



CRISPR/Cas9-Based Gene Editing Using Egg Cell-Specific Promoters in Arabidopsis and Soybean

Na Zheng^{1,2}, Ting Li³, Jaime D. Dittman⁴, Jianbin Su², Riqing Li², Walter Gassmann², Deliang Peng¹, Steven A. Whitham^{4*}, Shiming Liu^{1*} and Bing Yang^{2,5*}

¹ State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China, ² Division of Plant Sciences, Christopher S. Bond Life Sciences Center, and Interdisciplinary Plant Group, University of Missouri, Columbia, MO, United States, ³ The Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, United States, ⁴ Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA, United States, ⁵ Donald Danforth Plant Science Center, St. Louis, MO, United States

OPEN ACCESS

Edited by:

Sandeep Kumar,
Corteva Agriscience™, United States

Reviewed by:

Yanfei Mao,
Shanghai Institutes for Biological
Sciences (CAS), China
Qi-Jun Chen,
China Agricultural University, China

*Correspondence:

Steven A. Whitham
swhitham@iastate.edu
Shiming Liu
smliuhn@yahoo.com
Bing Yang
yangbi@missouri.edu

Specialty section:

This article was submitted to
Plant Biotechnology,
a section of the journal
Frontiers in Plant Science

Received: 11 April 2020

Accepted: 19 May 2020

Published: 16 June 2020

Citation:

Zheng N, Li T, Dittman JD, Su J,
Li R, Gassmann W, Peng D,
Whitham SA, Liu S and Yang B (2020)
CRISPR/Cas9-Based Gene Editing
Using Egg Cell-Specific Promoters
in Arabidopsis and Soybean.
Front. Plant Sci. 11:800.
doi: 10.3389/fpls.2020.00800

CRISPR/Cas9-based systems are efficient genome editing tools in a variety of plant species including soybean. Most of the gene edits in soybean plants are somatic and non-transmissible when Cas9 is expressed under control of constitutive promoters. Tremendous effort, therefore, must be spent to identify the inheritable edits occurring at lower frequencies in plants of successive generations. Here, we report the development and validation of genome editing systems in soybean and Arabidopsis based on Cas9 driven under four different egg-cell specific promoters. A soybean ubiquitin gene promoter driving expression of green fluorescent protein (GFP) is incorporated in the CRISPR/Cas9 constructs for visually selecting transgenic plants and transgene-evicted edited lines. In Arabidopsis, the four systems all produced a collection of mutations in the T2 generation at frequencies ranging from 8.3 to 42.9%, with egg cell-specific promoter AtEC1.2e1.1p being the highest. In soybean, function of the gRNAs and Cas9 expressed under control of the CaMV double 35S promoter (2x35S) in soybean hairy roots was tested prior to making stable transgenic plants. The 2x35S:Cas9 constructs yielded a high somatic mutation frequency in soybean hairy roots. In stable transgenic soybean T1 plants, AtEC1.2e1.1p:Cas9 yielded a mutation rate of 26.8%, while Cas9 expression driven by the other three egg cell-specific promoters did not produce any detected mutations. Furthermore, the mutations were inheritable in the T2 generation. Our study provides CRISPR gene-editing platforms to generate inheritable mutants of Arabidopsis and soybean without the complication of somatic mutagenesis, which can be used to characterize genes of interest in Arabidopsis and soybean.

Keywords: CRISPR/Cas9, gene editing, egg cell-specific promoter, Arabidopsis, *Glycine max*, soybean

INTRODUCTION

With the advent of genome editing technologies, such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated protein (Cas) (CRISPR/Cas) in particular, targeted mutagenesis and precise base changes in genomes of interest can be achieved in ways that

were unimaginable 10 years ago (Zhang et al., 2018). These engineered nucleases can generate double-stranded DNA breaks (DSBs) at pre-chosen genomic loci, and repairs to the DSBs *in vivo* lead to site-specific genetic alterations. Two main pathways are used to repair DSBs *in vivo*: error-prone non-homologous end joining (NHEJ) and error-free homology-directed repair (HDR) in the presence of template DNA. The former is the predominant event that introduces insertions/deletions (indels) that range from one to hundreds of base pairs (Voytas, 2013).

The type II CRISPR/Cas9 system from *Streptococcus pyogenes* is the first described and the most popular CRISPR system for genome editing. CRISPR/Cas9 consists of two components, the Cas9 nuclease and a chimeric single guide RNA (gRNA) derived from the fusion of a crRNA (CRISPR RNA) and a trans-activating crRNA preceded by a spacer (or guide) sequence of 18–20 nucleotides complementary to the target DNA (or protospacer). The Cas9 protein cleaves the target DNA to cause DSBs predominantly located 3 bp upstream of the protospacer adjacent motif (PAM) sequence (5'-NGG-3'). Due to the ease of assembly and high frequency of inducing mutations, the CRISPR/Cas9 system is widely used for gene editing in various organisms, including yeast, mouse, fish, human cells, and plant species (Jinek et al., 2012; Cong et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Li et al., 2013; Mali et al., 2013; Nekrasov et al., 2013; Shan et al., 2013).

Soybean (*Glycine max*) is one of the most economically important crops for food, vegetable oil and animal feed. As a paleopolyploid species, gene function studies in soybean are frequently complicated by genetic redundancy, in addition to low frequencies of genetic transformation. Nevertheless, CRISPR/Cas9 genome editing systems have been successfully utilized in soybean since the first demonstrations in 2015. It has been applied to create targeted mutations in hairy roots, somatic embryos and stable transgenic plants (Jacobs et al., 2015; Li et al., 2015, 2019; Michno et al., 2015; Sun et al., 2015; Du et al., 2016; Chilcoat et al., 2017; Cai et al., 2018, 2020; Kanazashi et al., 2018; Al Amin et al., 2019; Bao et al., 2019; Campbell et al., 2019; Cheng et al., 2019; Do et al., 2019; Zhang et al., 2019; Michno et al., 2020; Wang et al., 2020; Wu et al., 2020). Various promoters have been deployed for expression of Cas9. For example, the constitutive cauliflower mosaic virus (CaMV) 35S promoter (35S) is the most used, especially in hairy roots (Jacobs et al., 2015). The soybean SCREAM M4 promoter (pM4) (Bai et al., 2020), parsley ubiquitin promoter (Kanazashi et al., 2018), and translation elongation factor 1 alpha 2 (*EF1A2*) promoter (Li et al., 2015) have also been used to constitutively activate Cas9 expression.

Compared with the constitutive promoters, germline specific promoters for Cas9 expression can improve the frequency and heritability of mutations significantly in Arabidopsis. Expression of Cas9 under control of the *DD45* (egg cell and early embryo), *Yao* (shoot apical and root meristem-active), tomato *Lat52* (pollen) and *EC* (egg cells, embryo) promoters can increase the frequency of heritable edits in the T2 generation. These promoters also reduce the rate of somatic mutations (Wang et al., 2015; Yan et al., 2015; Mao et al., 2016). The lower frequency of chimerism reduces the need to screen large numbers of individual

plants and conduct multiple generational analyses to acquire the desired mutants in Arabidopsis. The efficiency of *Agrobacterium*-mediated soybean transformation is very low, labor-intensive and time-consuming, so a high-efficiency CRISPR/Cas9 system based on germline specific promoters may reduce the chimerism and thus, the workload of characterizing edited plants.

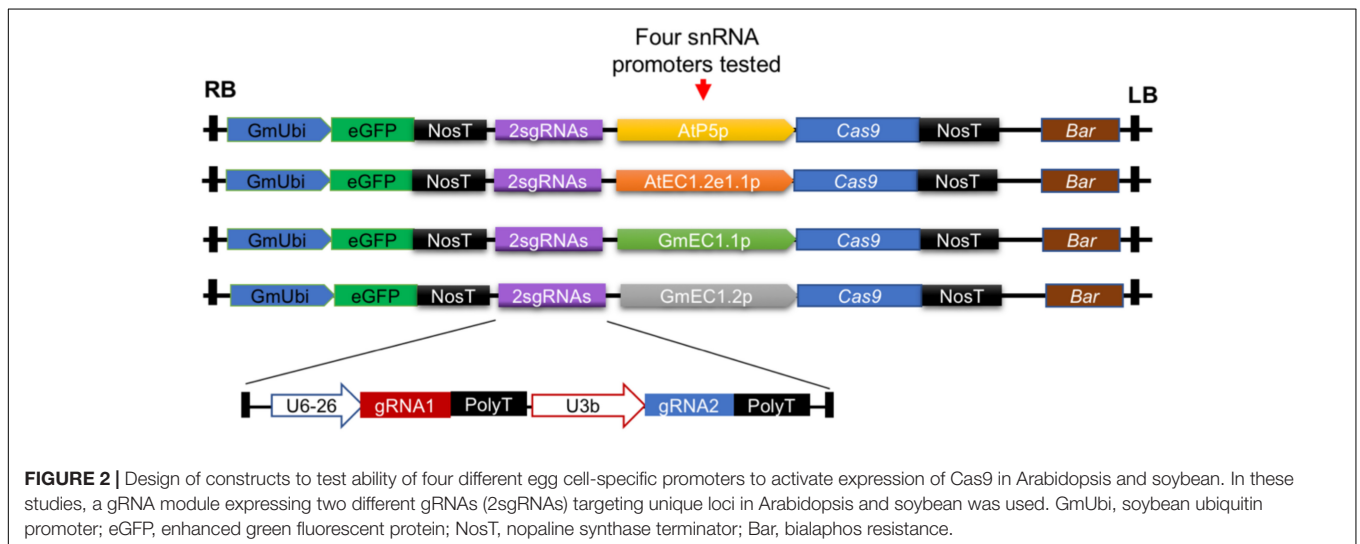
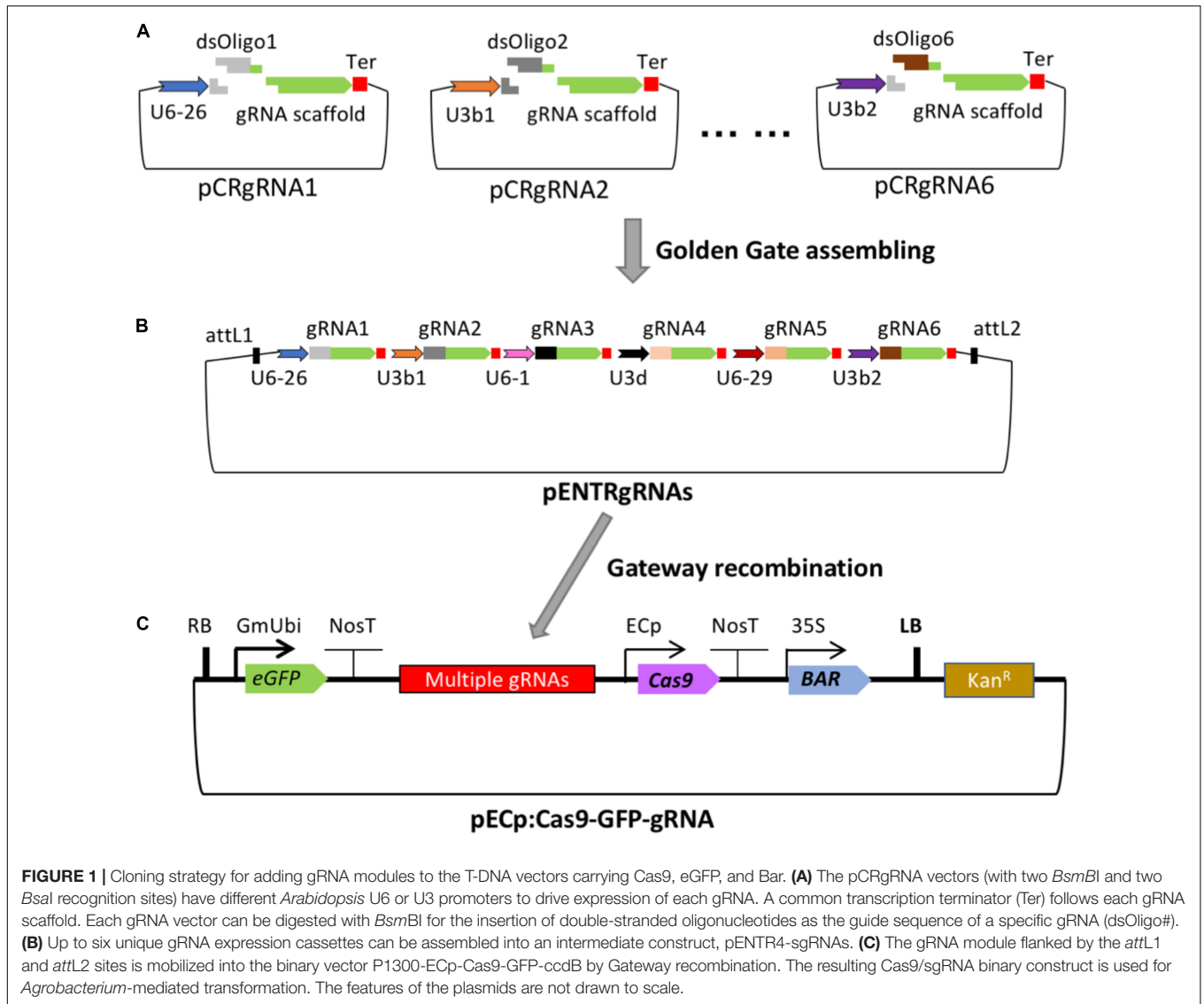
In this work, we present easy-to-use binary vector systems with Cas9 driven by egg cell-specific promoters (ECp) for efficient site-specific mutagenesis in Arabidopsis and soybean based on *Agrobacterium*-mediated transformation. In the system, a GFP marker can also be used to identify transgenic and transgene-free plants. We validated the ECp-Cas9 systems and *Agrobacterium*-mediated protocol by targeting two genes each in Arabidopsis and soybean. Our results showed that egg cell-specific promoters can induce mutations of endogenous genes in Arabidopsis and soybean, and multiple, independent mutations can be obtained from the progeny of individual single lines. We confirmed that the continuous presence of the Cas9/gRNA construct in transgenic plants can cause mutagenesis of target genes of interest in subsequent generations.

