Site-directed mutagenesis by biolistic transformation efficiently generates inheritable mutations in a targeted locus in soybean somatic embryos and transgene-free descendants in the T_1 generation

Kohei Adachi[1], Aya Hirose[1], Yuhei Kanazashi[1], Miki Hibara[1], Toshiyuki Hirata[2], Masafumi Mikami[3], Masaki Endo[3], Sakiko Hirose[3], Nobuyuki Maruyama[4], Masao Ishimoto[5], Jun Abe[1], Tetsuya Yamada[1*]

[1]Graduate School of Agriculture, Hokkaido University, Kita 9, Nishi 9, Kita-ku, Sapporo, Hokkaido 060-8589, Japan

[2]Field Science Center Northern Biosphere, Hokkaido University, Kita 11, Nishi 10, Kita-ku, Sapporo, Hokkaido 060-0811, Japan

[3]Plant Genome Engineering Research Unit, Institute of Agrobiological Sciences, National Agricultural and Food Research Organization, 1-2, Owashi, Tsukuba, Ibaraki 305-8634, Japan

[4]Graduate School of Agriculture, Kyoto University, Uji, Kyoto 611-0011, Japan

[5]Plant Genome Engineering Research Unit, Institute of Agrobiological Sciences, National Agricultural and Food Research Organization, 2-1-2 Kannondai, Tsukuba, Ibaraki, 305-8602, Japan

[*]Correspondence: tetsuyay@res.agr.hokudai.ac.jp
Graduate School of Agriculture, Hokkaido University, Kita 9, Nishi 9, Kita-ku, Sapporo, Hokkaido 060-8589, Japan

16-digit ORCID of the corresponding author, Tetsuya Yamada: 0000-0001-6671-4995

Abstract

The clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated endonuclease 9 (Cas9) system is being rapidly developed for mutagenesis in higher plants. Ideally, foreign DNA introduced by this system is removed in the breeding of edible crops and vegetables. Here, we report an efficient generation of *Cas9*-free mutants lacking an allergenic gene, *Gly m Bd 30K*, using biolistic transformation and the CRISPR/Cas9 system. Five transgenic embryo lines were selected on the basis of hygromycin resistance. Cleaved amplified polymorphic sequence (CAPS) analysis detected only two different mutations in e all of the lines. These results indicate that mutations were induced in the target gene immediately after the delivery of the exogenous gene into the embryo cells. Soybean plantlets (T_0 plants) were regenerated from two of the transgenic embryo lines. The segregation pattern of the *Cas9* gene in the T_1 generation, which included *Cas9*-free plants, revealed that a single copy number of transgene was integrated in both lines. Immunoblot analysis demonstrated that no Gly m Bd 30K protein accumulated in the *Cas9*-free plants. Gene expression analysis indicated that nonsense mRNA decay might have occurred in mature mutant seeds. Due to the efficient induction of inheritable mutations and the low integrated transgene copy number in the T_0 plants, we could remove foreign DNA easily by genetic segregation in the T_1 generation. Our results demonstrate that biolistic transformation of soybean embryos is useful for CRISPR/Cas9-mediated site-directed mutagenesis of soybean for human consumption.

Keywords

CRISPR/Cas9, *Glycine max*, transgenic embryo, null-segregant, nonsense mRNA decay

Introduction

Soybean (*Glycine max*, $2n = 2x = 40$) is one of the most economically important crops worldwide, because its seeds contain high-quality proteins suitable for food and forage and large amounts of lipids that can be used as sources of vegetable and industrial oil (Liu 1997). Soybean is commonly genetically modified to add desirable traits such as resistance to herbicide and/or insect damage: approximately 78% of soybean grown worldwide has been developed by genetic engineering (ISAAA, http://www.isaaa.org/). Successful and repeatable production of transgenic soybean has been achieved by using cotyledonary node explants from young seedlings or imbibed mature seeds for *Agrobacterium*-mediated transformation (Hinchee et al. 1988; Parrot et al. 1989; Olhoft and Somers 2001; Paz et al. 2006; Dang and Wei 2007). *Agrobacterium*-mediated transformation has been widely optimized for various soybean genotypes (Hinchee et al. 1988; Paz et al. 2006; Sato et al. 2007; Liu et al 2008). Recently, site-directed mutagenesis using transcription activator–like effector nucleases (TALENs) or the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated endonuclease 9 (Cas9) system is being rapidly developed as a genetic modification technique for higher plants (Cermak et al. 2011; Li et al. 2013; Nekrasov et al. 2013; Shan et al. 2013); this technique is often performed via *Agrobacterium*-mediated transformation in soybean (Haun et al. 2014; Cai et al. 2018; Curtin et al. 2018; Kanazashi et al. 2018; Bonawitz et al. 2019; Do et al. 2019).

When developing soybean with modified agronomic traits, especially for human consumption, it is desirable that foreign genes are removed from the mutant genome through genetic segregation (Cai et al. 2018; Curtin et al. 2018; Kanazashi et al. 2018). However, site-directed mutagenesis of soybean by the CRISPR/Cas9 system via *Agrobacterium*-mediated transformation sometimes produces mutations in the T_0 plant that are not transmitted to the T_1 generation (Cai et al. 2018; Curtin et al. 2018; Kanazashi et al. 2018); this occurs when the germ cells of the T_0 plant do not contain mutant alleles. Continuous induction of mutations through proceeding generations of transgenic soybean plants increases the occurrence of mutations in germ cells, ensuring the transmission of mutations to the next generation (Kanazashi et al. 2018); however, it is necessary to select plants in which the mutation is induced in the targeted locus while avoiding the genetic fixation of exogenous genes (Kanazashi et al. 2018). Therefore, development of a method for efficient induction of inheritable mutations in the T_0 plants is crucial for a simple and rapid removal of transgenes in the site-directed mutagenesis of soybean.

