1 **Efficient, selectable marker free gene targeting in soybean**
2 **using novel *Ochrobactrum haywardense*-mediated delivery**

3

4 Sandeep Kumar*, Zhan-Bin Liu, Nathalie Sanyour-Doyel, Brian Lenderts, Andrew

5 Worden, Ajith Anand, Hyeon-Je Cho, Joy Bolar, Charlotte Harris, Lingxia Huang,

6 Aiqiu Xing, Alexandra Richardson

7

8 Corteva Agriscience™, Johnston, IA 50131, USA

9 *Corresponding author sandeep.kumar@corteva.com

10

Abstract

We report robust selectable marker-free gene targeting (GT) system in soybean, one of the most economically important crops. A novel efficient *Ochrobactrum haywardense*-mediated embryonic axis transformation method was used for the delivery of CRISPR-Cas9 components and donor template to regenerate T0 plants in 6-8 weeks after transformation. This approach generated up to 3.4% targeted insertion of the donor sequence into the target locus in T0 plants, with ~ 90% mutation rate observed at the genomic target site. The GT was demonstrated in two genomic sites using two different donor DNA templates without a need of a selectable marker within the template. High-resolution Southern by Sequencing (SbS) analysis identified T1 plants with precise targeted insertion and without unintended plasmid DNA. Unlike previous low-frequency GT reports in soybean that involved particle bombardment-mediated delivery and extensive selection, the method described here is fast, efficient, reproducible, does not require selectable marker within the donor DNA, and generates non-chimeric plants with heritable GT.

Introduction

Soybean (*Glycine max*) has become one of the most economically important legume seed crops. In addition to being a major oilseed crop, it also provides more than a quarter of the total protein for the world's food and animal feed (Graham and Vance, 2003). Due to population growth along with increased societal interest in plant-based protein diets, the demand for soybean is gradually expanding (Ray et al., 2013). However, unlike considerable progress made in increasing the yields of rice, wheat, and maize through the Green Revolution, only limited improvements have

been made for soybean(Liu et al., 2020). The yield and other trait improvement in soybean will require complex and precise genetic manipulations to obtain desired plant height, node number, internode length, branch number, disease resistance, and seed size supplemented with high oil and protein contents. Conventional plant breeding has played a major role in making genetic improvement of soybean and will continue to do so in the future. However, genome editing technology could expedite and revolutionize traditional breeding processes and enable markedly improved precision in soybean breeding.

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9(Jinek et al., 2012) is a robust and versatile genome editing tool to make targeted DNA double-strand breaks (DSBs)(Kim et al., 1996; Epinat et al., 2003; Christian et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012). Genomic DSBs in somatic cells are repaired predominantly by non-homologous end joining (NHEJ), which is an error-prone repair pathway often resulting in small insertions or deletions (indels) at the DSB site. Homologous recombination (HR)(Wood, 1996), in contrast, is a precise repair pathway that requires a DNA repair template with flanking sequences homologous to those flanking the genomic DSB. Unlike NHEJ, HR is a very low frequency process in somatic cells used for plant transformation. Consequently, NHEJ-mediated gene mutations, which can be recovered in 30-100% of regenerated plants, have become routine(Puchta, 2017), whereas precise gene targeting (GT) via HR remains inefficient and challenging. Some monogenic simple traits that depend on endogenous gene mutation can be generated through NHEJ-mediated DNA repair (Li et al., 2016; Li et al., 2016; Liu et al., 2017; Nonaka et al., 2017; Li et al., 2018). Complex yield traits require precise GT via HR to make large modifications in the genome to expedite the breeding process(Shi et al., 2017; Yu et al., 2017).

The majority of genome editing work reported in soybean describes successful NHEJ-mediated targeted mutagenesis (Al Amin et al., 2019; Bao et al., 2019; Campbell et al., 2019; Cheng et al., 2019; Di et al., 2019; Do et al., 2019; Han et al., 2019; Li et al., 2019; Cai et al., 2020; Michno et al., 2020; Zheng et al., 2020). So far, very limited success has been obtained in HR-mediated GT in soybean, with the majority of regenerated plants being chimeric and mutations not inherited to the next generation (Li et al., 2015).

