl3a (Glyma.04g057400), GmDcl12b (Glyma.09g025300) and GmDrb2ab (Glyma.09g0257400). Furthermore, the TAL-effector Nuclease Targeter 2.0 (TALE-NT) software was used to identify suitable repeat variable domains (RVD) array configurations (Doyle et al., 2012; Zhang et al., 2013), and the configured arrays were rapidly assembled using the Golden Gate TAL-effector nuclease (TALEN) kit. The assembled arrays were next tested for their respective cleavage activities using a yeast-based assay (Figure 1c; Method S1) (Cermak et al., 2011; Christian et al., 2010), and TAL arrays determined to have a significant activity were incorporated into the plant expression vectors (Figure S3a–b).

**Successful mutagenesis and mutation transmission of the soya bean Drb2ab homoeologues**

A Gmubi promoter-driven CRISPR/Cas9 reagent with two gRNAs targeting sites within both soya bean Drb2a paralog copies was constructed (Drb2a, Glyma.12g075700, and Drb2b, Glyma.11g145900; referred to as GmDrb2a and GmDrb2b herein) (Figure 2a). The U6-driven gRNA targeted GmDrb2a and GmDrb2b (‘Target2’), while the 7sL-driven gRNA was designed to target four Drb genes, namely GmDrb2a and GmDrb2b and two additional Drb encoding loci (Glyma.12g172100 and Glyma.13g325600) (‘Target1’) (Figure 2b). As the 7sL promoter performs poorly in soya bean and M. truncatula, we did not expect to detect any activity associated with the 7sL-driven gRNAs. However, the resulting reagent transgene was still introduced into soya bean via Agrobacterium-mediated transformation and four T0 plants, WPT590-1, WPT590-2, WPT590-3, and WPT590-4, were recovered from this transformation experiment (Figure S4a–d). A PCR digestion assay was used to screen for introduced mutation and this approach showed that three of the four T0 plants returned digestion-resistant amplicons (plants, WPT590-1, WPT590-2 and WPT590-4) (Figure S4a). A heteroduplex and TACAS assay were applied to confirm the mutation status of these three T0 plant lines with the two T1 plants, WPT590-1 and WPT590-4, determined to harbour fully mutated alleles for both the GmDrb2a and GmDrb2b loci (Figures 2c and S4–S5), while mutated and wild-type amplicon sequences were isolated from T0 plant, WPT590-2 (Figures S4–S5). Sequence data generated from the TACAS assay also detected the presence of extra mutant amplicons sizes from the individual plants assessed. For example, three mutated GmDrb2a alleles (Δ8-bp, Δ7-bp and a Δ1-bp) were observed at high frequency across multiple leaf samples (Figures 2c and S5).

To confirm heritable transmission of the introduced mutations, T0 plants were self-fertilized and progeny screened (T1 plants) using the heteroduplex and TACAS approach. This approach revealed that the targeted mutations in the WPT590-1 and WPT590-4 T0 plants had successfully transmitted to the T1 generation (Figure S4b,d). However, this analysis also showed that no mutation or reagent transgene was transmitted to the T1 progeny of the WPT590-2 plant (Figures S4c and S5). Whole-genome sequencing (WGS) on the genomic DNA extracted from T0 plants, WPT590-1 and WPT590-4, and on two T1 progeny of each T0 plant (including T1 plants WPT590-1-1, WPT590-1-2, WPT590-4-1, and WPT590-4-2, respectively) was used to investigate the targeted loci in greater detail. The WGS approach confirmed the previously identified mutant alleles of a Δ7-bp and Δ1-bp for GmDrb2a and a Δ7-bp and Δ4-bp for GmDrb2b in both the T0 and T1 generations of the WPT590-1 and WPT590-4 backgrounds (Table S1; Figures S6 and S7). Interestingly, the four transmitted mutant alleles were identical between the T1 progeny derived from T0 plants WPT590-1 and WPT590-4. This finding indicates that the two T0 plants may have been independent regenerants of the same transformation and mutagenesis event. Furthermore, TACAS assay assessment of WPT590-4 T1 progeny
indicated that the four identified mutant alleles were segregating in the T1 generation (Figure 2d).

Upon successful heritable transmission of each mutant allele, additional somatic mutations in T1 plants are not expected to be detected due to the removal of the target sites by the biallelic mutation. This was determined to be the case in these analyses, as no new mutations were identified at the targeted sites of either the *GmDrb2a* or *GmDrb2b* locus. A single wild-type read was detected for the WPT590-1-1 sample; however, it is assumed likely that this read originated from contamination of the assessed T1 sample. For the T1 progeny of the T0 plant, WPT590-4, no somatic mutations nor the presence of any wild-type alleles was detected (Table S1). In conclusion, the WPT590-1 and WPT590-4 plants have fully mutated *GmDrb2a* (*Δ*7-bp and a *Δ*1-bp) and *GmDrb2b* (*Δ*7-bp and a *Δ*4-bp) alleles that result in the creation of in-frame premature stop codons (Figure S8).