Biolistic transformation (i.e., transformation mediated by high-speed particle bombardment) has been achieved in various soybean tissues such as immature seed meristem (McCabe et al. 1988), somatic embryogenic tissue (Finer et al. 1992), and apical meristem (Aragão et al. 2000). Among these, somatic embryogenic tissue is often used as an explant for soybean transformation by the biolistic method (reviewed by Yamada et al. 2011). Mutations in the targeted locus have been shown to be induced in transgenic embryos 2–4 weeks after bombardment with the CRISPR/Cas9 expression

module (Jacobs et al. 2015), indicating that the mutations occurred immediately after the biolistic transformation. Early mutation induction is expected to increase the occurrence of mutations in germ cells, which would ensure the transmission of mutations to the next generations. This would remove the need for continuous induction of mutations through proceeding generations and facilitate the removal of the transgene by segregation.

Here, we used the CRISPR/Cas9 system for site-directed mutagenesis of an allergenic gene, *Gly m Bd 30K*, which encodes an oil body–associated protein (Ogawa et al. 1993; Tsuji et al. 1997). A mutant deficient in *Gly m Bd 30K* was found as a soybean spontaneous variant (Joseph et al. 2006). To develop hypoallergenic soybeans, the mutant *Gly m Bd 30K* allele in this variant has been used for gene stacking with mutant alleles of genes encoding Kunitz trypsin inhibitor and lectins (Schmidt et al. 2015). If mutant alleles of *Gly m Bd 30K* can be efficiently generated through site-directed mutagenesis, hypoallergenic varieties could be directly developed from elite soybean varieties, substantially shortening the breeding period. Here, we used biolistic transformation to achieve site-directed mutagenesis of *Gly m Bd 30K*. We evaluated induction of mutations in *Gly m Bd 30K* in transgenic embryo lines, the transmissibility of these mutations to the next generation, and the removal of transgenes in the next generation. The effects of the induced mutations were examined at the transcript and protein levels. These analyses revealed that *Cas9*-free mutants had no detectable expression of the *Gly m Bd 30K* gene and no detectable accumulation of Gly m Bd 30K protein.

Materials and methods

Vector construction

We constructed a gRNA vector using pUC19_GmU6 oligonucleotide (Kanazashi et al. 2018) and the frame sequence of the gRNA scaffold from the vector pEn-Chimera (Fauser et al. 2014). The 20-nucleotide sequence (5'-ACCCAAGTAAAGTACCAAGG-3'), which is identical to a site in exon 1 of the *Gly m Bd 30K* gene (Glyma.08G116300.1; Fig. 1), was selected as the gRNA sequence by the web-based CRISPR-P 2.0 (http://crispr.hzau.edu.cn/CRISPR2/). The expression module containing the gRNA and gRNA scaffold was digested with I-SceI restriction enzyme (NEB, Ipswich, MA, USA). We replaced the 2×P35S::OsADH5'UTR:: Cas9::Tpea3A fragment in pZH_OsU3gYSA_MMCa9 (Mikami et al., 2015) with the Pcubi::Cas9::Tpea3A fragment from pDe-CAS9 (Fauser et al. 2014) to produce a binary vector called pZH_OsU3gYSA_FFCas9. Using the I-SceI sites, we then replaced the OsU3gYSA-encoding gRNA targeting the rice YSA gene with the gRNA targeting *Gly m Bd 30K*. The resultant expression vector was named p30K-hyg.

Soybean transformation

Soybean variety ‘Jack’, which possesses a high ability to undergo somatic embryo induction and regeneration from immature cotyledons (Tomlin et al. 2002), was used for transformation. Biolistic

transformation was performed according to El-Shemy et al. (2004). All tissue culture was conducted under 16 h light: 8 h dark (photosynthetic photon flux density: 20–50 μmol m^{-2} s^{-1}) at 26°C. Transgenic plants were grown in commercial soil (Katakura Chikkarin Co., Tokyo, Japan) at 25°C in an isolated greenhouse for transgenic plants at Hokkaido University. Transgenic embryo lines were selected based on their viability in liquid medium containing 15 mg/L of hygromycin 3 months after bombardment.

Cleaved amplified polymorphic sequence (CAPS) analysis

To extract leaf genomic DNA, leaf pieces (approximately 5 mm × 5 mm) or parts of trasngeic embryos (approximately 2 mm × 2 mm) were homogenized in 200 μL of extraction buffer A [2% (w/v) CTAB (hexadecyltrimethyl-ammonium bromide), 100 mM Tris-HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl, and 0.07% 2-mercaptethanol] in a BioMasher II tube (Nippi, Tokyo, Japan). To extract genomic DNA from mature seeds, a segment of cotyledon was powdered and approximately 5 mg of powder was stirred in extraction buffer B [10 mM Tris-HCl (pH 8.0), 5 mM EDTA, 0.5% SDS, 0.5% NP-40, 0.5% Tween 20, and 80 mg/L proteinase-K (Wako, Osaka, Japan)]. The mixture was incubated at 50°C for 1 h. DNA extracts were deproteinized with a mixture of phenol, chloroform, and isoamyl alcohol [25:24:1 (v/v/v)]. DNA was precipitated from the supernatant with 2-propanol. The targeted regions in the *Gly m Bd 30K* gene were amplified by PCR with specific primers (5′-GCAAGCTCCCAAGGATGTG-3′ and 5′-ACGCCCAACCGCTTCCTAT-3′) using 30 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 60 s in a thermal cycler T100 (BioRad, Hercules, CA, USA). The amplified products were digested with BsaJI restriction enzyme (NEB) and separated by electrophoresis in 2% agarose gels. The DNA fragments with the digestion pattern expected if the target region did or did not carry mutations were considered as mutant type and wild type, respectively. DNA fragments of unexpected size were also regarded as mutant type.