Here we report a novel *O. haywardense*-mediated embryonic axis transformation system for high frequency GT in soybean, where the resultant HR-modified target locus is selectable-marker-free. Using two different donor designs 2-3.4% GT was achieved in two soybean genomic sites. Analyses of T1 progeny confirmed Mendelian inheritance for the majority of GT events, while randomly integrated T-DNA segregated independently of the GT locus resulting in T-DNA-free GT-positive T1 progeny. The method being simple, robust, highly efficient and reproducible, could expedite soybean precision breeding for yield and other traits.

Results

Gene Targeting (GT) experiment design

An overview of the GT construct design used in this study is described in Figure 1. The T-DNA construct (Figure 1A and Supplementary Figure 1A) targeting soybean DD38 (Cigan et al., 2016) genomic site contained five expression cassettes. The expression cassette nearest the LB encoded a *hygromycin*-resistance gene (*hptII*) donor DNA repair template flanked by homology arms, HR1 and HR2, which were in turn flanked by two Cas9 cut sites matching the genomic target site. The T-DNA construct also contained in order between the RB and *hptII* cassette a single guide

Selection-free gene targeting in soybean

RNA expression cassette, a Cas9 expression cassette, a *spectinomycin*-resistance gene (*spcN*), and a *DsRED* color marker cassette. Following T-DNA transfer, transient Cas9 expression results in two DSBs at the Cas9 cut sites next to the two homology arms flanking *hptII*, leading to release of the donor DNA repair template. The T-DNA containing *spcN* as the plant selectable marker, which is randomly inserted in the genome, is used for positive selection and regeneration of transformed cells. In addition, a DSB generated at the genomic target locus, which could then be repaired via HR using the donor template released from T-DNA, resulting in GT. The T-DNA could also be initially randomly integrated into the genome, with Cas9 expression leading to the release of the donor DNA repair template resulting in intragenomic GT as described previously (Barone et al., 2020). The donor DNA repair template in T-DNA construct (Figure 1B and Supplementary Figure 1B) targeting soybean DD51 (Cigan et al., 2016) genomic site contained yellow fluorescent protein (*YFP*) gene.

GT in soybean

A plant transformation construct pDD38 containing constitutively expressed *Cas9* was used for GT at soybean DD38 site (Figure 1A and Supplementary Figure 1A). *O. haywardense*-mediated embryonic axis transformation was carried out to generate stably transformed shoots ~1.5 cm in size using spectinomycin selection as described in Materials and Methods. In 6-8 weeks after transformation, a total of 466 shoots were regenerated and analyzed for GT and targeted NHEJ-mediated mutation at DD38 site. HR-mediated GT was detected by HR1 and HR2 junction qPCR as depicted in Figure 1A. The junction PCR was designed for quick screening of putative GT positive plants, further analyses for perfect GT was performed in T1 generation. Out of 466 shoots analyzed, HR1 junction-PCR positive product was observed in 48

Selection-free gene targeting in soybean

(10%) shoots, while 43 (9.2%) shoots were found positive for HR2 junction- PCR (Figure 2 and Table 1). Positive PCR products for both HR1 and HR2 junctions were obtained in 16 plants indicating 3.4% putative GT at DD38 site. Targeted insertion mutation analysis of the target site revealed 93.3% of shoots containing mutations in at least one allele, while >7% of the plants were observed with bi-allelic mutations (Table 1).

After having successfully demonstrated GT at DD38 site, we created another donor construct pDD51 (Figure 1B and Supplementary Figure 1B) targeting DD51 site in soybean genome. The construct pDD51 contained *YFP* cassette as donor repair template while *DsRed* was not included in the T-DNA, the rest of the construct design was similar to pDD38. Using *O. haywardense*-mediated embryonic axis transformation 690 shoots were generated on spectinomycin selection. HR1 and HR2 junction PCR revealed 2.8% and 2% positive plants (Table 2), respectively. Fourteen plants were observed positive for both HR1 and HR2 junctions demonstrating 2% GT at DD51 site. Similar to the DD38 site, high frequency (89%) NHEJ-mediated mutations were detected at the DD51 site, with 9.5% of plants showing bi-allelic mutations.