Identification of null-segregant and phenotype analysis of *drb2adrb2b* mutant plants

Having established the double-mutation transmission in the T1 progeny of WPT590-1 and WPT590-4 T0 plants, PCR-based assays were next used to screen greater numbers (*n*/C25 = 50) of each T1 generation to identify null-segregant, transgene-free plants. However, this approach failed to identify any null-segregant plants (Figure S4b–d), suggesting that the WPT590-1 and WPT590-4 backgrounds harboured three or more transgenic events. We therefore used orphaned reads (reads where only one of the two paired ends map to the transgene cassette) generated as part of our WGS analysis to identify the genomic locations of transgene insertions in T0 plants. The WGS orphaned read approach failed to identify a transgene insertion event in the WPT590-2 background. However, this approach identified three independent transgene insertions on soya bean chromosomes.
four, thirteen and fifteen in the WPT590-1 and WPT590-4 plants (Figure 3a). Each integration site was found adjacent to small chromosomal DNA deletions (ranging from 11 to 31 bp; Figures 3a and S9–S11) and three of the flanking regions included small genomic DNA duplication or insertion events (Figure 3a). Moreover, the three identified transgene insertions were mapped to identical genomic coordinates in the WPT590-1 and WPT590-4 backgrounds to again indicate that these two T0 plants were independent regenerants of the same transformation and mutagenesis event.

The genomic coordinates of each inserted transgene were next used to screen for null-segregant T1 individuals and this approach revealed that T1 plant, WPT590-4-28, harboured only one of the three transgene insertions (Figure 3b). The applied screening also revealed that the transgene located on soya bean chromosome 4 was heterozygous (Figure 3b). Using the same approach to assess the T2 progeny of the T1 plant, WPT590-4-28, we identified a null-segregant (WPT590-4-28-5) that harboured no transgene-derived sequences (Figure 3b) and confirmed via sequencing the mutant status of this plant (Figure 3c). A preliminary phenotype analysis was conducted on twenty-four plants each of the WPT590-4-28 mutant (T1 progeny used in this analysis) and wild-type ‘Bert’ backgrounds to assess the sensitivity/tolerance of the drb2ab double mutant to drought stress. Drought stress was selected as an abiotic stress of interest due to the large number of proteins associated with ‘drought stress’ or ‘osmotic stress’ responses demonstrated to have altered abundance following proteomic profiling of the Arabidopsis drb2 (Atdrb2) mutant (Reis et al., 2015b). Reis et al. (2015b) demonstrated that Atdrb2 plants were putatively more resistant to salt stress (specifically, 100 and 150 mM NaCl) than either wild-type Arabidopsis (ecotype Columbia-0; Col-0), or the Arabidopsis drb1 mutant. However, we observed that 21-day-old Gmdrb2ab mutant plants were significantly (P = 0.01) more sensitive to drought stress than wild-type soya bean plants when water was withheld for a 7-day period (Figure S12a). Scanning electron microscopy was next used to analyse the adaxial and abaxial leaf surfaces of 14-day-old plants, but this analysis failed to detect any obvious structural defects (Figure S12c). More specifically, the stomata of Gmdr-b2ab and wild-type leaves were determined to be of a similar size.

Figure 2 CRISPR/Cas9 mutagenesis and inheritance of mutations for the soya bean paralogous gene pair GmDrb2a and GmDrb2b. (a) Schematic representation of the CRISPR/Cas9 reagent transgene. (b) The genomic arrangement of GmDrb2a and GmDrb2b. The ‘Target1’ and ‘Target2’ represent gRNAs target sites, and the green arrows represent primer sites used to screen mutations (Table S5). (c) The multiple mutated alleles were observed in the WPT590-1 and WPT590-4 T0 plants from the both the TACAS and WGS analyses at the Drb2a and Drb2b loci. The target sequences are indicated in boldface, along with the FAM site (red). The ‘Target 1’ gRNA also targets two additional (although uncharacterized) Drb genes; however, no activity was observed at these sites. (d) The heritable transmission of mutated alleles was confirmed in WPT590-1 and WPT590-4 T1 plants. The four T1 progeny of the TPT590-4 T0 plant shown here are examples of segregation of the Δ7-bp and Δ1-bp alleles for Gmdrb2a and the Δ7-bp and Δ4-bp alleles for Gmdrb2b.
and had a comparable aperture. We also quantified the abundance of several microRNA (miRNA) species in the Gmdrb2ab double mutant due to this mutant line displaying dark green leaf coloration (Figure S5), a hallmark vegetative phenotype expressed by the Atdrb2 mutant (Curtin et al., 2008; Eamens et al., 2012a).

A stem-loop primer-based quantitative reverse transcription PCR (SL-qPCR) approach was used to quantify the abundance of six miRNAs in the mutant and wild-type backgrounds (Turner et al., 2013). Encouragingly, minor perturbations to the abundance of three of the six quantified miRNAs, miR156, miR163 and miR168, were detected in Gmdrb2ab leaves (Figure S12b).