DNA sequencing

The targeted and flanking regions of *Gly m Bd 30K* were amplified with specific primers (Table S1). The amplified products were cloned into the pGEM-T-Easy vector (Promega, Madison, WI, USA), and then sequenced with the Big Dye terminator cycle method by using an ABI3100 or ABI3130 Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing was performed by the Instrumental Analysis Division, Graduate School of Agriculture, Hokkaido University.

Evaluation of off-target effect of gRNA designed in this study

Genome DNA of five embyogenic lines and ‘Jack’ was used to evaluated off-target effect of gRNA used in this study. Off-target sites were estimated by *in silico* (CRISPR-P 2.0) analysis. PCR analysis was performed using primers specific for the targeted gene (Table S1). The amplified products were

sequenced and evaluated off-target effect of the gRNA.

Selection of *Cas9*-free plants

To confirm the integration of the Cas9 and gRNA expression module in T_0–T_2 generations, PCR analysis was performed using primers specific for the *Cas9* gene (Table S1). PCR was also performed to simultaneously amplify endogenous Glyma.01G214600 as a positive control. The PCR was performed using 30 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 30 s. The presence of the *Cas9* gene was detected by the PCR amplification of a product of the expected size.

Protein analyses of mature seeds

Soybean meal was collected from mature seeds of mutants and 'Jack'. The extraction of crude protein and the separation of proteins were performed as described by Yamada et al. (2014). Crude proteins were separated by SDS-PAGE in a precast 5%–12% gradient gel (ATTO, Tokyo, Japan), and then transferred onto a PVDF membrane (Hybond-P; GE Healthcare, Little Chalfont, UK). Membranes were blocked with 5% skim milk (Wako, Osaka, Japan) overnight at 4°C. Recombinant Gly m Bd 30K protein was produced using the baculovirus expression system as described in Maruyama et al. (2018). Antisera were raised in rabbits against the recombinant protein as described in Nishizawa et al. (2003). Western blot was performed with the antisera and the ECL Plus Western Blotting system (GE Healthcare).

Expression analysis by semi-quantitative RT-PCR

Total RNA was extracted from mature T_3 seeds of mutants and 'Jack' by the LiCl precipitation procedure (Dwiyanti et al. 2011) with several modifications described below. Frozen matured seeds were disrupted with a Multi-Beads Shocker (Yasui Kikai Co., Osaka, Japan). The fine powder was mixed with 600 µL of RNA extraction buffer [10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 100 mM NaCl, 1% SDS) and 300 µL of Tris-saturated phenol (pH 8.0). Then 300 µL of chloroform:isoamyl alcohol [24:1 (v/v]) was added to the sample, the solution was mixed vigorously, and the aqueous and organic layers were separated by centrifugation at 20,000 ×*g* for 10 min. RNA was precipitated by the addition of 0.3 volumes of 8 M LiCl. Removal of DNA by DNase I and cDNA synthesis from purified total RNA were performed as described by Dwiyanti et al. (2011). Semi-quantitative RT-PCR was conducted in a 20-µL volume using 30 cycles of 94°C for 30 s, 57°C for 30 s, and 72°C for 10 s. The transcript level of the *Gly m Bd 30K* gene was evaluated relative to that of the 18S rRNA gene (XR_003264275).

Elemental analysis

Mature seeds of mutants and 'Jack' were kept for more than 10 days in a desiccator to remove extra

moisture. This analysis was provided with three biological replicates. Seed coats were removed by a surgical knife and only cotyledons were analyzed. A total of five seeds were disrupted with a Multi-Beads Shocker (Yasui Kikai) to produce soybean meal. Contents of carbon, hydrogen, and nitrogen in soybean meal were measured with a Vario ELIII elemental analyzer (Elementar, Hesse, Germany).

Statistical analysis

Pearson's chi-square test was performed to evaluate the segregation pattern of transgene in the T_1 generation. The Welch's *t* test were used to examine whether mutant line means are significantly from control plant in carbon, hydrogen, and nitrogen content of mature seeds. All statistic data was evaluated significantly different from the control when *P* values were < 0.05.

Results

Generation of transgenic embryo lines harboring the CRISPR/Cas9 expression module

To conduct site-directed mutagenesis of soybean with the CRISPR/Cas9 system, we used biolistic transformation of soybean somatic embryos with p30K-hyg (Fig. 1). Five embryogenic lines (30K-1 to -5) were selected 3 months after bombardment. Genomic DNA was extracted from these embryogenic lines. Integration of the CRISPR/Cas9 expression module was confirmed by PCR analysis with *Cas9*-specific primers in all five embryogenic lines (Fig. 2a). Subsequently, induction of mutation in the *Gly m Bd 30K* locus was evaluated by cleaved amplified polymorphic sequence (CAPS) analysis. The DNA fragments were divided into wild-type (386 bp) and mutant-type (approximate 510 bp) based on expected sizes; fragments of unexpected size were also detected and considered to be mutant-type (Fig. 2b). All the embryo lines showed mutant-type fragments (Fig. 2b). Fragments of unexpected size were detected in the 30K-1 and 30K-3 lines (Fig. 2b). Both wild-type and mutant-type fragments were detected in the 30K-5 line (Fig. 2b).

Screening for mutations in the targeted locus in transgenic embryo lines

PCR products amplified using *Gly m Bd 30K*–specific primers were cloned and sequenced. Mutations in the targeted locus were detected in all five transgenic embryo lines (Fig. 3). Insertions of 600 and 133 nucleotides, whose sequences corresponded to the vector frame of p30K-hyg, were detected in 30K-1 and 30K-3 lines, respectively (Fig. 3). Deletions were the other predominant mutation type in the targeted locus (Fig. 3). Only two mutant variants were observed per transgenic embryo line (Fig. 3). Although the CAPS analysis profile for the 30K-5 line contained a fragment whose size corresponded to the wild-type fragment (Fig. 2b), no wild-type allele was detected by sequencing (Fig. 3); rather, the sequencing analysis revealed that a single-nucleotide deletion generated a de novo recognition site of the BsaJI restriction enzyme (Fig. S1). These results suggest that all the embryo lines contained biallelic mutations and none were chimeric in the target region.