Inheritance and segregation analysis of GT plants

To study inheritance and segregation of GT, four GT positive T0 plants at DD38 site (designated as P1 – P4) were self-pollinated to generate T1 progeny. Approximately 90 T1 plants from each of the 4 GT-positive T0 plants were analyzed for the presence of GT and T-DNA components. To confirm GT and validate integrity of the insertion, HR1/HR2 junction PCR data were supplemented with Southern by Sequencing (SbS(Zastrow-Hayes et al., 2015)) (Table 3). Approximately 75% progeny from 3 GT-positive T0 plants (P2, P3 and P4) were observed positive for Selection-free gene targeting in soybean

both HR1 and HR2 junction PCR, conformed to expected Mendelian inheritance of the GT insertion at DD38 locus. These plants contained 1-2 copies of randomly inserted T-DNA components that segregated 1:2:1 (homo:hemi:null) as expected for a single locus insertion. For two of the four DD38 insertion lines, the number of T1 plants positive for both the HR1 and HR2 junctions and also null for the randomly inserted T-DNA, conformed to the expected Mendelian ratios of ~25% (16 nulls out of 67 for P3, 22 nulls out of 71 for P4). Only one T-DNA-free GT T1 plant was observed from P2 plant. The T1 progeny from P1 T0 plant did not show expected Mendelian inheritance, only 7 plants (~7%) were observed GT positive and no T-DNA-free GT T1 plant was obtained (Table 3).

Further molecular characterization of T-DNA component-free GT plants was done using SbS, which utilizes in-solution sequence capture coupled with NGS(Zastrow-Hayes et al., 2015). The method is highly sensitive in detection of construct-to-genome and construct-to-construct novel junctions that could have been created during the transformation process. The junction sequence data is then used to detect unintentionally inserted sequences from the transformation plasmid and help identify small T-DNA fragments, truncations of the intended donor DNA for GT, or presence of transformation plasmid backbone sequences. One T1 plant each from P1 and P2 and five T1 plants each from P3 and P4 T0 plants were sampled for SbS (Table 3). The P1 plant although PCR-negative for HR1 was included to confirm sensitivity of SbS. In addition, samples from wild-type negative control and plasmid-spiked positive control were also included.

The sequence coverage graphs of the GT plants and controls were mapped to expected schematic of precise donor insertion at DD38 genomic site (Figure 3 A) and to schematic of the construct used (Figure 3B) to confirm precise donor DNA

insertion and absence of unintentional randomly inserted construct sequences. The SbS analyses revealed T-DNA-free precise GT in P2 and four P4 plants A representative result is shown in Figure 3AIII and BIII. One T1 plant from P4 was observed to contain unexpected unique junction at DD38 target site indicating insertion of multiple copies of rearranged donor DNA (Figure 3 A VI). Similarly, all five P3 progeny plants were found to contain rearranged donor DNA at DD38 target site (Figure 3 A V). In addition, these plants also contained small T-DNA and backbone fragments (Figure 3 B V) that were undetected by quick PCR screening conducted in T0 generation. As expected from the HR1 PCR analysis, one P1 progeny plant was observed to contain partial donor DNA insertion lacking majority of intended donor DNA sequence (Figure 3 A IV).

Discussion

Being a major source of plant proteins and oils, soybean is an economically important food and oilseed crop. The recent societal shift towards the plant-based meat alternatives (Choudhury et al.) has further increased the demand for soybean, making this crop even more important than ever. While Green Revolution has brought a considerable increase in the yields of rice, wheat and maize, only modest yield improvements have been made for soybean(Liu et al., 2020). The unique architecture of soybean plants with leaves, inflorescences, and pods growing at each node, and the determined internode number greatly affect the final yield. Therefore, soybean yield being a multifactorial, complex trait, is likely to require simultaneous and precise modification of several components to positively affect the yield. Modern precision breeding tools, including genome editing, present an opportunity for such targeted breeding approach. In addition, genome editing would be extremely useful in making

pathway engineering needed to develop soybean varieties with enhanced protein and oil contents(Aili Bao, 2020; Subedi et al., 2020).