Successful mutagenesis of the GmdDrb locus followed by failure to transmit the mutation and transgene

A Gmubi promoter-driven CRISPR/Cas9 reagent with a single gRNA targeting the GmdDrb locus was transformed into soya bean, and this transformation experiment generated five T0 plants (Figure 4a–b). Two T0 plants, WPT527-1 and WPT527-2, exhibited
digestion-resistant amplicons typical of germ-line heterozygous mutations (Figure S13a). Sequencing of amplicons revealed Δ13-bp and Δ1-bp (~18%) and Δ4-bp and Δ1-bp (~12%) mutations to the GmdCl3a locus in T0 plants, WPT527-1 and WPT527-2, respectively (figure 4c). Initial analysis of target site amplicons recovered mostly wild-type alleles, rather than the expected 50% recovery of mutant amplicons normally associated with a heterozygous germ-line mutation (Figure S14). We next screened the progeny of the WPT527-1 and WPT527-2 T0 plants (approximately 60 T1 plants from each T0 plant) and failed to confirm heritable transmission of any mutant allele, or the presence of the reagent transgene in either T1 population (Figure S13b). Whole-genome sequencing was therefore next used to investigate each T0 event to determine both the fate of the mutant alleles, and to identify the genomic location of a putative reagent transgene. A comparable mutation frequency was obtained using WGS analysis to the TACAS assay for T0 plant, WPT527-1 (17.7%) (Table S2; Figures S14 and S15) and mutations in the WPT527-2 T0 plant were confirmed in both the GmdCl3a locus (9.1%) and the GmdCl2b pseudogene (17.8%) (Table S2; Figure S16). WGS orphaned reads identified a putative transgene location at Chr09:36 274 816–36 275 314 for WPT527-1; however, these same reads mapped to several genomic positions of low homology (Figure S17), and furthermore, a PCR-based assay failed to detect any transgene sequences at the expected chromosome 9 coordinates (Figure S18). Finally, transgene analysis of the WPT527-2 T0 plant line revealed low read counts and these lowly abundant reads failed to successfully map to any specific region of the soya bean genome.

**Targeted mutagenesis of the Hen1 locus in soya bean and Medicago truncatula**

A Gmubi and UBQ10 promoter-driven CRISPR/Cas9 reagent, each containing two gRNAs targeting both homoeologue copies of soya bean Hen1 locus (GmHen1a, Glyma.08g081600; GmHen1b, Glyma.05g126600) and two sites in the single copy MtHen1 gene (Medtr4g094545) were transformed into soya bean and M. truncatula (figure 5a–b). Two T0 soya bean plants, WPT589-1 and WPT589-2, were screened in tissue culture for the introduction of targeted mutations using a PCR digestion assay and the WPT589-2, were confirmed in both the GmHen1a and GmHen1b loci by performing both a TACAS and heteroduplex assay across both target sites using enriched and unenriched template; however, no evidence of mutagenesis was observed (Figure S19c–d). Twelve M. truncatula T0 plants were recovered from tissue culture and four T0 plants, namely WPT222-1, WPT222-6, WPT222-10 and WPT222-12, returned digestion-resistant amplicons indicative of germ-line heterozygous mutations. The progeny of these four plants were screened and segregating germ-line mutations were confirmed in the WPT222-1, WPT222-10 and WPT222-12 T1 plants (Figures S20a–d and S20). Multiple WPT222-1 and WPT222-12 T1 plants were found to harbour an identical Δ9-bp deletion, suggesting that these plants originated from the same transgenic event (Figure 5e). Screening of WPT222-10 T1 plants confirmed a +9-bp insertion (Figure 5e), and null-segregant plants were identified from each line at a frequency that suggested each was the result of a single transgene insertion events (Figure S20).

**Targeted mutagenesis of soya bean genes using TAL-effector nuclease (TALEN)**

In previous work, the soya bean dicer-like genes GmDcl2a, GmDcl2b and GmDcl3a proved recalcitrant to mutagenesis using the zinc-finger nuclease platform. Furthermore, and as outlined in Figure 4, CRISPR/Cas9 mutagenesis of the soya bean GmDcl3a gene. (a) A schematic representation of the single gRNA CRISPR/Cas9 reagent targeting GmDcl3a. (b) The genomic arrangement of the GmDcl3a gene from exon 15 to exon 18, including the target site at exon 17. Green arrows represent primers used to screen mutant plants. (c) Sequence data generated from TACAS assays of the WPT527-1 and WPT527-2 T0 plants.