Off-target analysis in transgenic embryo lines

Off-target sites were estimated by *in silico* analysis. Three sites with high DNA identity to the gRNA sequence were found in the 3′-UTR of Glyma.06G082000, an intron of Glyma.14G205500, and the 5′-UTR of Glyma.18G050700. These off-target sites contained 4-nucleotide mismatches against the gRNA designed in this study (Fig. S2). Sequencing analysis revealed that sequences of the three off-target sites in all five transgenic embryo lines completely matched the wild-type sequence in 'Jack' (Fig. S2), indicating that no mutations were induced at these off-target sites.

Regeneration from transgenic embryo lines

All five transgenic embryo lines possessed proliferative activity even 7 months after bombardment (Fig. S3a, b). Mature embryos were produced from each transgenic embryo line (Fig. S3c, d). To stimulate shoot regeneration and rooting, these mature embryos were transferred to rooting medium. Mature embryos of two transgenic embryo lines (30K-3 and -5) developed into a total of 10 and 20 young plantlets (T_0 plants), respectively (Fig. S3e); no T_0 plants were obtained from the other transgenic embryo lines. The T_0 plants were grown (Fig. S3f). Those generated from the same embryogenic line were considered as clonal plants.

Transmission of mutations and the exogenous gene to the T_1 generation

T_1 seeds were collected from T_0 plants derived from 30K-3 and 30K-5 transgenic embryo lines. Genomic DNA extracted from cotyledons of T_1 seeds was used to evaluate the inheritability of mutations in the targeted locus in T_1 progenies. The 30K-3 line possessed a 133-nucleotide insertion (*i133*) and 7-nucleotide deletion (*d7*) in the targeted locus (Fig. 3). The insertion could be easily evaluated without CAPS analysis by conducting PCR analysis with *Gly m Bd 30K*–specific primers (Fig. S4a). Sequencing analysis confirmed that the amplified product that corresponded in size to that of 'Jack' in the T_1 progenies of the 30K-3 line actually contained the *d7* allele (Table S2). Therefore, 30K-3 progenies were genotyped by PCR alone in the rest of the study. On the other hand, the 30K-5 line possessed mutant alleles of a single-nucleotide deletion (*d1*) and a 9-nucleotide deletion (*d9*) in the targeted locus (Fig. 3). The *d9* allele could be evaluated as a mutant-type fragment by CAPS analysis (Fig. S4b). As mentioned in a previous subsection, the restriction fragment corresponding in size to that of 'Jack' in T_1 progenies of a 30K-5 T_0 plant was confirmed to be the *d1* allele by sequencing analysis (Table S3). Therefore, genotyping of 30K-5 progenies was performed by CAPS analysis only in the rest of the study.

Transmission of the *Cas9* gene to the T_1 generation was evaluated by PCR analysis using *Cas9*-specific primers (Fig. S5). A single T_0 plant derived from the 30K-3 line produced 78 *Cas9*-positive and 28 *Cas9*-negative seeds. The segregation ratio was compatible with 3:1 ($p = 0.74$). In contrast, a

single T_0 plant derived from the 30K-5 line produced 59 *Cas9*-positive and 35 *Cas9*-negative seeds, which is not compatible with 3:1 segregation ($p = 0.0062$; Table 1), indicating a bias in segregation. These results show that the frequency of *Cas9*-free seeds is high.

Selection of homozygous mutant alleles

Cas9-free plants with homozygous mutant alleles in the targeted locus were selected among the T_2 progenies through PCR or CAPS analyses. Genotypes containing homologous mutant alleles of the *d1*, *d7* and *d9*, and *i133* were named as *del-1*, *del-7*, *del-9*, and *ins-133*, respectively.

Level of Gly m Bd 30K protein in mutant mature seeds

In SDS-PAGE analysis, the accumulation of putative wild-type protein (34 kDa) was detected in the crude proteins of 'Jack' mature seeds, but not mature seeds with a mutant genotype (*del-1*, *del-7*, *del-9*, or *ins-133*) (Fig. 4a). SDS-PAGE analysis also showed no difference in the composition of the main storage proteins, such as 7S and 11S globulins, between the mutant and 'Jack' mature seeds (Fig. 4a). To confirm the absence of Gly m Bd 30K in the mutants, we performed immunoblot analysis using polyclonal antibody specific to this protein. Although a strong signal band corresponding to Gly m Bd 30K was obtained in 'Jack', all four mutants showed no accumulation of Gly m Bd 30K protein (Fig. 4b). Weak signal bands of unexpected size detected in the *del-9* mutant and 'Jack' were considered as non-specific products (Fig. 4b).

Expression level of the *Gly m Bd 30K* gene

There was a possibility that smaller proteins than that of 'Jack' are detected in mutants, because the polyclonal antibody against Gly m Bd 30K protein was used in the immunoblot analysis. No accumulation of these proteins was detected in all mutants (Fig. 4b). Therefore, we also examined the expression level of the targeted gene. To evaluate the expression level of the *Gly m Bd 30K* gene, we extracted total RNA from the mature T_3 seeds of the four mutants and 'Jack', and conducted semi-quantitative RT-PCR analysis of the region up-stream of the mutation site (Fig. S6). Although amplified products of the 18S ribosomal RNA were detected at similar levels in all mature seeds of the mutants and 'Jack' (Fig. 5a), three mutant genotypes (*del-1*, *del-7*, and *ins-133*) showed no expression of the *Gly m Bd 30K* gene (Fig. 5b). On the other hand, *Gly m Bd 30K* gene expression was detected in the *del-9* genotype (Fig 5b). Based on the expression level of the 18S ribosomal RNA, the expression level of the *Gly m Bd30K* gene in the *del-9* genotype was similar to that of 'Jack'.