RESULTS

Construction of an ECp-Cas9/gRNA System for Genome Editing in Arabidopsis and Soybean

The cloning strategy for building the series of gRNA vectors for insertion into the destination ECp-Cas9/gRNA binary vector suitable for *Agrobacterium*-mediated transformation of Arabidopsis and soybean is shown in **Figure 1**. The gRNA vectors (pCRgRNA1 to pCRgRNA6) were designed and constructed to generate individual gRNA units each consisting of a small nuclear RNA (snRNA, U6 or U3) promoter and a gRNA followed by a poly-T terminator. The multiple cloning sites between the U6 or U3 and the gRNA scaffold sequence contain two *BsmBI* sites which facilitate an insertion of a double-stranded DNA fragment with two unique overhangs generated after annealing two complementary oligonucleotides. The inserted sequence in each gRNA vector forms the spacer or guide sequence of a gRNA gene designed specifically to target the genomic locus of interest (**Figure 1A**). The six gRNA cassettes were designed and constructed such that, after digestion with *BsaI*, each unit has overhanging ends compatible to the ends of the adjacent units, which are assembled into the intermediate pENTR_ccdB vector using Golden Gate assembly (**Supplementary Table S1**). In order to have the flexibility of making constructs containing fewer than 6 gRNA units, two additional vectors (pCRgRNA2T and pCRgRNA4T) were designed and constructed to enable 2- and 4-gRNA units (**Supplementary Table S1**). Therefore, the vectors can be used to construct gRNA modules consisting of 2, 4, or 6 gRNAs (**Figure 1B**). In this study, we used the 2-gRNA version to target two different genes at once for mutagenesis (**Figure 2**).

For egg cell-specific expression of Cas9, promoters of four egg cell-specific genes were fused to the Cas9 coding sequence to establish the four CRISPR/Cas9 systems used in this study



(Figure 2). This strategy was used to allow us to identify the promoter that worked best for genome editing in both Arabidopsis and soybean. The promoter from the AT1G71470 locus of Arabidopsis, referred to as AtP5p, was used for expression of the Arabidopsis codon-optimized Cas9. Another Arabidopsis promoter, referred to as AtEC1.2e1.1p, was adapted from the fusion of AtEC1.1 and AtEC1.2 *cis*-regulatory elements (Wang et al., 2015). Additionally, two soybean promoters [GmEC1.1p and GmEC1.2p from loci encoding *G. max* egg cell-secreted protein 1.1 (LOC100801164) and *G. max* egg cell-secreted protein 1.2 (LOC102670289), respectively] were identified and used for Cas9 expression. The Cas9 expression plasmids each contained a cassette of the *ccdB* gene flanked by the Gateway recombination sequences *attL1* and *attL2* (Supplementary Figure S1). The gRNA modules from pENTR-gRNAs could be mobilized to the individual destination vectors pGW-ECp:Cas9-GmUbi:GFP through Gateway recombination, resulting in a single binary vector pECp:Cas9/gRNA for *Agrobacterium*-mediated plant transformation with the *bar* gene driven by the CaMV 35S promoter as the transformation selection marker (Figures 1C, 2).

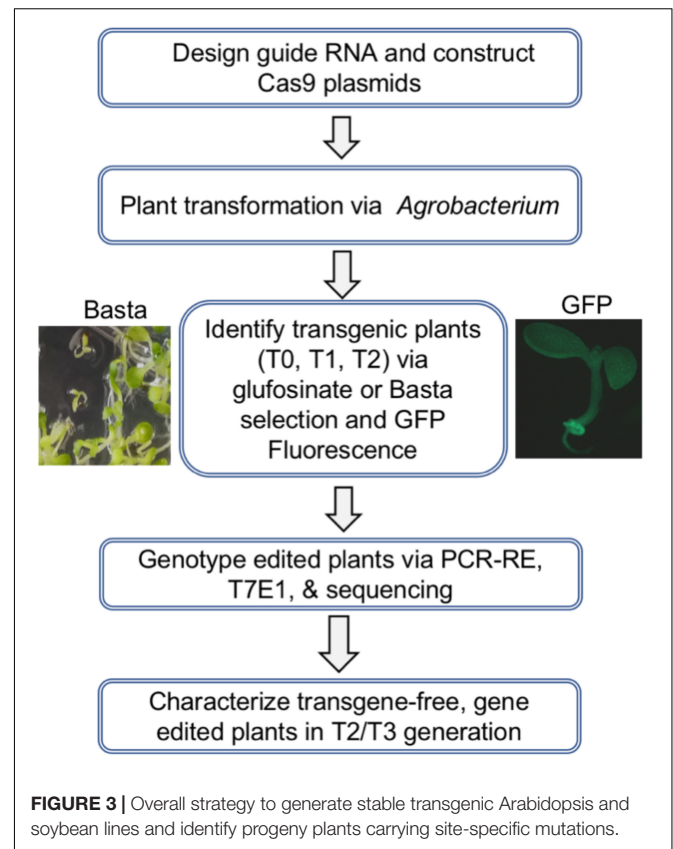
The guide sequences of gRNA genes were selected and designed based on the *Arabidopsis thaliana* ecotype Wassilewskija-2 (WS-2) or soybean cultivar Williams 82 genome sequences using the CRISPR Genome Analysis Tool (Brazelton et al., 2015)¹. The corresponding target genomic regions were PCR-amplified and confirmed by Sanger sequencing prior to gRNA design. All pCRgRNA constructs were confirmed for sequence accuracy at the insertion sites and the flanking regions by Sanger sequencing. The binary plasmids were mobilized into *Agrobacterium* strain GV3101 or EHA105 for transformation of Arabidopsis and soybean, respectively.

The overall strategy to design and apply a CRISPR/Cas9 construct for targeted mutagenesis in Arabidopsis and soybean is illustrated in Figure 3. The main steps include designing and constructing individual gRNA gene units, assembling the gRNA units into a guide RNA cassette, transferring the gRNAs into ECp:Cas9 vectors, transforming *Agrobacterium*, and performing *Agrobacterium*-mediated plant transformation. Basta resistance and GFP fluorescence were used to screen and select for the transgenic or transgene-free plants of the T0, T1, and T2 generations. The final step involves genotyping edited plants through PCR-amplification of the targeted regions followed by restriction enzyme digestion analysis (PCR-RE) or T7 endonuclease I (T7E1) assay and further confirmation of edits by Sanger sequencing (Figure 3).

Targeted Mutagenesis of *AtRPS4* and *AtRPS4B* in Arabidopsis

To investigate whether the ECp-Cas9/gRNA systems could induce site-specific mutations, we first tested the ability of each to induce mutations in Arabidopsis. The Arabidopsis genes encoding the TIR-NB-LRR proteins *AtRPS4* and *AtRPS4B* were chosen as the targets. These resistance proteins activate effector-triggered host immunity upon recognizing two bacterial effectors,