© 2017 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 16, 1125-1137.
above, we also failed to generate a transmissible allele for GmDcl3a using the CRISPR/Cas9 platform. Therefore, we next attempted to use a TALEN platform to generate mutations in each of these soya bean Dcl genes. TAL arrays that showed a significant cleavage activity were cloned into a rolD expression vector (pSC218RG) and introduced into soya bean using a whole-plant Agrobacterium-mediated transformation assay (Table S3; Figure 6a). In total, 25 herbicide-resistant T0 plants were recovered from tissue culture and included (i) 20 plants from the GmDcl2a TALEN transformation (plants WPT383-1 to WPT383-20), (ii) 4 plants from the GmDcl2b TALEN transformation (plants WPT384-1 to WPT384-4) and (iii) 1 plant from the GmDcl3a TALEN transformation, the WPT423-1 T0 plant. Screening of GmDcl2a and GmDcl3a T0 plants failed to identify mutations at the targeted loci. However, three of the four T0 plants targeting Dcl2b (WPT384-1, WPT384-2 and WPT384-3) exhibited digestion-resistant amplicons indicative of a mutated target sequence (Figures 6b and S21a–b). Amplicons from the Dcl2b target site were sequenced and monoallelic deletions of Δ3-bp for plant WPT384-1, Δ5-bp for plant WPT384-2 and Δ6-bp for plant WPT384-3 were confirmed (Figure 6c). T0 plants were self-fertilized, and their T1 progeny screened to confirm heritable transmission of the mutagenesis transgenes. Sequencing of the target sites in T1 progeny of the WPT384-1 T0 plant confirmed segregation of the Δ3-bp deletion, with six heterozygous, four homozygous and two wild-type plants, including two transgene-free heterozygous lines, plants WPT384-1-8 and WPT384-1-12 (Figure S22). These results confirmed the previous findings of Haun et al. (2014) and Demorest et al. (2016) that TALENs can generate nontransgenic heritable mutant plants in soya bean. A screening of T1 progeny from the T0 plants, WPT384-2 and WPT384-3, failed to confirm heritable transmission of the T0

Figure 5  CRISPR/Cas9 mutagenesis of the soya bean and Medicago truncatula Hen1 genes. (a) A schematic representation of the reagent transgene targeting two sites in both homeologue copies of the soya bean Hen1 genes, GmHen1a and GmHen1b. The reagent transgene targeting sites in the single copy M. truncatula Hen1 gene, MtHen1, uses the AtUBQ10 promoter to drive Cas9 expression. (b) The genomic arrangement of the targeted soya bean GmHen1a and GmHen1b genes and the ‘Target1’ and ‘Target2’ sites. Green arrows represent primers used to screen mutant plants (c) The phenotypes of wild type (left) and the putative Gmhen1ab double-mutant seed (right). Seeds with the shrunk developmental phenotype germinated poorly and did not progress past the stage of radicle emergence, and also displayed a mottled phenotype (bottom panel). (d) The genomic arrangement in M. truncatula gene, MtHen1, and the location of the ‘Target1’ and ‘Target2’ sites. Green arrows represent primers used to screen mutant plants. (e) The sequence confirmation of germ-line mutations in MtHen1 WPT222-10 and WPT222-12 mutant plants. The in-frame mutations fail to disrupt the open-reading frame of MtHen1, and no obvious hen1-like mutant phenotype was observed.
mutations. Instead, screening identified mostly wild-type genotypes and some somatic mutations (Figure S23).

Generating combinatorial soya bean mutant plants by cross-fertilization

We have previously reported the generation of single nontransgenic Gmdcl1a, Gmdcl1b and Gmdrb2ab soya bean mutants via the use of the CoDA ZFN platform (Curtin et al., 2011, 2016). Mutant alleles from two nontransgenic plants, namely DCLA4a/DCLA4a/dcr2b/DCLA4a/dcr2b, and dcl1a/DCLA4a/dcr1b/dcl1a/DCLA4a/dcr1b, were combined by cross-fertilization to generate a combinatorial Gmdcl1a/dcr1b double mutant (Figure S24a; Method S2). The resulting seeds were bulked, and the mutant status of F1 plants was confirmed by a PCR digestion assay that confirmed wild-type, dcl1a and dcr2b allelic backgrounds (Figure S24b). The F2 progeny of these F1 plants were next screened and the homozygous double mutant was identified (Figures 7a and S24–S25). A preliminary phenotypic analysis comparing seed weight of the mutant and wild-type plants was determined to be significantly reduced in the double mutant (P < 0.00001). These results demonstrate that the stacking of single mutants to generate a combinatorial dcl1a/dcr1b double mutant is a successful approach for the generation of novel mutant backgrounds that display unique phenotypes. We next combined the Gmdcl4b and an additional Gmdcl1a, and Gmdcl1b mutant alleles with the GmDrb2a/trb2b, GmDrb2a/trb2b, and GmDrb2a/trb2b, double mutant, described above (Figures S26–S30). The F1 plants resulting from this crossing event had their mutational status confirmed by a sequence-based approach and the F2 progeny bulked for the identification of single, double and triple mutant combinations of Gmdcl1, Gmdcl1b, Gmdcl4b, GmDrb2a and GmDrb2b plants (Table 1).