Morphological characteristics and seed components of mutants

To assess the consequences of the site-directed mutagenesis in the targeted loci or the affection of somaclonal variation induced through tissue culture, we examined the morphological characteristics

of the plant body and seeds in the four mutants and 'Jack' and detected no morphological differences (Figs. 6 and 7). In addition, the contents of nitrogen, carbon, and hydrogen in the seeds were also evaluated. No significant difference was detected in these components between each mutant and 'Jack' (Fig. S7).

Discussion

Here, we used biolistic transformation and the CRISPR/Cas9 system to conduct site-directed mutagenesis of a target gene, *Gly m Bd 30K*, in soybean embryos. After the resultant transgenic embryo lines were cultured for 6 months, all lines contained biallelic mutations and no wild-type allele in the targeted locus (Fig. 3). If mutagenesis occurs during division and proliferation of a transformed cell, various mutant variants in the targeted locus would be expected to be detected in transgenic embryo lines. Therefore, our finding of two mutant variants per line indicates that mutagenesis in the targeted locus likely occurred in a single-transgenic cell immediately after the exogenous gene was delivered into the embryo. Our finding and those of studies that delivered preassembled CRISPR/Cas9 ribonucleoprotein complexes into the protoplasts of various plant species (Woo et al. 2015, Murovec et al. 2018, Liu et al. 2020) suggest that it is possible to induce mutations in a targeted locus in a cell harboring the CRISPR/Cas9 module before division of that cell. In contrast to our previous study of site-directed mutagenesis using *Agrobacterium*-mediated transformation and the CRISPR/Cas9 system in which most mutations induced in the targeted locus, even in the T_1 generation, were heterozygous and/or chimeric in nature (Kanazashi et al. 2018), in the current study all the transgenic embryo lines certainty transmitted the mutant alleles found in the T_0 generation to the T_1 generation (Tables S2, S3; Fig. S4). These transformation systems differ in the starting status of tissue culture whether it undergoes embryogenesis or organogenesis. Mutagenesis immediately after introgression of the CRISPR/Cas9 expression module might be depended on these differences.

The opportunity to remove transgenes from mutants by genetic segregation strongly depends on the number of exogenous genes integrated into the soybean genome. If this number is small, the transgenes are easily removed in the next generation by genetic segregation. Although biolistic transformation in soybean tends to insert multiple copies of exogenous gene in the host genome (Reddy et al. 2003), our previous study demonstrated that a low copy number of gene integration can be achieved in transgenic soybean plants generated by biolistic transformation (Ishimoto et al. 2010). Only two among five transgenic embryo lines generated plantlets from their mature embryos (Fig. S3). The integration ratio of transgene in the T_1 progenies showed a simple Mendelian inheritance manner, although the number of tandem insertion of the CRISPR/Cas9 expression module was not evaluated in this study. The segregation ratio of presence of transgenes in the T_1 progenies of a 30K-3 T_0 plants was compatible with 3:1 segregation ($p = 0.74$; Table 1), and approximately 37% of the T_1 progenies of a 30K-5 T_0 plant were *Cas9* free (Table 1). These results indicate that both 30K-3 and 30K-5 transgenic embryo

lines likely possessed a single transgene copy or single locus of multiple copies. The transformation conditions such as concentration of gold particles and coated DNA, and bombardment distance might be suitable to the integration of single copy of transgenes. On the other hand, it is possible that the integration of the large copy number of transgenes might also have prevented plant regeneration from the other three transgenic embryo lines (30K-1, -2, and -4). Generating of many transgenic embryo lines and analyzing them might help understanding these differences.

Unintentional DNA insertions of 600 bp and 133 bp were detected in the targeted locus of 30K-1 and 30K-3 transgenic embryo lines, respectively (Fig. 3). These inserted DNAs, which originated from the expression vector frame used in this study, were precisely inserted at a predicted cleavage site by the Cas9 enzyme (Fig. 3), suggesting that foreign DNAs might be integrated into the targeted locus through non-homologous end joining (NHEJ). Insertion of foreign DNAs at a targeted site by the CRISPR/Cas9 system is generally mediated by homology-directed repair (HDR), which requires that these DNAs have desirable specific DNA sequence and homologous sequence identical to the flanking region of the targeted site (Li et al. 2015, Zhao et al. 2016, Butt et al. 2017, Hahn et al. 2018). It was not possible to explain the potential as a microhomology provider of the protospacer region, when the protospacer sequence was compared with the end and its adjacent sequences of DNA fragments inserted into the cleaved sites. Bonawitz et al. (2019) demonstrated that the transgenes were introduced into the targeted site of soybean genome by NHEJ when a zinc finger nuclease–mediated system was used for genome editing. Although there are few reports of the insertion of DNA fragment mediated by NHEJ in higher plants, these results indicate that the double strand break by the Cas9 enzyme allows intentional delivery or replacement of desirable DNA fragments in soybean plants. Delivery or replacement of desirable DNA fragments mediated by NHEJ might become a major tool for molecular breeding of crops and vegetables if biolistic transformation is optimized for various plant species.

Immunoblot analysis revealed that no mutant seeds accumulated wild-type Gly m Bd 30K protein (Fig. 5). For genotype *del-9*, the mutant *Gly m Bd 30K* allele had an in-frame mutation and semi-quantitative RT-PCR analysis of *Gly m Bd 30K* exon 1 showed an expression level similar to that in wild-type seeds (Fig. 5, Fig. S6). However, the immunoblot analysis in genotype *del-9* showed no accumulation of protein reacted with polyclonal antibody against Gly m Bd 30K protein (Fig. 4b). Further study might resolve these contradictions.