CRISPR-Cas9 due to flexibility and ease of the design has become an important genome engineering tool for creating targeted DSBs required for precision breeding applications(Wiedenheft et al., 2011; Jinek et al., 2012; Cong et al., 2013; Xing et al., 2014). DSB at desired target site triggers induction of a DNA repair pathway. In somatic cells NHEJ remains a dominant pathway, which mainly involves ligation of unrelated sequences or sequences with micro-homologies (Salomon and Puchta, 1998; Puchta, 1999; Orel et al., 2003). Therefore, CRISPR-Cas9-mediated targeted mutagenesis, which is based on the NHEJ repair, is now becoming routine in plants that are amenable for transformation including soybean(Al Amin et al., 2019; Bao et al., 2019; Campbell et al., 2019; Cheng et al., 2019; Di et al., 2019; Do et al., 2019; Han et al., 2019; Li et al., 2019; Cai et al., 2020; Michno et al., 2020; Zheng et al., 2020). In contrast, GT for targeted insertion or allele replacement remains a huge challenge mainly due to low-frequency HR in somatic cells coupled with complexity of making externally supplied donor template accessible at DSB site(Mao et al., 2019; Rozov et al., 2019). The complex genetic changes needed to improve soybean yield, oil and protein traits are likely to require precise insertions and replacements of DNA sequences, which can only be accomplished by HR-mediated GT. The challenge of inefficient GT process is further confounded by the need of a selectable marker to recover, enrich and regenerate plants from rare GT cell(s), while the presence of the selectable marker in the final GT plants is undesirable for commercial deployment of the technology. Therefore, development of a GT system which does not have a selectable marker in the final plant is highly desirable for application of GT for commercial product development.

Here we report novel *O. haywardense* -mediated delivery of CRISPR-Cas9 and donor DNA template to obtain high frequency GT in soybean. The delivery method utilized embryonic axis derived from soybean seed to generate stably transformed shoots (Cho et al. Manuscript submitted). First, we demonstrated efficient GT at soybean DD38 genomic site. While up to 10% targeted insertion was observed by HR1 or HR2 junction PCR, 3.4% putative true GT was obtained as indicated by positive PCR for both HR1 and HR2 junctions. Second, to establish reproducibility of our GT method we targeted another genomic site (DD51) using a different construct design. Similar to DD38 site, we successfully generated 2% GT at DD51 site indicating that GT method described in this report is reliable and can be applicable for different genomic sites. High frequency (~90%) targeted mutagenesis both at DD38 and DD51 sites further confirm that our CRISPR-Cas9 genome editing is robust and efficient.

While the donor DNA template contained *hptII* or *YFP* gene cassette, they were not used for selection; instead, *spcN* located on the T-DNA outside the donor DNA sequence was used to identify plants with randomly inserted T-DNA. Such construct design enables the donor DNA to contain only the trait gene(s) and the GT at genomic target site to be free from selectable marker. The randomly inserted T-DNA containing *spcN* cassette can be segregated out in next generation to obtain clean GT plants without unintended T-DNA sequences. The unique donor construct design with Cas9 cut sites flanking donor template (Peterson et al., 2021) enabled such indirect selection using randomly inserted *spcN* selection cassette while GT at genomic target site remains selection free. Recently, we have applied similar construct design for an intra-genomic GT in maize, however, it utilized a heat-shock inducible expression of Cas9 (Barone et al., 2020). The GT work on soybean

described in this paper is based on the constitutive expression of Cas9. While we assume that selective regeneration advantage for the Cas9-expressing cells with donor released from T-DNA provided preferential enrichment for GT events, intra-genomic GT with donor released from the T-DNA randomly integrated into the plant genome cannot be ruled out. Therefore, similar to our maize intra-genomic GT(Barone et al., 2020), inducible Cas9 expression coupled with selectable marker activation would be an attractive approach to further enhance GT in soybean.