Discussion

Here, we report the generation of targeted mutations in loci central to the small RNA pathways of soya bean and M. truncatula using the ZFN, TALEN and CRISPR/Cas9 nucleases platforms. This approach enabled the recovery of a wide range of stable, nontransgenic plants, including those with in-frame, frame-shift, single homozygous and double homozygous mutations. The CRISPR/Cas9 constructs proved straightforward to produce and were efficient at generating targeted mutations in planta in both species. One of the key findings of this study was the ability of CRISPR/Cas9 reagents to generate double mutants with fully mutated alleles in the initial T0 generation, as evidenced by the GmDrb2a and GmDrb2b paralog pair. Soya bean is a highly duplicated palaeo-polyploid plant where approximately 70% of the protein-coding sequences encoded by the genome are further represented by a duplicated copy (Schmutz et al., 2010). An important requirement for any reagent designed for use in soya bean is the ability to simultaneously mutate multiple gene copies in a single transformation event. While the capacity of the CRISPR/Cas9 reagent to readily generate mutants is quite remarkable, we have also identified double mutants in a single generation using the ZFN platform (Curtin et al., 2016). However, ZFNs have a reduced workable target range in most plant species and require considerably more effort to design and construct (Sander et al., 2011).

We also report here on the generation of two T0 double bi-allelic mutations of the GmDrb2a and GmDrb2b loci, and further, heritable transmission of these mutations was demonstrated in T1 progeny. The mutated alleles in these two plants were identical (this was true for both gene targets, GmDrb2a and GmDrb2b), and at least one CRISPR/Cas9 transgene insertion was transmitted to all T1 progeny. Our initial suspicion that these plants arose from the same transgenic explants was confirmed by WGS analysis that identified three transgene insertions in exactly the same genomic coordinates in both T0 plants. Using these genomic coordinates, null-segregant transgene-free plants were identified and deregulated article status granted by USDA-APHIS, thereby paving the way for future field testing of this line (https://www.aphis.usda.gov/aphis/outfocus/biotechnology/am-i-regulated) (Figure S31). The drb2ab double-mutant plant will also be a valuable genetic resource for studying the function of DRB2 function in soya bean, having previously been demonstrated to play functionally diverse roles in multiple small RNA-directed RNA silencing pathways in Arabidopsis, including the production stages of the repeat-associated small-interfering RNA (rasiRNAs) and trans-acting small-interfering RNAs (tasRNAs) pathways (Pelissier et al., 2011) and the inhibition of viral replication (Barton et al., 2017). Arabidopsis DRB2 has also been shown to be an important requirement for the production of a cohort of miRNAs in developmentally important Arabidopsis tissues, namely the shoot apex and maturing pollen (Eamens et al., 2012a,b). Many of the target genes of this ATDR2-dependent miRNA cohort appear to be involved in stress responses and are not involved in normal plant
development pathways (Reis et al., 2015b). Therefore, in view of these findings, the generation of the Gmdrb2ab double mutant provides an excellent opportunity to characterize miRNA-directed responses to stress in soya bean and for the subsequent comparison of GmDRB2 function to the well-documented functionality of the AtDRB2 protein.

The transgenesis and mutagenesis of the GmDcl3a locus with our CRISPR/Cas9 reagent generated curious results. First, we identified two putative T0 mutant plants that initially appeared to be germ-line heterozygous mutants (Figure S13a). However, the quantity of mutated amplicons obtained was significantly less than that expected for a heterozygous mutant plant, which normally returns an approximate mutant to wild-type amplicon ratio of 1 : 1 (i.e. 50% mutant and 50% wild-type amplicons). Secondly, and most importantly, neither the targeted mutation nor the introduced transgene was heritably transmitted to the T1 generation (Figure S13b). We used the generated WGS data, specifically orphaned reads, to identify the genomic location of the transgene insertion on chromosome nine for the WPT527-1 T0 plant. However, we failed to identify a transgene insertion for the WPT527-2 T0 plant, despite orphaned reads mapping to the transgene at a very low frequency. There are at least three

Figure 7 Combining soya bean dicer-like mutated alleles. (a) Sequence confirmation of a double mutant developed via a standard genetic crossing approach using the confirmed soya bean dcl1a mutant line and the dcl4b mutant plant. (b) In a preliminary phenotype analysis, the dcl1a \textsuperscript{a6}/dcl1a \textsuperscript{a6}/dcl4b \textsuperscript{a2}/dcl4b \textsuperscript{a2} double-mutant seed has a significantly reduced seed weight compared with wild type.

Table 1 The generation of combinatorial dcl1ab/drb2ab and dcl4b/drb2ab mutant plants