Five immunodominant epitopes have been reported for the Gly m Bd 30K protein (Helm et al. 1998, Helm et al. 2000). If mutant Gly m Bd 30K proteins are produced from *del-1*, *del-7*, *del-9*, and *ins-133* genotypes, they would be expected to contain some or all of the epitopes. Furthermore, a mutation induced in the *Gly m Bd 30K* locus might produce a de novo allergen. Therefore, the expression level of mutant alleles is important for developing mutants with the hypoallergenic trait. Frame-shift mutations in the targeted locus in the *del-1*, *del-7*, and *ins-133* mutants prevented not only the accumulation of the Gly m Bd 30K protein but also the expression of the *Gly m Bd 30K* gene (Figs.

4b and 5). These findings suggest that the frame-shift mutations produce aberrant mRNAs from the targeted locus, which induced nonsense mRNA decay (NMD), like in a site-directed mutagenesis study conducted in *Brassica carinata* using the hairy root transformation system (Kirchner et al. 2017). Consequently, the *del-1*, *del-7*, and *ins-133* mutants generated in this study are suitable as low-allergen lines for Gly m Bd 30K.

In this study, we demonstrated the efficient induction of inheritable mutations in T_0 plants enabled us to remove the foreign DNAs easily by genetic segregation in the T_1 generation. However, the high embryogenesis ability which is essential for the success of the biolistic transformation is limited among commercial soybean varieties. Song et al. (2010) have reported a highly effective QTL underlying somatic embryogenesis capacity in soybean. Identification of the responsible gene for this QTL and selection of the appropriate genotypes might lead to an efficient site-directed mutagenesis using the biolistic transformation system in many commercial soybean varieties.

Abbreviations

EDTA: ethylenediaminetetraacetic acid; SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis; PVDF: polyvinylidene difluoride; UTR: untranslated region.

Acknowledgements

We thank Professor H. Puchta (Karlsruhe Institute of Technology) for permission to use plasmid DNA of pEn-Chimera and pDe-CAS9, and M. Suzuki, S. Noguchi, and Y. Kitsui (Hokkaido University), and J. Kamiya (National Agricultural and Food Research Organization) for technical assistance of plant transformation and tissue culture.

Declarations

Funding

This work was supported by funds from the Cabinet Office, Government of Japan [the Cross-ministerial Strategic Innovation Promotion Program (SIP)] for TY.

Conflicts of interest/Competing interests

The authors declare that they have no competing interests.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The dataset supporting this study is included within the manuscript and its additional files. The expression vector p30K-hyg developed in this study are available from the corresponding author on reasonable request.

Code availability

Not applicable.

Authors' contributions

KA, JA, and TY conceived and designed the experiments; AH, YK, MM, SH, and ME constructed vectors; KA, AH, YK, and MI performed the site-directed mutagenesis experiment; KA, MH, TH, and NM performed the mutational analyses in transgenic soybean; KA, JA, and TY contributed to the writing of the manuscript. All authors have read and approved the manuscript.

References

Aragão FJL, Sarokin L, Vianna GR, Rech EL (2000) Selection of transgenic meristematic cells utilizing a herbicidal molecule results in the recovery of fertile transgenic soybean *Glycine max* (L.) Merril plants at a high frequency. Theor Appl Genet 101:1-6

Bonawitz ND, Ainley WM, Itaya A, Chennareddy SR, Cicak T, Effinger K, Jiang K, Mall TK, Marri PR, Samue JP, Sardesai N, Simpson M, Folkerts O, Sarria R, Webb SR, Gonzalez DO, Simmonds DH, Pareddy DR (2019) Zinc finger nuclease-mediated targeting of multiple transgenes to an endogenous soybean genomic locus via non-homologous end joining. Plant Biotechnol J 17:750-761

Butt H, Eid A, Ali Z, Atia MAM, Mokhtar MM, Hassan N, Lee CM, Bao G, Mahfouz MM (2017) Efficient CRISPR/Cas9-mediated genome editing using a chimeric single-guide RNA molecule. Frontiers in Plant Science 8:1441

Cai Y, Chen L, Liu X, Guo C, Sun S, Wu C, Jiang B, Han T, Hou W (2018) CRISPR/Cas9-mediated targeted mutagenesis of *GmFT2a* delays flowering time in soya bean. Plant Biotechnol J 16:176-185

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-

based constructs for DNA targeting. Nucleic Acids Res 39:82

Curtin SJ, Xiong Y, Michno J-M, Campbell BW, Stec AO, Cermak T, Starker C, Voytas DF, Eamens AL, Stupar RM (2018) CRISPR/Cas9 and TALENs generate heritable mutations for genes involved in small RNA processing of *Glycine max* and *Medicago truncatula*. Plant Biotechnol J 16:1125-1137

Dang W, Wei ZM (2007) An optimized *Agrobacterium*-mediated transformation for soybean for expression of binary insect resistance genes. Plant Sci 173:381-389

Do PT, Nguyen CX, Bui HT, Tran LTN, Stacey G, Gillman JD, Zhang ZYJ, Stacey MG (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

Dwiyanti MS, Yamada T, Sato M, Abe J, Kitamura K (2011) Genetic variation of γ-tocopherol methyltransferase gene contributes to elevated α-tocopherol content in soybean seeds. BMC Plant Biol 11:152

El-Shemy HA, Teraishi M, Khalafalla MM, Katsube-Tanaka T, Utsumi S, Ishimoto M (2004) Isolation of soybean plants with stable transgene expression by visual selection based on green fluorescent protein. Mol Breed 14:227-238