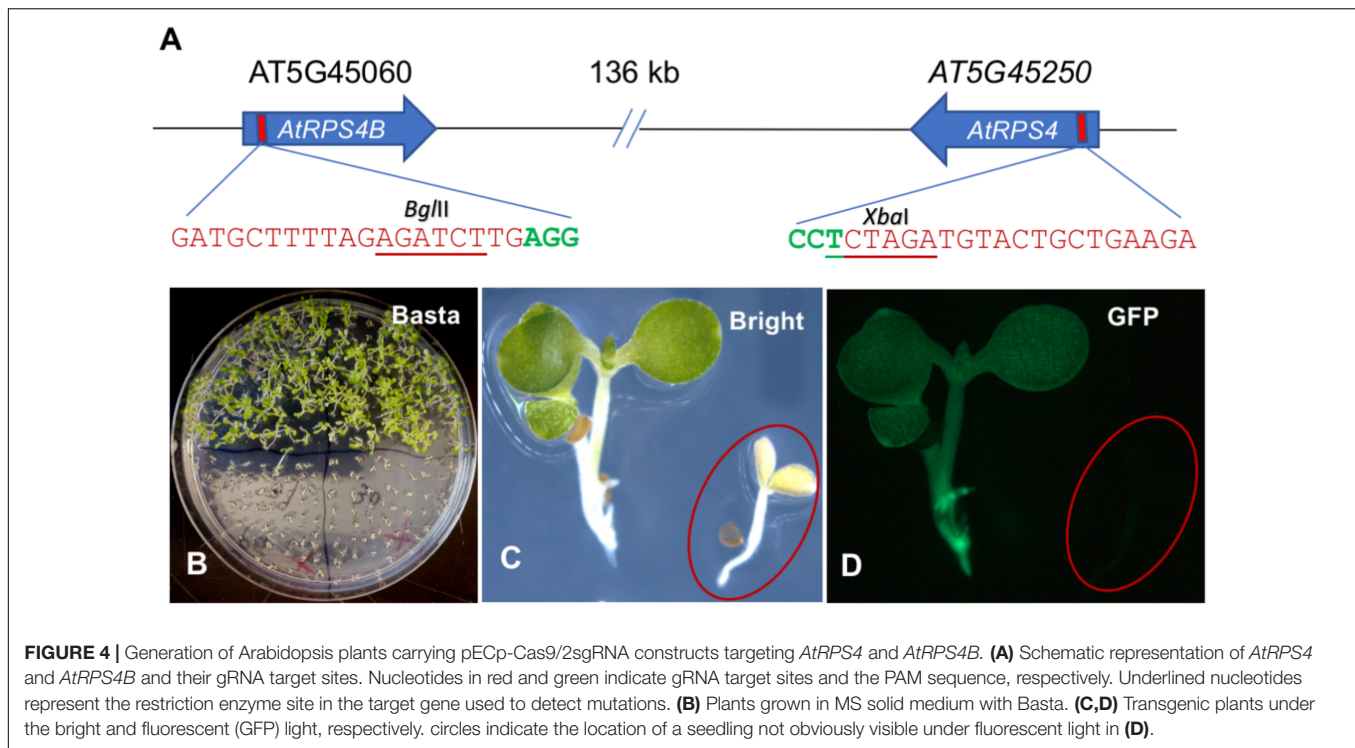
¹<http://cbc.gdcb.iastate.edu/cgat/>



AvrRps4 from *Pseudomonas syringae* and PopP2 from *Ralstonia solanacearum* (Saucet et al., 2015). The two gRNA genes (gRPS4 and gRPS4B, one gRNA for each target gene) were designed and constructed into one gRNA cassette using pCRgRNA1 and pCRgRNA2T as intermediate cloning vectors. The Cas9 cleavage site (three nucleotides upstream of the NGG PAM) of each target gene overlapped with a restriction enzyme recognition site (*Xba*I for *AtRPS4* and *Bgl*II for *AtRPS4B*), which facilitated genotyping using the PCR-RE approach (Figure 4A).

The same gRNA cassette was mobilized individually into pGW-ECp:Cas9-GmUbi:GFP recipient vectors, resulting in four different ECp:Cas9/gRNA constructs that were individually introduced into Arabidopsis through *Agrobacterium*-mediated transformation using the floral-dipping method. The *Bar* and GFP markers were used to select the transgene positive plants. Compared with non-transgenic Arabidopsis WS-2 plants, the transgenic plants grew well on the MS solid medium supplemented with Basta and showed strong green fluorescence under fluorescence microscopy (Figures 4B–D).

We first established a way to identify true transgenic plants by combining Basta resistance and GFP fluorescence presence to screen the T1 and T2 generations. Several transgenic T1 plants obtained from four different ECp:Cas9/gRNA constructs were subjected to Basta, GFP fluorescence and Cas9/gRNA-PCR screening (Supplementary Table S2). Compared to Basta screening, the GFP fluorescence was more consistent with the presence of the gRNA genes and Cas9 as further confirmed



by the PCR approach (**Supplementary Table S2**). These results indicated that transgenic plants could be best identified by using the two methods together.

We next examined the extent of gene editing in the T1 populations by PCR-amplifying the relevant genomic regions and digesting the amplicons with the restriction enzymes *Xba*I for *AtRPS4* and *Bgl*II for *AtRPS4B* (PCR-RE approach) followed by Sanger sequencing some of the randomly selected amplicons. For AtEC1.2e1.1p, among 20 transgenic T1 lines selected, 3 lines contained mutations at one target locus (*AtRPS4*) (**Supplementary Table S2**). Two mutations were heterozygous, while one was bi-allelic. The frequency of mutations for *AtRPS4* was 15%, suggesting that the T-DNA transferred into zygotes of Arabidopsis flowers started to function at an early stage. No mutation was detected at the *AtRPS4B* target locus, suggesting a low efficiency of the guide-RNA for *AtRPS4B*. However, no mutation was detected at the two target loci with the three constructs containing other three ECp driving Cas9 expression. These results indicate that site-specific mutants can be obtained in the T1 generation with one of four egg cell promoters.

Since ECp:Cas9/gRNA should continue to function in egg cells, zygotes and the early stage embryos in the reproductive stage of transgenic plants in the T1 or successive generations, we focused on detecting mutations at the target loci in the T2 generation derived from each CRISPR construct. We counted T2 plants that survived Basta selection in MS medium and with GFP fluorescence and calculated the segregation ratio from individual T1 lines (**Supplementary Table S3**). The T2 plants from several T1 lines with a 3:1 segregation ratio for each Cas9/gRNA construct were selected for genotyping using the PCR-RE approach (**Figure 5**). For AtEC1.2e1.1p, the

mutation frequencies ranged from 18.2 to 63.6% for *AtRPS4* and from 0 to 18.2% for *AtRPS4B* (**Table 1**). Genotyping and sequencing analysis of T2 plants showed that AtP5p, GmEC1.1p and GmEC1.2p induced site-specific mutations at 12.8, 25.2, and 0%, respectively, for *AtRPS4*, but induced no mutation at the *AtRPS4B* target locus (**Table 1**). The AtEC1.2e1.1p:Cas9 construct induced the highest number of mutations at both loci in comparison with the other egg cell-specific promoters. These results demonstrated that AtEC1.2e1.1p was the most efficient promoter compared to AtP5p, GmEC1.1p and GmEC1.2p for gene editing in Arabidopsis.

We also determined the exact identities of edits in individual T2 plants from the single T1 lines. Sanger sequencing was performed for seven plants from line #18 of AtEC1.2e1.1p:Cas9 construct, and five plants contained single mutations for *AtRPS4* and two contained double mutations for *AtRPS4* and *AtRPS4B*. A variety of mutations (heterozygous, homozygous and bi-allelic) occurred in the *AtRPS4* single mutants (**Table 2**). The results demonstrate that multiple independent mutations can be obtained from just a single T1 line, which could be useful for generating allelic series.

The GFP fluorescence and PCR-RE approach were also used to select transgene-free mutants in the T3 generation (**Figure 6**). Through genetic segregation, non-GFP (transgene-free) T3 plants could be recovered that carried mutations in either *AtRPS4*, *AtRPS4B* or both genes. These results suggest that it is feasible to use visible GFP fluorescence to remove the transgenic plants, then use the PCR-RE approach to identify the mutants from the previous generation through genetic segregation. This strategy can be used to obtain transgene-free mutants from either the T2 or T3 generation.

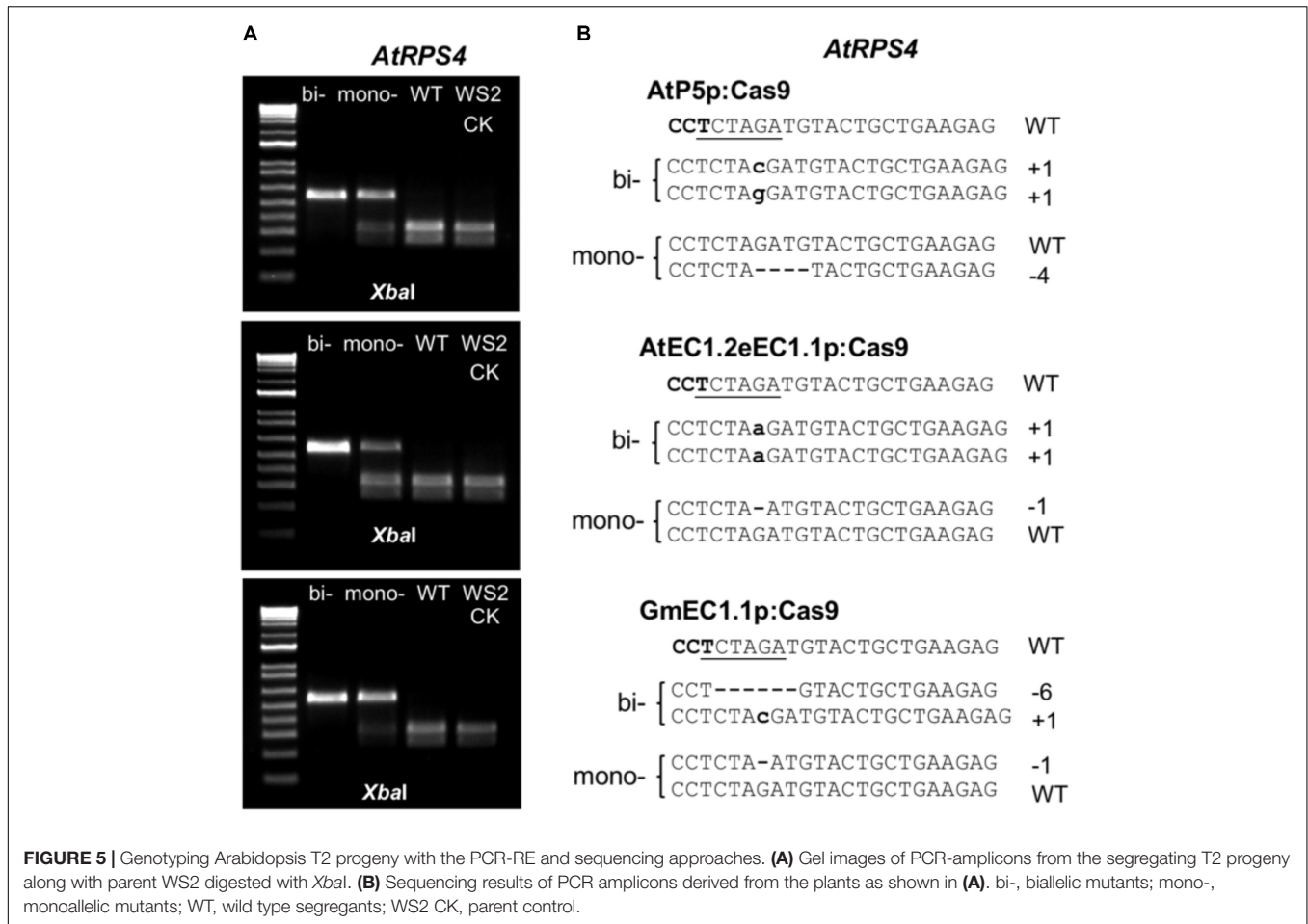


TABLE 1 | Efficiency of targeted mutagenesis by Cas9 driven by four different promoters in T2 generation of Arabidopsis.