| Reference | Direction of cross | Mutant alleles | Mutants |
|-----------|------------------|----------------|--------|
| M13-719   | dcl1a × dcl4b    | dcl1a \textsuperscript{a6}, dcl4b \textsuperscript{a2} | dcl1a/dcl4b-1\textsuperscript{†} |
| M17-725   | db2a × wt        | db2a \textsuperscript{a1}, db2b \textsuperscript{a2} | db2a-1, db2b-1\textsuperscript{†} |
| M17-726   | wt × db2ab       | N.D.           |        |
| M17-727   | db2a × dcl1b     | db2a \textsuperscript{a6}, db2b \textsuperscript{a1}, dcl1b \textsuperscript{a15} | db2a-1, db2b-1\textsuperscript{†}, db2a/dcl1b-1\textsuperscript{†}, db2a/db2b/dcl1b-1\textsuperscript{†} |
| M17-728   | dcl1b × db2ab    | N.D.           |        |
| M17-729   | db2ab × dcl4b    | dcl1a \textsuperscript{a7}, dcl4b \textsuperscript{a2} | dcl1a-2, db2ab-2\textsuperscript{†}, db2a/dcl4b-1\textsuperscript{†}, db2a/db2b/dcl4b-1\textsuperscript{†} |
| M17-730   | dcl4b × db2ab    | N.D.           |        |
| M17-733   | dcl1a × db2ab\textsuperscript{a*} | db2a \textsuperscript{a7}, db2b \textsuperscript{a8}, dcl1a \textsuperscript{a7} | db2a-2, db2b-1, db2a/dcl1a-1, db2a/db2b/dcl1a-1 |
| M17-734   | dcl1a × db2ab\textsuperscript{a*} | N.D.           |        |
| M17-735   | dcl4b × dcl1a\textsuperscript{a*} | dcl1a \textsuperscript{a7}, dcl4b \textsuperscript{a2} | dcl1a/dcl4b-2 |
| M17-736   | dcl1a × dcl4b\textsuperscript{a*} | N.D.           |        |

N.D. not determined.
*M1 screening in progress.
*M2 screening in progress.

ª2017 The Authors. Plant Biotechnology Journal published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 16, 1125–1137
been used in cl1a and GmDcl1a Cas9 constructs used in this study showed mutation activity in transmittable mutation. In comparison, each of the four CRISPR/Cas9 reagents exhibited activity in yeast and just one producing a and types associated with mutation of the M. truncatula reagent platform to introduce mutations to other targeted and rice (Abe et al., 2010; Chen et al., 2002). Using an identical reagent platform to introduce mutations to other targeted M. truncatula genes, we have previously observed high mutation rates of 50%–75% in the T0 generation plants (Curtin et al., 2017). It was therefore surprising that only in-frame mutations were recovered at reduced mutation efficiencies. Taken together, these results suggest that a heterozygous frame-shift knockout of Hen1 may be embryo lethal in legumes.

The TALENs and CRISPR/Cas9 platforms are complementary genome engineering technologies, with the CRISPR/Cas9 platform recently emerging as the platform of choice for most applications. The TALEN platform has a limitless targeting range that might be advantageous for targeting a highly specific locus (Joung and Sander, 2013). However, the striking disadvantage of the TALEN approach, in our limited experience in soya bean, is the inferior mutational efficiency offered by this platform. Here, nine TALEN constructs that targeted three Dcl loci (GmDcl2a, GmDcl2b and GmDcl3a) were engineered. However, three of these reagents exhibited activity in yeast and just one producing a transmittable mutation. In comparison, each of the four CRISPR/Cas9 constructs used in this study showed mutation activity in whole plants and two of them produced transmittable mutations.

A combinatorial Gmdc11a/dcl4b mutant plant was identified by combining single Gmdc11a and Gmdc4b mutants previously generated via the use of the ZFN platform. A similar approach has been used in Arabidopsis to establish both hierarchical and redundant roles for each of the five members of the D8R gene family in the parallel RNA silencing pathways (Curtin et al., 2008; Eamens et al., 2009, 2012a,b). We also confirmed the functional activity of Gmdc22b and Gmdc33a TALEN and CRISPR/Cas9 reagents that could be introduced into the nontransgenic Gmdc11a/dcl4b double mutant to quickly combine additional mutations. Similarly, the Gmdrb2ab double mutant was crossed with Gmdc11a, Gmdc11b and Gmdc4b single mutants, and single, double and triple mutant combinations are currently being screened. The generation of this particular suite of soya bean mutant germplasm is of high priority when Arabidopsis research is considered; that is, AtDRB2 appears to be able to bind both imperfectly and perfectly dsRNA templates, and to form functional partnerships with AtDCL1 and AtDCL4 for miRNA and siRNA (rasiRNAs and tasRNAs) production, respectively (Eamens et al., 2012a; Pelissier et al., 2011; Reis et al., 2015a). Therefore, this collection of small RNA-associated mutants forms a valuable genetic resource, not only for the legume research community but also for the greater plant biology community. Several recent reports have used whole-genome, RNA and small RNA sequence analyses to document the roles played by small RNA-directed gene expression regulation in both development and response to various environmental stresses (Arik et al., 2014; Curtin et al., 2012; Zhai et al., 2011). To date however, small RNA studies in plant species outside of Arabidopsis have been largely limited to cataloging alterations in the accumulation of various species of small RNAs in developmentally distinct tissues, or after exposure to stress. Therefore, in legumes, as in other plant species, there has been a conspicuous lack of mutant analyses to establish gene function for loci encoding the orthologous central machinery proteins involved in either the production or activity stages of small RNA-directed RNA silencing pathways.