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

Finer JJ, Vain P, Jones MW, McMullen MD (1992) Development of the particle inflow gun for DNA delivery to plant-cells. Plant Cell Rep 11:323-328

Hahn F, Eisenhut M, Mantegazza O, Weber APM (2018) Homology-directed repair of a defective *Glabrous* gene in *Arabidopsis* with Cas9-based gene targeting. Front Plant Sci 9:424

Haun W, Coffman A, Clasen BM, Demorest ZL, Lowy A, Ray E, Retterath A, Stoddard T, Juillerat A, Cedrone F, Mathis L, Voytas DF, Zhang F (2014) Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnol J 12:934-940

Helm RM, Cockrell G, Connaughton C, West CM, Herman E, Sampson HA, Bannon GA, Burks AW (2000) Mutational analysis of the IgE-binding epitopes of P34/Gly m Bd 30K. J Allergy Clinic Immun 105:378-384

Helm RM, Cockrell G, Herman E, Burks AW, Sampson HA, Bannon GA (1998) Cellular and molecular characterization of a major soybean allergen. Int Arch Allergy Imm 117:29-37

Hinchee MAW, Connorward DV, Newell CA, McDonnell RE, Sato SJ, Gasser CS, Fischhoff DA, Re DB, Fraley RT, Horsch RB (1988) Production of transgenic soybean plants using *Agrobacterium*-mediated DNA transfer. Bio-Technol 6:915-921

Ishimoto M, Rahman SM, Hanafy MS, Khalafalla MM, El-Shemy HA, Nakamoto Y, Kita Y, Takanashi K, Matsuda F, Murano Y, Funabashi T, Miyagawa H, Wakasa K (2010) Evaluation of

amino acid content and nutritional quality of transgenic soybean seeds with high-level tryptophan accumulation. Mol Breed 25:313-326

Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA (2015) Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnol 15:16

Joseph LM, Hymowitz T, Schmidt MA, Herman EM (2006) Evaluation of *Glycine* germplasm for nulls of the immunodominant allergen P34/Gly m Bd 30k. Crop Sci 46:1755-1763

Kanazashi Y, Hirose A, Takahashi I, Mikami M, Endo M, Hirose S, Toki S, Kaga A, Naito K, Ishimoto M, Abe J, Yamada T (2018) Simultaneous site-directed mutagenesis of duplicated loci in soybean using a single guide RNA. Plant Cell Rep 37:553-563

Kirchner TW, Niehaus M, Debener T, Schenk MK, Herde M (2017) Efficient generation of mutations mediated by CRISPR/Cas9 in the hairy root transformation system of *Brassica carinata*. PLoS One 12:e0185429

Li JF, Norville JE, Aach J, McCormack M, Zhang DD, Bush J, Church GM, Sheen J (2013) Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. Nature Biotechnol 31:688-691

Li ZS, Liu ZB, Xing AQ, Moon BP, Koellhoffer JP, Huang LX, Ward RT, Clifton E, Falco SC, Cigan AM (2015) Cas9-guide RNA directed genome editing in soybean. Plant Physiol 169:960-970

Liu KS (1997) Soybeans: Chemistry, technology, and utilization. Chapman & Hall, USA, pp 36-113

Liu SJ, Wei ZM, Huang JQ (2008) The effect of co-cultivation and selection parameters on *Agrobacterium*-mediated transformation of Chinese soybean varieties. Plant Cell Rep 27:489-498

Liu WS, Rudis MR, Cheplick MH, Millwood RJ, Yang JP, Ondzighi-Assoume CA, Montgomery GA, Burris KP, Mazarei M, Chesnut JD, Stewart CN (2020) Lipofection-mediated genome editing using DNA-free delivery of the Cas9/gRNA ribonucleoprotein into plant cells. Plant Cell Rep 39:245-257

Maruyama N, Sato S, Cabanos C, Tanaka A, Ito K, Ebisawa M (2018) Gly m 5/Gly m 8 fusion component as a potential novel candidate molecule for diagnosing soya bean allergy in Japanese children. Clin Exp Allergy 48:1726-1734

McCabe DE, Swain WF, Martinell BJ, Christou P (1988) Stable transformation of soybean (*Glycine-max*) by particle-acceleration. Bio-Technol 6:923-926

Mikami M, Toki S, Endo M (2015) Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. Plant Mol Biol 88:561-572

Murovec J, Gucek K, Bohanec B, Avbelj M, Jerala R (2018) DNA-free genome editing of *Brassica oleracea* and *B. rapa* protoplasts using CRISPR-Cas9 ribonucleoprotein complexes. Front Plant Sci 9:1594

Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. Nature Biotechnol

31:691-693

Nishizawa K, Maruyama N, Satoh R, Fuchikami Y, Higasa T, Utsumi S (2003) A C-terminal sequence of soybean β-conglycinin α' subunit acts as a vacuolar sorting determinant in seed cells. Plant J 34:647-659

Ogawa T, Tsuji H, Bando N, Kitamura K, Zhu YL, Hirano H, Nishikawa K (1993) Identification of the soybean allergenic protein, Gly-m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. Biosci Biotechnol Bioch 57:1030-1033

Olhoft PM, Somers DA (2001) L-cysteine increases *Agrobacterium*-mediated T-DNA delivery into soybean cotyledonary-node cells. Plant Cell Rep 20:706-711

Parrott WA, Hoffman LM, Hildebrand DF, Williams EG, Collins GB (1989) Recovery of primary transformants of soybean. Plant Cell Rep 7:615-617

Paz MM, Martinez JC, Kalvig AB, Fonger TM, Wang K (2006) Improved cotyledonary node method using an alternative explant derived from mature seed for efficient *Agrobacterium*-mediated soybean transformation. Plant Cell Rep 25:206-213

Reddy MSS, Dinkins RD, Collins GB (2003) Gene silencing in transgenic soybean plants transformed via particle bombardment. Plant Cell Rep 21:676-683