Lack of fast, efficient and reproducible transformation method has been a key bottleneck for soybean genome editing(Altpeter et al., 2016). The common transformation method used to transform soybean utilizes *Agrobacterium*-mediated DNA delivery to cotyledonary-node explants(Paz et al., 2006), which is not only inefficient (~4% transformation frequency) but often leads to chimerism and non-transgenic escape plants, making the transformation process labor-intensive and expensive(Zheng et al., 2020). The current status of soybean transformation probably explains the lack of a viable GT system for commercial product development. The only previously published report on GT in soybean utilized particle bombardment-mediated delivery, showing GT in callus stage with majority being chimeric and not inherited to next generation(Li et al., 2015). In contrast, novel *O. haywardense* H1-8 mediated delivery used here is fast, highly efficient (~35% transformation frequency), reproducible, and generates non-chimeric plants (Cho et al. Manuscript submitted). The transformation method coupled with the construct design described in this report was instrumental for successful demonstration of efficient, heritable, and selectable marker free GT system paving the way for next-generation precision breeding in soybean.

After demonstrating a successful GT in T0 plants, the next step was to validate that our GT method produces non-chimeric plants with stable and precise insertion, while unintentionally randomly inserted T-DNA sequences are segregated away to generate transgene-free GT plants in T1 generation. Among self-pollinated T1 progeny planted from 4 GT-positive T0 plants, Mendelian inheritance of GT at DD38 locus and randomly inserted T-DNA was observed for progeny from 3 T0 plants. Expected 25% T-DNA-free GT positive T1 plants were obtained from two T0 GT plants. Only one T-DNA-free GT plant was observed from the third plant indicating genetic linkage between DD38 target site and T-DNA insertion site in this line. The T1 progeny from one plant showed chimeric GT and no T-DNA-free GT T1 plant was obtained in this line. The next goal was to confirm the fidelity of PCR-positive GT call and ensure selected GT T1 plants are free from any unintentionally inserted T-DNA sequences. We employed high-resolution SbS(Zastrow-Hayes et al., 2015) to further characterize selected transgene-free GT positive T1 plants. Compared to the whole-genome sequencing, SbS utilizes sequence capture technology, which reduces sequence analysis complexity by enriching the target sequences of interest, in turn decreasing the amount of sequence data generated by NGS technology. Bioinformatic analysis of the targeted sequencing identifies novel plasmid-to-genome or plasmid-to-plasmid junctions providing comprehensive information about the number of unique insertion loci, potential rearrangements of the inserted DNA, and the presence or absence of plasmid backbone sequences. SBS analyses revealed random T-DNA-free precise GT in T1 plants from two T0 lines. Some T1 plants from these lines were also observed with insertion of multiple copies of rearranged donor DNA. All T1 plants screened for one T0 line had rearranged donor DNA at DD38 target site, in addition to containing small T-DNA and backbone fragments that were undetected by quick PCR

screening in T0 generation. More extensive diagnostic PCR in T0 could have detected these imperfect GT or T-DNA fragment-containing events.

In summary, using a novel *O. haywardense* H1-8 mediated delivery we have developed efficient and reliable GT system in a major dicot crop plant. This work also provides general analytical guidelines for T0 to T1 generations post transformation for selecting plants with precise GT and void of any unintentionally inserted plasmid DNA sequences. Given gene targeting without a selectable marker gene or any transgene in the final plant is required for non-transgenic applications of genome editing technology, successful first demonstration of the system design for a selectable marker-free GT in this report is a major step towards commercial application of CRISPR-Cas9 genome editing in soybean precision breeding.

Methods

Plasmids and reagents used for plant transformation

Standard cloning methods were used to construct DNA plasmids used in this study. All plasmids were quality controlled through deep sequencing and 100% identity was confirmed. Supplementary Table 1 and Supplementary Figure 2 provide the details of expression and other elements used in different constructs described in Figure 1 and Supplementary Figure 1.

Soybean transformation

Corteva Agriscience elite soybean variety 93Y21 was used. Mature dry seeds were surface-sterilized for 16 hours using chlorine gas as described by Di et al. (Di et al., 1996). Disinfected seeds were imbibed on solid agar medium containing 5 g/L sucrose and 6 g/L agar for 6-8 hrs and then the seeds were soaked in distilled sterile

water for overnight at room temperature in the dark. Intact embryonic axes were isolated using a scalpel blade.