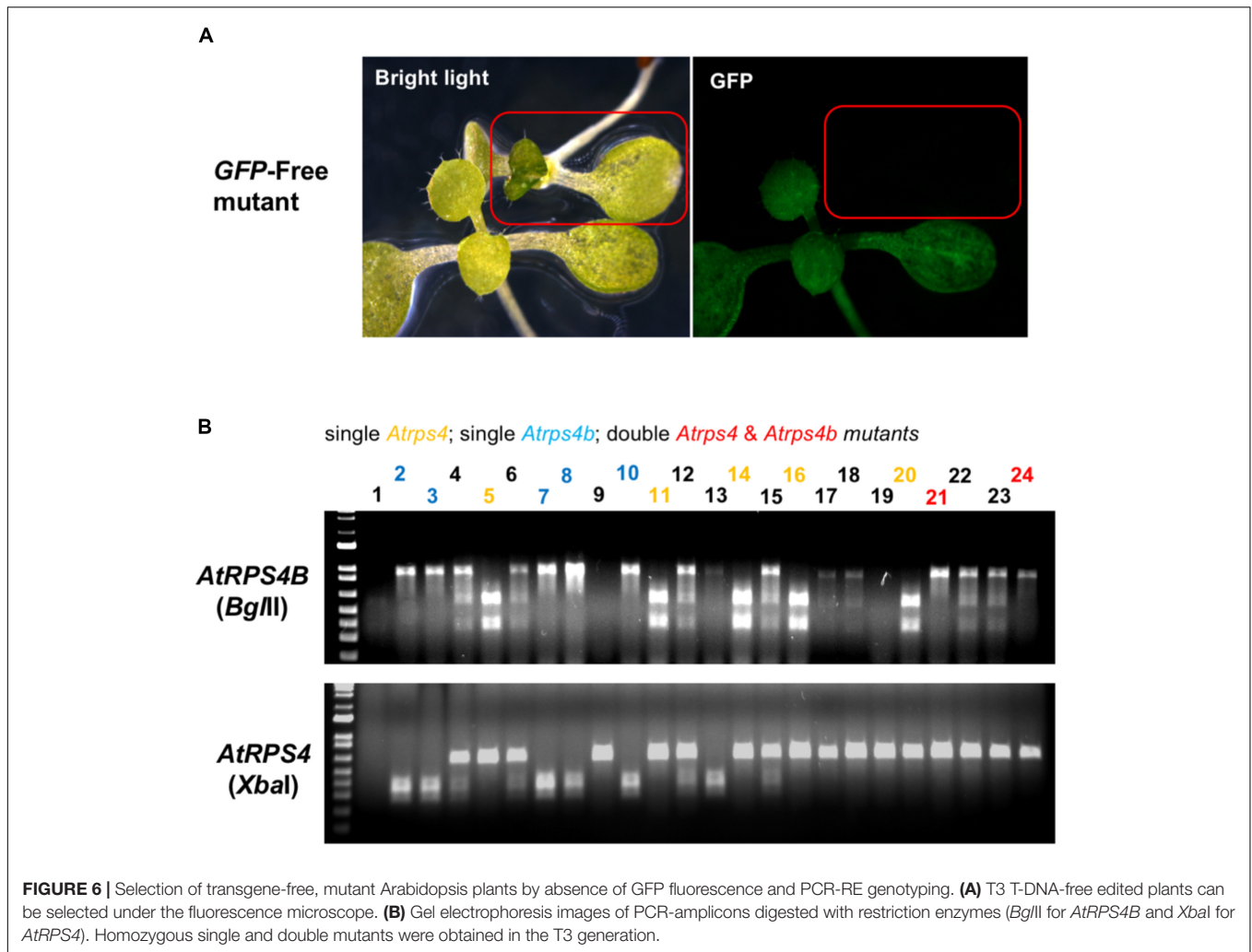
Promoter for Cas9	T1 Line	AtRPS4 number of plants	Editing efficiency	AtRPS4B number of plants	Editing efficiency	Total T2 plants
AtEC1.2e1.1p	#18	WT(4)/Heter-(3)/Homo-(2)/Bi-(2)	63.6%	WT(9)/Heter-(1)/Bi-(1)	18.2%	11
	#21	WT(4)/Heter-(3)/Bi-(2)	55.6%	WT(9)	0%	9
	#26	WT(8)/Heter-(1)/Bi-(1)	20%	WT(9)/Heter-(1)	10.0%	10
	#42	WT(9)/Bi-(2)	18.2%	WT(10)/Heter-(1)	9.0%	11
	#46	WT(3)/Heter-(1)/Bi-(3)	57.1%	WT(7)	0%	7
AtP5p	#5	WT(15)	0%	WT(15)	0%	15
	#11	WT(8)/Heter-(2)	20.0%	WT(10)	0%	10
	#20	WT(7)/Heter-(3)/Bi-(1)	36.4%	WT(11)	0%	11
	#27	WT(12)	0%	WT(12)	0%	12
GmEC1.1p	#41	WT(12)/Heter-(1)	7.8%	WT(13)	0%	13
	#10	WT(9)/Heter-(3)	25.0%	WT(12)	0%	12
	#36	WT(13)	0%	WT(13)	0%	13
	#42	WT(5)/Heter-(4)/Bi-(3)	63.6%	WT(12)	0%	12
GmEC1.2p	#52	WT(7)/Heter-(1)	12.5%	WT(8)	0%	8
	#72	WT(9)/Heter-(2)/Bi-(1)	25.0%	WT(12)	0%	12
	#2	WT(27)	0%	WT(27)	0%	27
	#3	WT(24)	0%	WT(24)	0%	24
	#4	WT(4)	0%	WT(4)	0%	4
	#5	WT(19)	0%	WT(19)	0%	19

Hetero-, heterozygous mutants; Bi-, biallelic mutants; WT, wildtype segregants.

TABLE 2 | The representative mutations induced by AtEC1.2e1.1p:Cas9 in T2 plants of #18.

	<i>AtRPS4</i>		<i>AtRPS4B</i>	
WS2	CCT CTAGATG TACT GCTGAAGAG	WT	GTGATGCTTTTAGAGATC TTGAGG	WT
AtEC #18-1	CCTCTA g GATG TACT GCTGAAGAG	+1	GTGATGCTTTTAGAGATC TTGAGG	WT
	CCTCTA g GATG TACT GCTGAAGAG	+1	GTGATGCTTTTAGAGATC TTGAGG	WT
AtEC #18-2	CCTCTA-TGTACTGCTGAAGAG	-2	GTGATGCTTTTAGAGATC TTGAGG	WT
	CCTCTAGATG TACT GCTGAAGAG	WT	GTGATGCTTTTAGAGATC TTGAGG	WT
AtEC #18-3	CCTCTA-ATGTACTGCTGAAGAG	-1	GTGATGCTTTTAGAGATC TTGAGG	WT
	CCTCTA-ATGTACTGCTGAAGAG	-1	GTGATGCTTTTAGAGATC TTGAGG	WT
AtEC #18-4	CCTCTA a GATG TACT GCTGAAGAG	+1	GTGATGCTTTTAGAGATC TTGAGG	WT
	CCTCTAGATG TACT GCTGAAGAG	WT	GTGATGCTTTTAGAGATC TTGAGG	WT
AtEC #18-5	CCTCTA a GATG TACT GCTGAAGAG	+1	GTGATGCTTTTAGAGATC TTGAGG	WT
	CCTCTA-ATGTACTGCTGAAGAG	-1	GTGATGCTTTTAGAGATC TTGAGG	WT
AtEC #18-6	CCTCTA-ATGTACTGCTGAAGAG	-1	GTGATGCTTTTAGAGAT-TGAGG	-2
	CCTCTAGATG TACT GCTGAAGAG	WT	GTGATGCTTTTAGAGATC TTGAGG	+1
AtEC #18-7	CCTCTA c GATG TACT GCTGAAGAG	+1	GTGATGCTTTTAGAGATC TTGAGG	WT
	CCTCTA g GATG TACT GCTGAAGAG	+1	GTGATGCTTTTAGAGAT-TTGAGG	-1

The green letters are PAM sequences of Cas9 target sites. Lower letters are insertions and dashed lines are deletions induced by Cas9.



ECp:Cas9/gRNA Constructs Induce Mutagenesis of Two *GmAGO7* Genes in Soybean

To test whether our ECp-Cas9/gRNA systems could introduce mutations at genomic target loci in soybean, we chose *GmAGO7a* (Glyma.01G053100) and *GmAGO7b* (Glyma.02G111600) for targeted mutagenesis. In plants, ARGONAUTE (AGO) proteins are associated with small RNA (sRNA) mediated repression of gene expression through either direct cleavage or other mechanisms, such as target destabilization or translational repression (Carbonell et al., 2012; Meister, 2013). ARGONAUTE7 (AGO7), a key regulator in the transacting small interfering RNAs (ta-siRNA) pathway, plays a conserved role in controlling leaf pattern among species. In *Arabidopsis*, the *ago7* mutants display increased leaf length and downward-curved leaf margin due to accelerated juvenile-to-adult transition; however, the mutants did not show obvious defects in leaf polarity (Hunter et al., 2003, 2006; Fahlgren et al., 2006). Overexpressing *SLAGO7* in tomato exhibited pleiotropic phenotypes, including improved axillary bud formation, altered leaf morphology and inflorescence architecture, and increased fruit yield (Lin et al., 2016). Loss-of-function of *Mtago7* resulted in lobed leaf margins and more widely spaced lateral organs in *Medicago truncatula* (Zhou et al., 2013). There are two *AGO7* paralogous genes (*GmAGO7a* and *GmAGO7b*) in the annotated Williams 82 genome².

We first verified the efficiency of the CRISPR-Cas9 system and activity of the gRNA in the soybean hairy root system. The CaMV double 35S (2x35S) promoter was used to drive expression of Cas9, and two gRNAs were designed to target a unique sequence in each gene. Cassettes expressing two different pairs of gRNAs targeting *GmAGO7a* and *GmAGO7b* – gAGO7a1 and gAGO7b2 (named gAGO7a1/b2) and gAGO7a2 and gAGO7b1 (named gAGO7a2/b1) – were constructed and each was recombined into the Cas9 binary vector, resulting in constructs that simultaneously targeted *GmAGO7a* and *GmAGO7b* (Figure 7).

The resulting two constructs were individually introduced into *Agrobacterium rhizogenes* strain K599 to induce hairy roots from infected soybean cotyledons. Hairy roots from individual cotyledons were collected and pooled for genomic DNA extraction. Individual DNA samples were used to PCR-amplify the relevant regions using the gene-specific primers for *GmAGO7a* and *GmAGO7b*. The T7 endonuclease I (T7E1) analysis, an assay that involves using T7E1 to digest the PCR amplicons of target genomic loci (Guschin et al., 2010), revealed a range of 80 to 100% mutation frequencies for gAGO7a1 and gAGO7a2, about 20% for gAGO7b1 and no detected mutation for gAGO7b2 (Supplementary Figure S2). The results indicated that the gRNAs (gAGO7a1 and gAGO7a2) for *GmAGO7a* and their promoters (U6-26 and U3b) were highly active, but the gAGO7b1 and gAGO7b2 gRNAs were less active.