Sato H, Yamada T, Kita Y, Ishimoto M, Kitamura K (2007) Production of transgenic plants and their early seed set in Japanese soybean variety, Kariyutaka. Plant Biotechnol 24:533-536

Schmidt MA, Hymowitz T, Herman EM (2015) Breeding and characterization of soybean Triple Null; a stack of recessive alleles of Kunitz Trypsin Inhibitor, Soybean Agglutinin, and P34 allergen nulls. Plant Breed 134:310-315

Shan QW, Wang YP, Li J, Zhang Y, Chen KL, Liang Z, Zhang K, Liu JX, Xi JJ, Qiu JL, Gao CX (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. Nature Biotechnol 31:686-688

Song XH, Han YP, Teng WL, Sun GL, Li WB (2010) Identification of QTL underlying somatic embryogenesis capacity of immature embryos in soybean (*Glycine max* (L.) Merr.). Plant Cell Rep 29:125-131

Tomlin ES, Branch SR, Chamberlain D, Gabe H, Wright MS, Stewart CN (2002) Screening of soybean, *Glycine max* (L.) Merrill, lines for somatic embryo induction and maturation capability from immature cotyledons. In Vitro Cell Dev -Pl 38:543-548

Tsuji H, Bando N, Hiemori M, Yamanishi R, Kimoto M, Nishikawa K, Ogawa T (1997) Purification and characterization of soybean allergen Gly m Bd 28K. Biosci Biotechnol Bioch 61:942-947

Yamada T, Mori Y, Yasue K, Maruyama N, Kitamura K, Abe J (2014) Knockdown of the 7S globulin subunits shifts distribution of nitrogen sources to the residual protein fraction in transgenic soybean seeds. Plant Cell Rep 33:1963-1976

Yamada T, Takagi K, Ishimoto M (2011) Recent advances in soybean transformation and their

application to molecular breeding and genomic analysis. Breed Sci 61:480-494

Woo JW, Kim J, Il Kwon S, Corvalan C, Cho SW, Kim H, Kim SG, Kim ST, Choe S, Kim JS (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nature Biotechnol 33:1162-1164

Zhao Y, Zhang C, Liu W, Gao W, Liu C, Song G, Li W-X, Mao L, Chen B, Xu Y, Li X, Xie C (2016) An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. Scientific Reports 6:23890.

Figure legends

Fig. 1 Schematic diagram of simultaneous site-directed mutagenesis using the CRISPR/Cas9 system in soybean. The upper panel denotes the exon–intron structure of the *Gly m Bd 30K* (Glyma.08G116300.1) gene. Boxes and pentagons, exons; bold lines, introns. Pink- and red-colored parts indicate untranslated and translated regions, respectively. Gray arrows, primer regions for CAPS analysis; yellow bar, position recognized by the gRNA. The middle panel shows sequences of the gRNA and its linkers. Red sequence, the gRNA designed in this study. The bottom panel denotes the structure of the CRISPR/Cas9 expression module for soybean transformation. *Hyg* cassette, marker gene unit for hygromycin selection; *AtHSP* ter, terminator of a heat shock protein gene from Arabidopsis; *Cas9*, *Cas9* from *Streptococcus pyogenes* (codon-optimized for Arabidopsis); *GmU6* pro, soybean U6 promoter; LB, left border; *PcUbi* pro, ubiquitin promoter from parsley (*Petroselinum crispum*); RB, right border.

Fig. 2 Detection of transgenes and mutations in the targeted locus in somatic embryo lines. (a) Products amplified using *Cas9*-specific primers in the somatic embryo lines. Black arrow, size expected for amplified products of the *Cas9* gene; gray arrow, size expected for amplified products of *GmSGR1* gene (control endogenous gene). (b) CAPS analysis of the targeted locus in somatic embryo lines. Black dotted arrow, size expected for mutant-type fragment (approximate 510 bp); gray dotted arrow, size expected for wild-type fragment (386 bp); black triangles, fragments of unexpected size (considered as mutant type). Jack, control embryo; p30K-hyg, expression vector used in this study; M, molecular weight marker (100-bp ladder).

Fig. 3 Mutant variants of the targeted locus in transgenic embryo lines. Red sequences, sites targeted by the gRNA designed in this study; underlined blue sequences, the proto-spacer adjacent motif (PAM) region; green numbers, numbers of nucleotides inserted into the targeted locus; black numbers in parentheses, numbers of clones in the sequencing analysis. The names of the mutant alleles are given

to the immediate right of the sequences; e.g. *d1*, a single-nucleotide deletion; *i133*, 133-nucleotide insertion; Jack, wild-type reference sequence.

Fig. 4 SDS-PAGE and immunoblot analyses of the crude proteins of mature seeds from the four mutants (*del-7*, *ins-133*, *del-1*, and *del-9*) and 'Jack'. (a) Proteins separated by SDS-PAGE and stained with Coomassie Brilliant Blue. Black triangle, putative band of Gly m Bd 30K protein; numbers at the left side indicate molecular weight. (b) Immunoblot analysis using polyclonal antibody against Gly m Bd 30K protein.

Fig. 5 Semi-quantitative RT-PCR of the *Gly m Bd 30K* gene in mature seeds of the four mutants (*del-7*, *ins-133*, *del-1*, and *del-9*) and 'Jack'. (a) Products amplified using primers specific for *18S rRNA* (control endogenous gene). (b) Products amplified using *Gly m Bd 30K*–specific primers. Jack, control plant; Water, negative control.

Fig. 6 Morphological characteristics of the four mutants (*del-7*, *ins-133*, *del-1*, and *del-9*) and 'Jack' plants. (a) Plant body. Scale bar, 10 cm. (b) Mature seeds. Scale bars, 1 cm.