Genome engineering reagents are rapidly becoming central tools for legume functional genomic studies as well as being used as an alternate approach to improve the soya bean germplasm (Cai et al., 2017; Curtin et al., 2017; Demorest et al., 2016). The limitations for targeted mutagenesis are no longer a result of inefficient genome engineering reagents. The bottleneck now appears to be the current plant transformation technologies that are highly labour intensive. Future efforts to overcome such barriers will come from the development of nontransgenic delivery methods that utilize morphogenic regulators to increase the efficiency and ease of transformation (Lowe et al., 2016; Mookkan et al., 2017). Another approach could deliver gRNAs directly to the plant that constitutively expresses Cas9 (Figure S32a–c). The ZFN, TALEN and CRISPR/Cas9 SSN platforms comprise a genome engineering toolkit that will be useful for future functional genomics projects in soya bean and in other plant species. The efficiency, multiplex capabilities and increased overall activity of the CRISPR/Cas9 platform will likely make it the preferred SSN. However, certain scenarios may arise, such as the targeting of a highly specific genomic sequence for which no CRISPR/Cas9 or ZFN target could be identified, and in this case, the TALEN platform may provide an alternate route to mutant generation.

Material and methods

Generation of plant expression vectors for site-specific nuclease delivery

The plant expression vectors were constructed using a previously reported binary vector pN896 (Curtin et al., 2011). A 35S:OCS cassette from pArt7 was cloned into the NotI sites of the T-DNA alongside the 35S::BAR cassette for phosphinothricin herbicide selection (pSC218). The pSC218 vector was digested with SacII along with the 35S::BAR cassette for phosphinothricin herbicide selection (pSC218). The pSC218 vector was digested with SacII and XhoI to remove the 35S promoter and replaced with either Gmubi, ATUBQ or roID promoters that were amplified from templates using primers with SacII and XhoI overhangs (Grefen et al., 2010; Hernandez-Garcia et al., 2010; Wally et al., 2008). The Gateway™ reading frame cassette was removed from the pTDTO (Valdes-Lopez et al., 2008) vectors by XhoI and SpeI digestion and cloned into the XhoI and XbaI sites to generate the pSC218GG, pSC218UG and pSC218RG vectors. The resulting plant expression vectors were used for delivery of the CRISPR/Cas9 or the TAL-effector nuclease reagents.
Hairy-root and whole-plant transformation assays of soybean and *M. truncatula*

Agrobacterium rhizogenes strain, K599, was used for rapid generation of transgenic plant tissue via a soybean hairy-root ex vitro transformation assay (Figure S2) (Curtin et al., 2011; Taylor et al., 2006). The disarmed A. rhizogenes strain, 18r12, was used for a previously reported whole-plant transformation assay (Curtin et al., 2011; Paz et al., 2006). All whole-plant soybean transformations were performed in the same genetic background, the Bert cultivar. Medicago truncatula whole-plant transformation was carried out in the cultivar, R108 (HM340), using the A. tumefaciens strain, EAH105 (Table S7; Cosson et al., 2006; Curtin et al., 2017).

**CRISPR/Cas9 design and assembly**

Targets sites were identified either by the ‘find’ function with the target template (G-N20–GG) in VectoNTI (Life Technologies, CA) or were designed using the Broad Institute’s sgRNA designer website (Table S4) (Doench et al., 2014). Guide RNAs for each target were engineered using a PNK oligo annealing assay and cloned into three binary vectors using Gateway™ cloning reactions. A detailed assembly method has been previously published (Curtin et al., 2017).

**PCR, target amplicon cloning and the sequence (TACAS) and heteroduplex assays**

All PCRs were carried out using the Hot Start Plus master mix (Qiagen) according to the manufacturer’s guidelines. For the TACAS assay, PCR amplicons were purified using the QiAquick PCR Purification Kit (Qiagen) and cloned into pGem-T-Easy cloning vector (Promega). Colony PCR amplicons were diluted 1 : 12 with sterile deionized water and directly sequenced. The heteroduplex assay has been previously reported (Jacobs et al., 2015; Zhu et al., 2014) and can viewed online under the Method S3 section. Primers used in this report are provided in Table S5.

**Analysis of transgene insertion sites and targeted and off-target mutations**

The whole-genome sequence (WGS) analyses used to identify mutations at targeted loci and to detect the genomic location of transgene integration sites have been previously reported (Anderson et al., 2016; Srivastava et al., 2014). A more detailed description of this protocol can be found online in the Method S4. Putative off-target mutagenesis activity in transformed plants was manually screened using WGS data, specifically via manually analysing the generated list of potential off-target sites for each construct. The list was generated using the CRISPR-P site (http://cibibi.hzau.edu.cn/crispr/) (Table S6; Lei et al., 2014).

**Phenotypic characterization of mutants**

Detailed methods for characterization mutant of phenotypes that include drought assay, SL-qPCR and SEM sample preparation can be viewed online under the Method S5.

**Acknowledgements**

We thank Michelle Christian, Yong Zhang, Nicholas Baltes and Paul Atkins for helpful TALEN and CRISPR/Cas9 construction advice and guidance. We would also like to thank Owen Wally for the roID promoter, Yiping Qi for the UBQ10 promoter, Gail Celio from the University Imaging Centers, University of Minnesota, for the imaging work and Diana Trujillo for photography. This work was supported by the Minnesota Research and Promotion Council (projects 501-6120-534 and 509-7520-534).