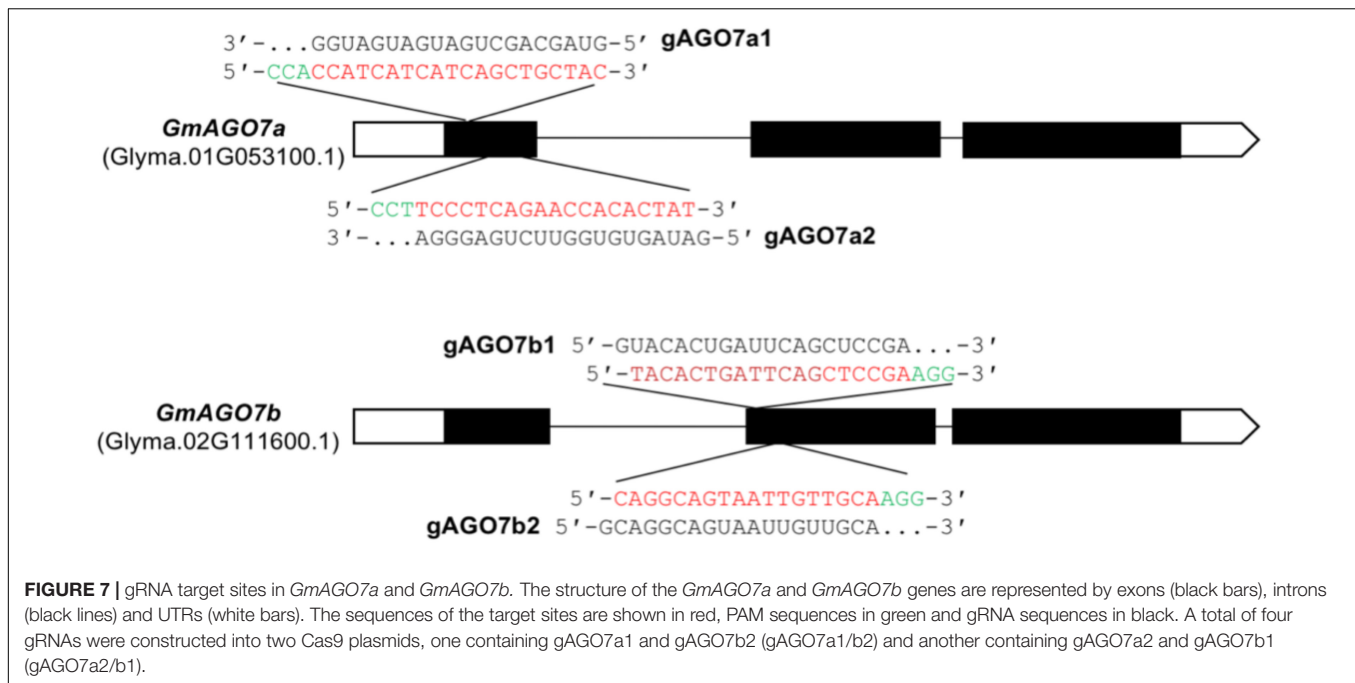
To validate the ability of ECp:Cas9/gRNA to induce mutagenesis and examine the correlation of gRNA activities

between hairy roots and stable transgenic plants, we used five constructs expressing Cas9 under different egg cell-specific promoters and gRNAs (gAGO7a2/b1 and gAGO7a1/b2) (pAtP5p:Cas9-gAGO7a1/b2, pAtP5p:Cas9-gAGO7a2/b1, pAtEC1.2e1.1p:Cas9-gAGO7a2/b1, pGmEC1.1p:Cas9-gAGO7a2/b1, and pGmEC1.2p:Cas9-gAGO7a2/b1) for stable soybean transformation. These five constructs were introduced individually into the Williams 82 cultivar of soybean by the *A. tumefaciens*-mediated transformation method. The GFP fluorescence was used to further screen the transgenic plants from the regenerated plants that survived from the herbicide glufosinate (Figures 8A–D). In addition, PCR-amplifications of Cas9 and gRNA genes were used to identify the transgenic plants (Figure 8E). Plants were screened by resistance to glufosinate (*bar*), GFP fluorescence and PCR-amplification, and 2 and 3 independent T0 transgenic plants were obtained for AtP5p:Cas9-gAGO7a1/b2 and AtP5p:Cas9-gAGO7a2/b1, respectively (Table 3). T0 plants were grown to maturity so that we could determine the mutation efficiency in the T1 generation. The three T0 AtP5p:Cas9+gGmAGO7a2/b1 lines (ST397-1, -4, and -5), produced transgenic progeny, but none that were tested carried a mutation at either gRNA target sites. As expected, the T0 AtP5p:Cas9-gAGO7a2/b1 plants (ST397-2 and -3) that survived Bar selection but were negative for GFP fluorescence and the PCR tests for the Cas9 and gRNA transgenes produced no transgenic progeny and carried no mutations at the *GmAGO7a* or *GmAGO7b* target sites (Table 3).

For the AtEC1.2e1.1p, GmEC1.1p and GmEC1.2p constructs, 13, 2, and 1 independent T0 transgenic lines were positive for glufosinate screen (*bar*) and the PCR amplification, respectively (Table 3). Out of the 13 T0 lines from the AtEC1.2e1.1p:Cas9-gAGO7a2/b1 construct, 10 lines produced transgenic T1 progeny, and of these, five lines produced T1 progeny some of which carried mutations in *GmAGO7a*, but none had mutations in *GmAGO7b*. Among the 97 T1 progeny plants from the five lines, 26 plants contained mutations for *GmAGO7a* a mutation efficiency of 26.8% based on the T7E1 assay (Table 3). For other non-transgenic plants, we also checked for occurrence of the mutation at both target loci, and as expected no mutation was found in these plants. For the GmEC1.1p:Cas9-gAGO7a2/b1, two transgenic lines (ST411-12, -18) produced transgenic progeny, but no mutation was detected at either target site. Unfortunately, no seed was produced from the only transgenic line derived from GmEC1.2p:Cas9-gAGO7a2/b1. These results indicate that the AtEC1.2e1.1p promoter is likely the most efficient for gene editing in soybean and that the frequencies of mutations induced by the gRNA were consistent with the hairy root system.

We expected that Cas9 would be activated in egg cells, zygotes or early embryos beyond the T1 generation. Therefore, we further examined the prevalence of mutations in T2 plants produced from T1 lines that were derived from AtEC1.2e1.1p:Cas9-gAGO7a2/b1 and carried the Cas9 and gRNA transgenes. A total of 31 T1 lines produced from 6 T0 lines were selected for further analysis in the T2 generation. Twenty-two T1 lines produced T2 plants that carried inheritable transgenes and mutations, while seven T1 lines (ST410-14-19, ST410-25-22,

²phytozome.jgi.doe.gov



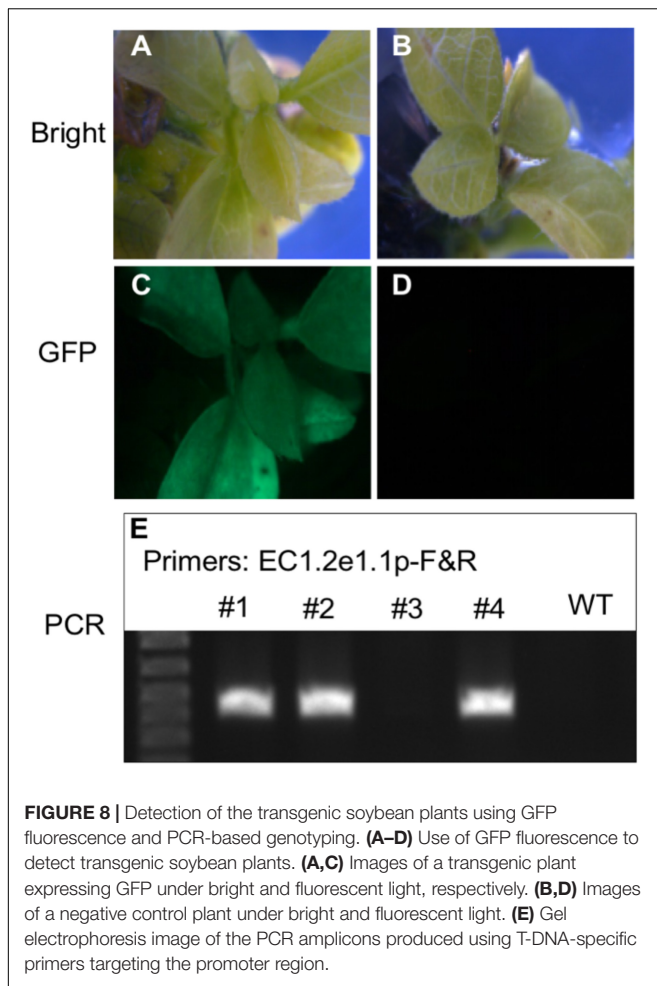
ST410-36-2, ST410-36-4, ST410-36-9, ST410-36-11, ST410-44-6) did not produce any seed (Table 4). A total of 145 out of 389 T2 plants tested positive for mutations in *GmAGO7a* by using the T7E1 assay. Sanger sequencing the PCR-amplicons from some of the T7E1 positive plants revealed site-specific mutations in *GmAGO7a* (Figure 9). Most mutants were heterozygous, but 11 were homozygous from T1 lines ST410-4-5, ST410-4-14, ST410-25-4, ST410-44-5) (Table 4). We found only one mutant for *GmAGO7b* among 389 T2 plants (ST410-14-3-7) (Table 4). We were unable to correlate a mutant phenotype with homozygous mutations in *GmAGO7a*, suggesting that there is a redundant function provided by *GmAGO7b*. The results indicate that Cas9 and gRNAs are still active in egg cells, zygotes and early embryos in the T1 and T2 generations. For construct GmEC1.1p:Cas9-gAGO7a2/b1, eight T1 lines were selected for analyses of transgene presence and occurrence of mutations in the T2 generation. No mutation was detected in either *GmAGO7a* or *GmAGO7b* among 140 T2 plants (Table 4). These results demonstrated that the AtEC1.2e1.1p promoter is the best promoter for egg cell-specific Cas9 expression and genome editing in Arabidopsis and soybean.

DISCUSSION

In the present study, we built a CRISPR/Cas9 platform for genome editing in soybean with the aid of proof-of-concept experiments in Arabidopsis. The system depends on the conditional expression of Cas9 at the reproductive stage by using an egg cell-specific promoter. The use of tissue-specific promoters is expected to have multiple benefits that includes reducing the potential toxicity associated with Cas9 expressed under the otherwise strong and constitutive promoters. In addition,

the expression of Cas9 in germ cells (egg cells, zygotes, and early embryos) results in heritable edits and reduces somatic mutations in plants derived from organogenesis, which has proven to be a feasible approach particularly in Arabidopsis (Wang et al., 2015; Yan et al., 2015; Mao et al., 2016). Our CRISPR system also features a visual marker (GFP) under a soybean ubiquitin gene promoter that can be used to avoid growing and maintaining escapes from soybean transformation projects. The system can be used for multiplex genome editing as up to 6 gRNA cassettes can be integrated into a single construct. Finally, we validated the system for site-specific mutagenesis by targeting two genes each in soybean and Arabidopsis up to the T2 and T3 generations, respectively.