All *O. haywardense* H1-8 -mediated soybean transformation was carried out prior to 2020 as described by Cho et al. (Cho et al. Manuscript submitted). *O. haywardense* H1-8 lines containing the vectors listed in Figure 1 and Supplementary Figure 1 were used for transformation. *O. haywardense* H1-8 suspension (OD 0.5 at 600 nm) in infection medium composed of 1/10X Gamborg B5 basal medium, 30 g/L sucrose, 20 mM MES, 0.25 mg/L GA3, 1.67 mg/L BAP, 200 μ M acetosyringone and 1 mM dithiothreitol in pH 5.4 was added to about 200-300 EAs on a 25 x 100 mm deep petri dish. The plates were sealed with parafilm, then sonicated for 30 seconds. After sonication, EAs were incubated 2 hrs at room temperature. After inoculation, about 200-300 EAs were transferred to a single layer of autoclaved sterile filter paper (Cat No. 28320-020, VWR) in 25 x 100 mm petri dish. The plates were sealed with Micropore tape (1530-0, 3M, St. Paul, MN, USA) and incubated under dim light (1-2 μ E/m²/s), cool white fluorescent lamps for 16 hours at 21°C for 3 days. After co-cultivation, the base of each embryonic axis was embedded in shoot induction medium (R7100, PhytoTech Labs) containing 30 g/L sucrose, 6 g/L agar, 25 mg/L spectinomycin as a selectable agent and 250 mg/L cefotaxime (GoldBio, ST Louis, MO, USA) in pH5.7. Shoot induction was carried out in growth room at 26°C with a photoperiod of 16 hours and a light intensity of 60 - 100 μ E/m²/s.

After 4-6 weeks in selection medium, the spectinomycin-resistant shoots were cut and transferred to ½ strength MS rooting medium (M404, PhytoTech Labs) containing 15 g/L sucrose, agar 6 g/L, 10 mg/L spectinomycin and 250 mg/L cefotaxime for further shoot and root elongations. Transformation efficiency was calculated based on the number of positive transgenic soybean T0 plants divided by

the total number of EAs. Transgenic soybean plantlets were transferred to moistened Berger BM2 soil (Berger, Saint-Modeste, QC, Canada), and hardened plantlets were potted in 2-gallon pots containing moistened SunGro 702 and grown to maturity for harvest in a greenhouse.

Molecular analysis

Plants were sampled by leaf punching at 7 days after transplantation for quantitative PCR (qPCR) to confirm the presence of the selectable marker (*spcN*), CRISPR-Cas9 (PSN1), the left (PSB1) and right border (PSA2) regions. The *spcN*, PSN1, PSB1, and PSA2 regions for qPCR are shown in T-DNA maps in Supplementary Figure 1 and PCR conditions are given in Wu et al. (2014)(Wu et al., 2014). The primers and probes used for qPCR assays are described in Supplementary Table 2. DNA was extracted from the 7 mm leaf disk with sbeadex™ chemistry (LGC Biosearch, Middlesex, United Kingdom) as per vendor recommendations and resulting DNA was analyzed for quality and quantity via DropSense (Unchained Labs, Pleasanton, CA). PCR to detect mutation and HR junction PCR have been described previously(Barone et al., 2020). Mutation frequencies at genomic target sites were analyzed by qPCR using primer pair and probes given in Supplementary Table 3. HR Junctions were analyzed by using HR spanning primers (F1 and R1) described in in Supplementary Table 3.

Southern-by-sequencing data were generated according to Zastrow-Hayes et al.(Zastrow-Hayes et al., 2015). Briefly, genomic DNA was isolated from 4-mm leaf punches using the Sbeadex Maxi Plant kit (LGC Genomics) then randomly sheared to an average size of ~400bps with a Covaris E210 ultrasonicator (Covaris). Indexed Illumina genomic libraries were generated with the KAPA HTP Library Preparation Selection-free gene targeting in soybean

Kit (Roche). Adapter-ligated DNA fragments were subsequently PCR-amplified for 8 cycles, pooled in equal molar ratios then normalized in nuclease-free water to 5ng/μl. Following library construction, pooling and normalization, target enrichment was performed using a modified Roche Nimblegen double capture protocol(Zastrow-Hayes et al., 2015), with biotinylated DNA probes designed from the entire plasmid sequences. Enriched DNA fragments were sequenced on an Illumina NextSeq sequencer, according to the manufacturer's instructions. The resulting DNA sequences were quality-trimmed and aligned to the plasmid sequence for coverage analysis and characterization of the event. Positive SbS controls were derived from wild-type genomic DNA combined with spiked-in plasmid DNA for Illumina genomic library construction and enrichment.