**Conflict of interest**

Čermák and Voytas are named inventors on several patents involving the use of TALENs to create targeted genome modifications.

**References**

Abe, M., Yoshikawa, T., Nosaka, M., Sakakibara, H., Sato, Y., Nagato, Y. and Itoh, J. (2010) WAVY LEAF1, an ortholog of Arabidopsis HEN1, regulates shoot development by maintaining MicroRNA and transcribing small interfering RNA accumulation in rice. *Plant Physiol.* 154, 1333–1346.

Anderson, J.E., Michno, J.-M., Kono, T.J.Y., Stec, A.O., Campbell, B.W., Curtin, S.J. and Stupar, R.M. (2016) Genomic variation and DNA repair associated with soybean transgenesis: a comparison to cultivars and mutated plants. *BMC Biotechnol.* 16, 41.

Ariket, S., Xia, R., Kakranam, A., Huang, K., Zhai, J., Yan, Z., Valdés-López, O. et al. (2014) An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. *Plant Cell* 26, 1–19.

Baltes, N.J., Gil-Humanes, J., Cermák, T., Atkins, P.A. and Voytas, D.F. (2014) DNA replications for plant genome engineering. *Plant Cell* 26, 151–163.

Barton, D.A., Roovers, E.F., Fouli, Q., da Fonseca, G.C., Reis, R.S., Jackson, C., Overall, R.L. et al. (2017) Live cell imaging reveals the relocation of siRNA binding. *Mol. Plant Microbe Interact.* 6, 435–443.

Bhaya, D., Davison, M. and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* 45, 273–297.

Bustos-Sanmamed, P., Babar, J., Hartmann, C., Crespi, M. and Lelandais-Briere, C. (2013) Small RNA pathways and diversity in model legumes: lessons from genomics. *Front. Plant Sci.* 4, 236.

Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., Han, T. et al. (2015) CRISPR/Cas9-mediated genome editing in soybean hairy roots. *PLoS ONE* 10, e0136064.

Cai, Y., Chen, L., Liu, X., Guo, C., Sun, S., Wu, C., Jiang, B. et al. (2017) CRISPR/Cas9-mediated targeted mutagenesis of GmFT2α delays flowering time in soybean. *Plant Biotechnol. J.* https://doi.org/10.1111/pbi.12758.

Cermák, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A. et al. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39, e82.

Cermák, T., Curtin, S., Gil-Humanes, J., Cegan, R., Starker, C.G., Kono, T.J.Y., Matthe, J.W. et al. (2017) A multi-purpose toolkit to enable advanced genome engineering in plants. *Plant Cell* 29, 1196–1217.

Chen, X., Liu, J., Cheng, Y. and Jia, D. (2002) HEN1 functions pleiotropically in Arabidopsis development and acts in C function in the flower. *Annu. Rev. Genet.* 36, 75–93.

Cho, Y.B., Jones, S.I. and Vodkin, L.O. (2017) Mutations in Argonaute5 illuminate epistatic interactions of the K and L loci leading to seedling color patterns in glycine max. *Plant Cell* 29, 708–725.

Christian, M., Cermák, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. et al. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186, 757–761.

Cong, L., Ran, F., Cox, D., Lin, S. and Barretto, R. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823.

Cosson, V., Durand, P., D’Erfurth, I., Kondorosi, A. and Ratet, P. (2006) Medicago truncatula transformation using leaf explants. *Methods Mol. Biol.* 343, 115–127.

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.
Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1 The plant expression vectors used for the delivery of CRISPR/Cas9 and TALEN site-specific nucleases.

Figure S2 Hairy-root transformation protocol for rapid introduction of transgenes into soybean.

Figure S3 The yeast cleavage assay used to screen TAL-effector nuclease activity.

Figure S4 PCR-digest and hetero-duplex gel analysis of WPT590 soybean plants.

Figure S5 TACAS assay mutation data for WPT590-1, WPT590-2 and WPT590-4 plants.

Figure S6 Whole genome sequence analysis of Drb2a (Glyma.12g075700) target sites in Bert, WPT590-1, WPT590-1-1, WPT590-1-2, WPT590-2, WPT590-4, WPT590-4-1 and WPT590-4-2 plants.

Figure S7 Whole genome sequence analysis of Drb2b (Glyma.11g1459900) target sites in Bert, WPT590-1, WPT590-1-1, WPT590-1-2, WPT590-2, WPT590-4, WPT590-4-1 and WPT590-4-2 plants.

Figure S8 The target site encodes the start of the highly conserved second double-stranded RNA-binding motif (dsRBM).

Figure S9 Coverage of reads mapping to the reference genome surrounding transgene at chromosome 4.

Figure S10 Coverage of reads mapping to the reference genome surrounding transgene at chromosome 13.

Figure S11 Coverage of reads mapping to the reference genome surrounding transgene at chromosome 15.

Figure S12 Phenotype analysis of the drb2ab mutant.

Figure S13 Screening of T0 and T1 soybean plants for CRISPR-mediated targeted mutations at Dcl3a.