Genome editing in soybean still faces several challenges although numerous studies have demonstrated its feasibility. First, efficient and reliable transformation technology in soybean has lagged behind some other crop species. Genetic transformation of soybean is still genotype dependent, and only a very limited number of genotypes are transformable (Yamada et al., 2012; Altpeter et al., 2016). The most widely and routinely used transformation platforms depend on organogenesis, a process that is based on cotyledonary-nodes as explants and *Agrobacterium*-mediated DNA delivery, to regenerate whole plants (Paz et al., 2006). This method often leads to a high occurrence of escape and chimeric plants, and it makes soybean transformation labor-intensive and expensive. The number of genuine transgenic plants is usually low from each transformation project. We observed evidence for escapes and chimerism during analyses of T0 and T1 plants in which some T0 plants that were Bar positive surviving glufosinate selection, but then tested negative for GFP and/or the Cas9 and gRNA transgenes (Table 3). In addition, some T0 lines that tested positive for Bar, GFP, and the Cas9 and gRNA transgenes were



likely chimeric, because their progeny neither inherited the transgenes nor carried mutations in the target genes (Table 3). Our soybean CRISPR/Cas9 system, therefore, is ideal, because it allows us to select the Cas9/gRNA positive plants that can be used to generate many progenies that each contain independent and inheritable mutated alleles.

We tested four different EC promoters and found only one (AtEC1.2e1.1p) was capable of inducing mutations in progenies of T1 and T2 generations in soybean. There are possible explanations for the failure of the other three promoters. First, Cas9 was expressed but not as levels sufficient to cause DSBs within the context of chromatin during the reproductive stage of soybean plants. It is worth to point out that two promoters (AtP5p and GmEC1.1p) are expected to be active, because they did enable Cas9 to induce mutagenesis in Arabidopsis. However, the third promoter (GmEC1.2p) may not be functional, because Cas9 failed to induce mutagenesis in both Arabidopsis and soybean. Future work is needed to determine whether Cas9 mRNA and protein expression is activated by the three promoters. Second, the guide RNAs used in the present study may not be the most active ones. More guide RNAs remain to be designed and tested. Finally, the number of T0 transgenic soybean plants derived from each of the three constructs was

small, which is a common limitation of soybean transformation. We obtained only eight, four and four T0 plants from AtP5p:Cas9, GmEC1.1p:Cas9, and Gm1.2p:Cas9, respectively, while 14 T0 plants were generated from AtEC1.2e1.1p:Cas9. The low number of T0 plants prevents us from making definitive conclusions about the three promoters that did not induce mutations in soybean. Future work to increase the efficiency of soybean transformation will make it possible to generate larger numbers of T0 plants with constructs carrying more gRNA and Cas9 configurations.

In the present study, we tested the correlation of gRNA activities among hairy roots and stable transgenic plants. Hairy roots yielded very high frequencies of mutagenesis for two of four gRNAs (gAGO7a1 and gAGO7a2) and relatively low or no activity by another two (gAGO7b1 and gAGO7b2). Similarly, the frequency of mutations induced by gAGO7a2 is much higher than gAGO7b1 in the stable transgenic plants. The findings suggest that the hairy root system is a quick way to test and select highly active gRNA before proceeding to stable transformation.

The second challenge facing genome editing in soybean is the genome size and complexity. Soybean, e.g., Williams 82, is a palaeopolyploid and contains a genome of ~1.15 gigabases (Schmutz et al., 2010), about eight times as larger than the *A. thaliana* genome (135 megabases) (Initiative, 2000). Prior to site-specific cleavage of the target DNA, Cas9 and gRNA first recognize the PAM sequence and then the PAM-proximal region through the complementarity between the gRNA and target strand (Sternberg et al., 2014). The implication of this mode of recognition is that the genome editing efficiency is expected to be negatively correlated to the genome size and complexity. Our data from Arabidopsis and soybean supports this notion. For instance, three of the egg cell-specific promoters driving Cas9 could produce mutants in Arabidopsis, while only one of three was able to do so in soybean at much lower percentage. Alternatively, the expression levels of Cas9 and gRNA may be higher in Arabidopsis than in soybean.

It is highly likely the editing efficiency would be correlated with the abundance of Cas9/gRNA ribonucleoprotein complexes. Based on this, to further improve our soybean CRISPR/Cas9 system, several strategies and approaches can be explored. First, we are interested in approaches to increase the abundance of Cas9 in egg cells, zygotes and early embryos of the T0 or T1 generation transgenic plants. For example, it was shown that zCas9 (maize codon-optimized Cas9) coding sequence followed by the pea ribulose-1,5-bisphosphate carboxylase small subunit (*rbcS*) E9 gene terminator (*rbcS* E9t) induced edits at higher rates than zCas9 followed by the *Agrobacterium Nos* gene terminator in Arabidopsis (Wang et al., 2015). The constructs described in the present study all carried Cas9 followed by the *Nos* terminator. In addition, the incorporation of introns in gene coding regions especially at the 5' end was shown to enhance gene expression (Le Hir et al., 2003). Therefore, use of a different terminator (e.g., pea *rbcS* E9 or soybean *rbcS1* terminator) along with soybean codon-optimized Cas9+intron could increase expression level of Cas9 transcripts expressed under control of the EC promoters. It may also be possible to use soybean U6 promoters for expressing gRNAs to improve the

TABLE 3 | Genotyping analysis of T0 and T1 transgenic soybean plants from five CRISPR constructs.

CRISPR Construct	Line	T0				T1	
		Bar	GFP	Cas9	gRNA	Transgenic vs. (Total plants)	# of mutant in <i>GmAGO7a</i>
AtP5p:Cas9 (gAGO7a1/b2)	ST398-4	+	+	+	+	0(19)	None
	ST398-5	+	+	+	+	0(27)	None
AtP5p:Cas9 (gAGO7a2/b1)	ST397-1	+	+	+	+	11(26)	None
	ST397-2	+	–	–	–	0(27)	None
	ST397-3	+	–	–	–	0(25)	None
	ST397-4	+	+	+	+	11(27)	None
	ST397-5	+	+	+	+	17(27)	None
	ST397-10	+	–	–	–		T0 died
AtEC1.2e1.1p:Cas9 (gAGO7a2/b1)	ST410-4	+	ND	+	+	6(15)	2
	ST410-7	+	ND	+	+	1(18)	None
	ST410-8	+	ND	+	+	5(15)	None
	ST410-9	+	ND	–	–		T0 died
	ST410-11	+	ND	+	+	0(17)	None
	ST410-13	+	ND	+	+	2(2)	None
	ST410-14	+	ND	+	+	21(23)	7
	ST410-18	+	ND	+	+	3(3)	None
	ST410-25	+	ND	+	+	30(36)	4
	ST410-32	+	ND	+	+	11(15)	None
	ST410-35	+	ND	+	+	0(14)	None
	ST410-36	+	ND	+	+	12(12)	7
	ST410-40	+	ND	+	+	0(17)	None
	ST410-44	+	ND	+	+	6(6)	6
GmEC1.1p:Cas9 (gAGO7a2/b1)	ST411-6	+	ND	–	–	0(6)	None
	ST411-12	+	ND	+	+	8(18)	None
	ST411-17	+	ND	–	–	0(9)	None
	ST411-18	+	ND	+	+	4(11)	None
GmEC1.2p:Cas9 (gAGO7a2/b1)	ST412-20	+	ND	+	+		T0 died
	ST412-29	+	ND	–	–		T0 died
	ST412-27	+	ND	–	–		T0 died
	ST412-31	+	ND	–	–	0(14)	None

ND, no GFP detection was performed.

genome editing efficiency (Di et al., 2019). Such modifications to increase the levels of activated Cas9/gRNA complexes in germline cells are expected to increase the frequency of targeted mutagenesis without sacrificing germline-specific expression.

MATERIALS AND METHODS

Plant Materials

Arabidopsis thaliana (ecotypes Wassilewskija-2, Ws-2) plants were grown vertically on half-strength Murashige and Skoog (1/2 MS) plates at pH 5.6–5.8 (adjusted with 1 N KOH), supplemented with 0.85% (w/v) agar and 1% (w/v) sucrose. All plants were grown at 22°C under long-day conditions (16 h light/8 h dark).

The soybean cultivar Williams 82 was used for hairy root transformation and whole plant stable transformation. Soybean seeds were surface-sterilized for 14–16 h with chlorine gas. Seeds were germinated on 1/4 Gamborg's solid medium under long-day conditions (16-h light/8-h dark, 28°C/24°C) in growth

chambers. After 5 days, healthy plants were selected for hairy root transformation. Soybean plants were grown in clay pots containing Pro-mix Bx Biofungicide potting mix supplemented with Osmocote slow-releasing fertilizer (14-14-14) and were grown in greenhouse to maturity under a photoperiod of 16 h light and 8 h dark at 28–24°C.

Vector Construction

Different ECp were used to drive Cas9 expression in this study. The *Arabidopsis* AtP5 promoter (919 bp, Chr1:26929821...26930831 in TAIR 10) fused to Cas9 (pCambia:AtP5p:Cas9) was constructed in pCambia3300. This vector was modified by inserting the cassette of *attR1-ccdB-attR2*, which is used to insert gRNA modules by Gateway recombination, resulting in the destination vector, pGW-AtP5p:Cas9. The AtEC1.2e1.1p ECp (1,362 bp, Chr2:9282423...9283302 + Chr1:28810535...28811064 in TAIR 10) was assembled through overlapping PCR amplification using *Arabidopsis* genomic DNA and primers. To identify soybean

TABLE 4 | Genotyping analysis of CRISPR/Cas9-induced mutations in T1 and T2 soybean plants.

CRISPR construct	T0 Line	T1		T2		Total plants		
		No.	Mutation (<i>GmAGO7a</i>)	Mutant plants (<i>GmAGO7a</i>)	Mutant plants (<i>GmAGO7b</i>)			
AtEC1.2e1.1p:Cas9 (<i>gAGO7a2/b1</i>)	ST410-4	#3	–	0	0	21		
		#4	–	0	0	18		
		#5	+	14 (3 Homo-)	0	20		
		#6	–	0	0	19		
		#14	+	11 (2 Homo-)	0	20		
		#15	–	0	0	15		
	ST410-14	#2	+	7	0	16		
		#3	+	0	1	15		
		#6	+	9	0	15		
		#19	+	ND	ND	ND		
	ST410-25	#3	+	7	0	16		
		#4	+	9 (2 Homo-)	0	14		
		#8	+	8	0	13		
		#22	+	ND	ND	ND		
	ST410-32	#1	–	0	0	17		
		#2	–	0	0	20		
		#5	–	0	0	17		
		#12	–	0	0	18		
	ST410-36	#1	+	1	0	14		
		#2	+	ND	ND	ND		
		#3	+	4	0	16		
		#4	–	ND	ND	ND		
		#6	+	9	0	19		
		#9	+	ND	ND	ND		
	ST410-44	#11	+	ND	ND	ND		
		#1	+	8	0	8		
		#2	+	9	0	9		
		#3	+	10	0	10		
		#4	+	19	0	19		
		#5	+	20 (4 Homo-)	0	20		
	Sum	6	31	22	145	1	389	
		GmEC1.1p:Cas9 (<i>gAGO7a2/b1</i>)	ST411-12	#2	–	0	0	19
				#5	–	0	0	17
				#17	–	0	0	18
				#18	–	0	0	18
			ST411-18	#6	–	0	0	18
#12	–			0	0	18		
#14	–			0	0	18		
#20	–			0	0	14		
Sum	2			8	0	0	0	140

ND, no seeds or DNA samples were collected.

ECp, the Arabidopsis egg cell secreted protein 1.1 (AT1G76750) was used as a query to BLAST search against the soybean reference genome to retrieve seven egg cell secreted protein sequences and their promoter sequences. Promoters of GmEC1.1 (1,446 bp, Chr20:40598179...40599767 in Wm82.a2v1) and GmEC1.2 (1,334 bp, Chr06:18296898...18298365 in Wm82.a2v1) were PCR-amplified using soybean genomic DNA as a template and gene-specific primers. GmEC1.1 and GmEC1.2 each represent one member of their clades among the seven

GmECs (**Supplementary Figure S3**). To construct a Cas9 expression cassette with the 2x35S promoter, the 2x35S promoter was PCR-amplified from the pTF101 plasmid. All the promoters were swapped with the AtP5p in pCambia:AtP5p:Cas9 at *NcoI* and *SpeI* through the Gibson cloning method, resulting in constructs containing each of the four different promoters for expression of Cas9. The GmUbi-GFP-NosT cassette was then inserted into each vector through restriction and ligation at the *HindIII* site using the standard molecular cloning

Williams 82	CTTATACCTCTTCAACAACCTTCCTTTGACTCCAC <u>CCT</u> <u>TCCCTCAGAACCACACTATC</u>	WT
ST410-4-5	CTTATACCTCTTCAACAACCTTCCTTTGACTCCACCCTTCC-TCAGAACCACACTATC	-1
ST410-4-14	CTTATACCTCTTCAACAACCTTCCTTTGACTCCACCCTTCC-----CACTATC	-10
ST410-14-19	CTTATACCTCTTCAACAACCTTCCTTTGACTCCACCCTTCC-TCAGAACCACACTATC	-1
ST410-25-4	CTTATACCTCTTCAACAACCTTCCTTTGACTCCACCCTTCC <u>c</u> CTCAGAACCACACTATC	+1
ST410-25-22	CTTATACCTCTTCAACAACCTTCCTTTGACTCCACCCTTCCCT---AACCACACTATC	-3
ST410-36-2	CTTATACCTCTTCAACAACCTTCCTTTGACTCCACCCTTCC-TCAGAACCACACTATC	-1
ST410-36-9	CTTATACCTCT-----ATC	-43
ST410-36-11	CTTATACCTCTTCAACAACCTTCCTTTGACTCCACCCTTC--TCAGAACCACACTATC	-2
ST410-44-3	CTTATACCTCTTCAACAACCTTCCTTTGACTCCACCCTTCC-TCAGAACCACACTATC	-1
ST410-44-5	CT-----CAGAACCACACTATC	-49

FIGURE 9 | Representative mutations induced by egg cell-specific expression of Cas9 in soybean plants derived from five independent T0 lines. Red letters and underlined letters in Williams 82 indicate the respective PAM and gRNA target sequences. Dashed lines and lowercase letters denote deletions and insertions, respectively, at the *GmAGO7a* target site.

methods (Ausubel et al., 1998), resulting in the GFP version of the plasmids. The primer information is provided in **Supplementary Table S4**.

For the construction of gRNA genes, the intermediate vectors pENTR4-ccdB which was modified by inserting the cassette of *attL1-ccdB-attL2*. Briefly, each gRNA vector (pCRgRNA) has a unique promoter and poly-T terminator. The specific gRNA spacer sequence was inserted at the two *BsmBI* restrict sites. To construct a specific gRNA gene, two 20 – 24 nt complementary oligonucleotides were annealed to produce a double-stranded DNA oligonucleotide (dsOligo). To make a 2-gRNAs cassette, the first dsOligo was designed with a 5' overhang of ATTG in the sense strand and a 5' of overhang AAAC in the antisense strand; and the second dsOligo was designed to contain a 5' overhang of GTCA in the sense strand and a 5' of overhang AAAC in the antisense strand. All oligonucleotides were synthesized and purchased from Integrated DNA Technology (Coralville, IA, United States). The individual gRNA cassettes were assembled into the pENTR4-ccdB vector by the Golden Gate assembly method using *BsaI*. After sequencing the guide RNA regions, the gRNA cassette was finally mobilized to four different pGW-ECp:Cas9-GFP-ccdB constructs by using Gateway LR Clonase (Thermo Fisher Scientific, Waltham, MA, United States).

Escherichia coli strain DH5 α and DB3.1 were used for molecular cloning of Cas9/gRNA constructs. *Agrobacterium tumefaciens* strain GV3101 and EHA105 were used for Arabidopsis and soybean transformation, respectively. *A. rhizogenes* strain K599 was used for the soybean hairy root transformation. *E. coli* cells were grown in Luria–Bertani (LB) medium supplemented with appropriate antibiotics at 37°C with a standard culture technique, while *Agrobacterium* strains were grown at 28°C in Luria–Bertani (LB) medium with appropriate antibiotics (Ausubel et al., 1998).

Hairy Root Transformation

Agrobacterium rhizogenes strain K599 containing the Cas9/gRNA binary constructs was used for the soybean hairy root induction.

Soybean cotyledons of cultivar Williams 82 were inoculated with the transformed K599 strain using a previously described protocol (Kereszt et al., 2007) with slight modifications. Bacterial cells were scraped from the plates and suspended in 1 mL sterile water. The blades of sterilized scissors were immersed into the bacterial suspension, then used to cut off the 1/4 of the cotyledon that was attached to the stem. The cut cotyledons were placed on a stack of one or two sterile pre-wet paper towels (with 1/4 Gamborg's liquid medium plus 200 μ g/ml Timentin) in an ice-cream box. The cotyledons were kept in a growth chamber with a photoperiod of 16-h light/8-h dark at 28°C/24°C. After cultivation for 2~3 weeks, hairy roots were collected for further analysis. Hairy roots induced by K599 lacking CRISPR construct were used as control.

Transformation of Arabidopsis and Soybean

Arabidopsis transformation was performed by using the floral dipping protocol as described (Pike et al., 2019). *Agrobacterium*-mediated transformation of a fixed number of half seed explants of Williams 82 genotype was performed at the Iowa State University Plant Transformation Facility as described (Paz et al., 2004). The plants were grown in greenhouses with a 16-h day at 28°C and an 8-h night at 24°C. The transgenic plants were further confirmed by GFP fluorescence and PCR assays on the Cas9 and gRNA transgenes.

Molecular Characterization of CRISPR Plants of Arabidopsis and Soybean

Genomic DNA samples were extracted from Arabidopsis leaves using the CTAB (cetyltrimethyl ammonium bromide) method (Murray and Thompson, 1980). Soybean genomic DNA was extracted from newly expanding primary leaves of T1 and T2 seedlings using the QIAamp Fast DNA Tissue Kit

(Qiagen, Germantown, MD, United States). Genomic DNA was used for PCR-amplification of relevant regions with specific primers flanking the target sites (**Supplementary Table S4**). PCR reaction conditions were optimized for each primer pair and are available upon request. PCR amplicons were assessed for mutations using the T7 endonuclease I (T7E1) assay or restrict enzyme digestion and Sanger sequencing. For the T7E1 assay, PCR-amplicons obtained from the transgenic tissues were mixed with the respective amplicon derived from wild type, denatured (95°C for 5 min) and reannealed (ramp down to 25°C at 5°C/min), then subjected to T7E1. For PCR-RE, the PCR-amplicons were used directly for digestion with appropriate restriction enzymes. For sequencing, the PCR amplicons derived from the T7E1-positive samples were treated with ExoSAP-IT (Affymetrix, Santa Clara, CA, United States) and subsequently evaluated by the Sanger sequencing method by the University of Missouri-Columbia DNA Core Facility³. The sequencing chromatograms were carefully examined for exact patterns that might indicate mono-allelic or bi-allelic mutations.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

³ <https://dnacore.missouri.edu>

REFERENCES

- Al Amin, N., Ahmad, N., Wu, X., Pu, T., Ma, Y., Du, X., et al. (2019). CRISPR-Cas9 mediated targeted disruption of FAD2-2 microsomal omega-6 desaturase in soybean (*Glycine max*.L). *BMC Biotechnol.* 19:9. doi: 10.1186/s12896-019-0501-2
- Altpeter, F., Springer, L. E., Bartley, A. E., Blechl, T. P., Brutnell, V., Citovsky, L. J., et al. (2016). Advancing crop transformation in the era of genome editing. *Plant Cell* 28, 1510–1520. doi: 10.1105/tpc.16.00196
- Ausubel, F., Brent, R., Kingston, D., Moore, J., Seidman, T., and Struhl, K. (1998). *Current Protocols in Molecular Biology*. New York, NY: John Wiley and Sons.
- Bai, M., Yuan, H., Kuang, P., Gong, S., Li, Z., Zhang, B., et al. (2020). Generation of a multiplex mutagenesis population via pooled CRISPR-Cas9 in soya bean. *Plant Biotechnol. J.* 18, 721–731. doi: 10.1111/pbi.13239
- Bao, A., Chen, L., Chen, S., Chen, Q., Hao, W., Guo, D., et al. (2019). CRISPR/Cas9-mediated targeted mutagenesis of GmSPL9 genes alters plant architecture in soybean. *BMC Plant Biol.* 19:131. doi: 10.1186/s12870-019-1746-6
- Brazelton, V. A., Zarecor, D. A., Wright, Y., Wang, J., Liu, K., Chen, B., et al. (2015). A quick guide to CRISPR sgRNA design tools. *GM Crops Food* 6, 266–276. doi: 10.1080/21645698.2015.1137690
- Cai, Y., Chen, X., Liu, C., Guo, S., Sun, C., Wu, B., et al. (2018). CRISPR/Cas9-mediated targeted mutagenesis of GmFT2a delays flowering time in soya bean. *Plant Biotechnol. J.* 16, 176–185. doi: 10.1111/pbi.12758
- Cai, Y., Wang, L., Chen, T., Wu, L., Liu, S., Sun, C., et al. (2020). Mutagenesis of GmFT2a and GmFT5a mediated by CRISPR/Cas9 contributes for expanding the regional adaptability of soybean. *Plant Biotechnol. J.* 18, 298–309. doi: 10.1111/pbi.13199
- Campbell, B. W., Hoyle, B., Bucciarelli, A. O., Stec, D. A., Samac, W. A., Parrott, A., et al. (2019). Functional analysis and development of a CRISPR/Cas9 allelic series for a CPR5 ortholog necessary for proper growth of soybean trichomes. *Sci. Rep.* 9:14757. doi: 10.1038/s41598-019-51240-7
- Carbonell, A., Fahlgren, H., Garcia-Ruiz, K. B., Gilbert, T. A., Montgomery, T., Nguyen, J. T., et al. (2012). Functional analysis of three *Arabidopsis*

AUTHOR CONTRIBUTIONS

BY, SW, and SL designed the research. NZ, TL, JD, JS, and RL performed the experiments. DP and WG analyzed the data. NZ and BY wrote the manuscript with input from all other authors.

FUNDING

Funding was provided by the Iowa Soybean Association (SW and BY), the National Natural Science Foundation of China (31972248 to SL), and the China Scholarship Council (NZ).

ACKNOWLEDGMENTS

We thank Diane Luth at the Iowa State University Plant Transformation Facility for soybean transformation. We thank the Iowa State University Crop Bioengineering Center for publication subvention.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.00800/full#supplementary-material>

- ARGONAUTES using slicer-defective mutants. *Plant Cell* 24, 3613–3629. doi: 10.1105/tpc.112.099945
- Cheng, Q., Dong, T., Su, T., Li, Z., Gan, H., Nan, S., et al. (2019). CRISPR/Cas9-mediated targeted mutagenesis of GmLHY genes alters plant height and internode length in soybean. *BMC Plant Biol.* 19:562. doi: 10.1186/s12870-019-2145-8
- Chilcoat, D., Liu, Z. B., and Sander, J. (2017). Use of CRISPR/Cas9 for crop improvement in maize and soybean. *Prog. Mol. Biol. Transl. Sci.* 149, 27–46. doi: 10.1016/bs.pmbts.2017.04.005
- Cong, L., Ran, D., Cox, S., Lin, R., Barretto, N., Habib, P., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Di, Y. H., Sun, Z., Hu, Q. Y., Jiang, G. H., Song, B., Zhang, S. S., et al. (2019). Enhancing the CRISPR/Cas9 system based on multiple GmU6 promoters in soybean. *Biochem. Biophys. Res. Commun.* 519, 819–823. doi: 10.1016/j.bbrc.2019.09.074
- Do, P. T., Nguyen, H. T., Bui, L. T. N., Tran, G., Stacey, J. D., Gillman, Z. J., et al. (2019). Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous GmFAD2-1A and GmFAD2-1B genes to yield a high oleic, low linoleic and α -linolenic acid phenotype in soybean. *BMC Plant Biol.* 19:311. doi: 10.1186/s12870-019-1906-8
- Du, H., Zeng, M., Zhao, X., Cui, Q., Wang, H., Yang, H., et al. (2016). Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *J. Biotechnol.* 217, 90–97. doi: 10.1016/j.jbiotec.2015.11.005
- Fahlgren, N., Montgomery, M. D., Howell, E., Allen, S. K., Dvorak, A. L., Alexander, L., et al. (2006). Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in *Arabidopsis*. *Curr. Biol.* 16, 939–944. doi: 10.1016/j.cub.2006.03.065
- Guschin, D. Y., Waite, G. E., Katibah, J. C., Miller, M. C., Holmes, K., and Rebar, E. J. (2010). A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* 649, 247–256. doi: 10.1007/978-1-60761-753-2_15

- Hunter, C., Sun, H., and Poethig, R. S. (2003). The *Arabidopsis* heterochronic gene ZIPPY is an ARGONAUTE family member. *Curr. Biol.* 13, 1734–1739. doi: 10.1016/j.cub.2003.09.004
- Hunter, C., Willmann, G., Wu, M., Yoshikawa, M., Gutiérrez-Nava, L., and Poethig, S. R. (2006). Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. *Development* 133, 2973–2981. doi: 10.1242/dev.02491
- Hwang, W., Fu, D., Reyon, M., Maeder, S., Tsai, J., Sander, R., et al. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 227–229. doi: 10.1038/nbt.2501
- Initiative, A. G. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- Jacobs, T., LaFayette, R., Schmitz, P., and Parrott, W. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* 15:16. doi: 10.1186/s12896-015-0131-2
- Jiang, W., Zhou, H., Bi, M., Fromm, B., Yang, N., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Res.* 41:e188. doi: 10.1093/nar/gkt780
- Jinek, M., Chylinski, I., Fonfara, M., Hauer, J. A., Doudna, K., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829
- Kanazashi, Y., Hirose, I., Takahashi, M., Mikami, M., Endo, S., Hirose, S., et al. (2018). Simultaneous site-directed mutagenesis of duplicated loci in soybean using a single guide RNA. *Plant Cell Rep.* 37, 553–563. doi: 10.1007/s00299-018-2251-3
- Kereszt, A., Li, A., Indrasumunar, C. D., Nguyen, S., Nontachaiyapoom, M., Kinkema, M., et al. (2007). *Agrobacterium* rhizogenes-mediated transformation of soybean to study root biology. *Nat. Protoc.* 2, 948–952. doi: 10.1038/nprot.2007.141
- Le Hir, H., Nott, A., and Moore, R. (2003). How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.* 28, 215–220. doi: 10.1016/S0968-0004(03)00052-5
- Li, C., Nguyen, J., Liu, W., Fu, C., Chen, K., Yu, R., et al. (2019). Mutagenesis of seed storage protein genes in soybean using CRISPR/Cas9. *BMC Res. Notes* 12:176. doi: 10.1186/s13104-019-4207-2
- Li, J. F., Norville, J., Aach, M., McCormack, D., Zhang, J., Bush, G. M., et al. (2013). Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* 31, 688–691. doi: 10.1038/nbt.2654
- Li, Z., Liu, A., Xing, B., Moon, J. P., Koellhoffer, L., Huang, R. T., et al. (2015). Cas9-guide RNA directed genome editing in soybean. *Plant Physiol.* 169, 960–970. doi: 10.1104/pp.15.00783
- Lin, D., Xiang, Z., Xian, Y., and Li, Z. (2016). Ectopic expression of SLAG07 alters leaf pattern and inflorescence architecture and increases fruit yield in tomato. *Physiol. Plant.* 157, 490–506. doi: 10.1111/ppl.12425
- Mali, P., Yang, K. M., Esvelt, J., Aach, M., Guell, J. E., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826.
- Mao, Y., Zhang, Z., Feng, P., Wei, H., Zhang, J., Botella, P., et al. (2016). Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in *Arabidopsis*. *Plant Biotechnol. J.* 14, 519–532. doi: 10.1111/pbi.12468
- Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* 14, 447–459. doi: 10.1038/nrg3462
- Michno, J. M., Virdi, A. O., Stec, J., Liu, X., Wang, Y., Xiong, R., et al. (2020). Integration, abundance, and transmission of mutations and transgenes in a series of CRISPR/Cas9 soybean lines. *BMC Biotechnol.* 20:10. doi: 10.1186/s12896-020-00604-3
- Michno, J. M., Wang, J., Liu, S. J., Curtin, T. J., Kono, T., and Stupar, R. M. (2015). CRISPR/Cas mutagenesis of soybean and *Medicago truncatula* using a new web-tool and a modified Cas9 enzyme. *GM Crops Food* 6, 243–252. doi: 10.1080/21645698.2015.1106063
- Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4325. doi: 10.1093/nar/8.19.4321
- Nekrasov, V., Staskawicz, D., Weigel, J. D., Jones, B., and Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 691–693. doi: 10.1038/nbt.2655
- Paz, M., Shou, Z., Guo, Z., Banerjee, A., and Wang, K. (2004). Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant. *Euphytica* 136, 167–179.
- Paz, M. M., Martinez, A. B., Kalvig, T. M., Fonger, M., and 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. doi: 10.1007/s00299-005-0048-7
- Pike, S., Gassmann, W., and Su, J. (2019). Generating transgenic *Arabidopsis* plants for functional analysis of pathogen effectors and corresponding R Proteins. *Methods Mol. Biol.* 1991, 199–206. doi: 10.1007/978-1-4939-9458-8_18
- Saucet, S. B., Ma, P., Sarris, O., Furzer, K. H., Sohn, H., and Jones, J. D. (2015). Two linked pairs of *Arabidopsis* TNL resistance genes independently confer recognition of bacterial effector AvrRps4. *Nat. Commun.* 6:6338. doi: 10.1038/ncomms7338
- Schmutz, J., Cannon, J., Schlueter, J., Ma, T., Mitros, W., Nelson, D. L., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183.
- Shan, Q., Wang, J., Li, Y., Zhang, K., Chen, Z., Liang, K., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi: 10.1038/nbt.2650
- Sternberg, S. H., Redding, M., Jinek, E. C., Greene, C., and Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62–67. doi: 10.1038/nature13011
- Sun, X., Hu, R., Chen, Q., Jiang, G., Song, H., Zhang, R., et al. (2015). Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Sci. Rep.* 5:10342. doi: 10.1038/srep10342
- Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant Biol.* 64, 327–350. doi: 10.1146/annurev-arplant-042811-105552
- Wang, L., Sun, T., Wu, L., Liu, X., Sun, Y., Cai, J., et al. (2020). Natural variation and CRISPR/Cas9-mediated mutation in GmPRR37 affect photoperiodic flowering and contribute to regional adaptation of soybean. *Plant Biotechnol. J.* doi: 10.1111/pbi.13346 [Epub ahead of print].
- Wang, Z., Xing, L., Dong, H., Zhang, C., Han, X., and Chen, Q. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. *Genome Biol.* 16:144. doi: 10.1186/s13059-015-0715-0
- Wu, N., Lu, P., Wang, Q., Zhang, J., Zhang, J., and Wang, N. (2020). Construction and Analysis of GmFAD1-1A and GmFAD2-2A soybean fatty acid desaturase mutants based on CRISPR/Cas9 technology. *Int. J. Mol. Sci.* 21:1104. doi: 10.3390/ijms21031104
- Yamada, T., Takagi, K., and Ishimoto, M. (2012). Recent advances in soybean transformation and their application to molecular breeding and genomic analysis. *Breed. Sci.* 61, 480–494. doi: 10.1270/jsbbs.61.480
- Yan, L., Wei, Y., Wu, R., Hu, H., Li, W., Yang, R., et al. (2015). High-efficiency genome editing in *Arabidopsis* using YAO promoter-driven CRISPR/Cas9 system. *Mol. Plant* 8, 1820–1823. doi: 10.1016/j.molp.2015.10.004
- Zhang, P., Du, J., Wang, Y., Pu, C., Yang, R., Yan, H., et al. (2019). Multiplex CRISPR/Cas9-mediated metabolic engineering increases soya bean isoflavone content and resistance to soya bean mosaic virus. *Plant Biotechnol. J.* doi: 10.1111/pbi.13302 [Epub ahead of print].
- Zhang, Y., Massel, I. D., Godwin, K., and Gao, C. (2018). Applications and potential of genome editing in crop improvement. *Genome Biol.* 19:210. doi: 10.1186/s13059-019-1622-6
- Zhou, C., Han, C., Fu, J., Wen, X., Cheng, J., Nakashima, J., et al. (2013). The transacting short interfering RNA3 pathway and no apical meristem antagonistically regulate leaf margin development and lateral organ separation, as revealed by analysis of an argonaute7/lobed leaflet1 mutant in *Medicago truncatula*. *Plant Cell* 25, 4845–4862. doi: 10.1105/tpc.113.117788

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zheng, Li, Dittman, Su, Li, Gassmann, Peng, Whitham, Liu and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.