Authors' contributions

S.K., Z.B.L., and A.A designed the experiment. N.S.D, J.B. and H.J.C. performed the soybean transformation. B.L., A.W. and A.R. designed and performed DNA analysis. S.K. led the project and wrote the manuscript. All the authors reviewed the manuscript.

Acknowledgements

We thank Lijuan Wang and Super Vector team for vector construction, Shujun Chang and Sam Ellis for soybean transformation support, Joel Van Wyk, and Becca Vickroy for greenhouse care, and Marv Taylor and Balaji Boovaraghan and the PCR Analysis and Characterization team for high throughput PCR support. The authors also thank Scott Betts, Maria Fedorova, and Sendil Devadas for critical review of the

manuscript. We acknowledge Clara Alarcon, Todd Jones and Doane Chilcoat for providing resources and support for this work.

Conflict of Interest

Authors are employees of Corteva Agriscience™. Some of the authors are inventors on pending applications on this work.

Data availability

The data supporting the findings of this study are available within the paper and its supplementary information files.

References

- Aili Bao CZ, Yi Huang, Haifeng Chen, Xinan Zhou, Dong Cao * (2020) Genome editing technology and application in soybean improvement. *Oil Crop Science* **5**: 31-40
- Al Amin N, Ahmad N, Wu N, Pu X, Ma T, Du Y, Bo X, Wang N, Sharif R, Wang P (2019) CRISPR-Cas9 mediated targeted disruption of FAD2-2 microsomal omega-6 desaturase in soybean (*Glycine max*.L). *BMC Biotechnol* **19**: 9
- Altpeter F, Springer NM, Bartley LE, Blechl AE, Brutnell TP, Citovsky V, Conrad LJ, Gelvin SB, Jackson DP, Kausch AP, Lemaux PG, Medford JI, Orozco-Cardenas ML, Tricoli DM, Van Eck J, Voytas DF, Walbot V, Wang K, Zhang ZJ, Stewart CN, Jr. (2016) Advancing Crop Transformation in the Era of Genome Editing. *Plant Cell* **28**: 1510-1520
- Selection-free gene targeting in soybean

- 403 **Bao A, Chen H, Chen L, Chen S, Hao Q, Guo W, Qiu D, Shan Z, Yang Z, Yuan**
 404 **S, Zhang C, Zhang X, Liu B, Kong F, Li X, Zhou X, Tran LP, Cao D**
 405 (2019) CRISPR/Cas9-mediated targeted mutagenesis of GmSPL9 genes alters
 406 plant architecture in soybean. *BMC Plant Biol* **19**: 131
- 407 **Barone P, Wu E, Lenderts B, Anand A, Gordon-Kamm W, Svitashhev S, Kumar**
 408 **S** (2020) Efficient Gene Targeting in Maize Using Inducible CRISPR-Cas9
 409 and Marker-free Donor Template. *Molecular Plant* **13**: 1219-1227
- 410 **Cai Y, Wang L, Chen L, Wu T, Liu L, Sun S, Wu C, Yao W, Jiang B, Yuan S,**
 411 **Han T, Hou W** (2020) Mutagenesis of GmFT2a and GmFT5a mediated by
 412 CRISPR/Cas9 contributes for expanding the regional adaptability of soybean.
 413 *Plant Biotechnol J* **18**: 298-309
- 414 **Campbell BW, Hoyle JW, Bucciarelli B, Stec AO, Samac DA, Parrott WA,**
 415 **Stupar RM** (2019) Functional analysis and development of a CRISPR/Cas9
 416 allelic series for a CPR5 ortholog necessary for proper growth of soybean
 417 trichomes. *Sci Rep* **9**: 14757
- 418 **Cheng Q, Dong L, Su T, Li T, Gan Z, Nan H, Lu S, Fang C, Kong L, Li H, Hou**
 419 **Z, Kou K, Tang Y, Lin X, Zhao X, Chen L, Liu B, Kong F** (2019)
 420 CRISPR/Cas9-mediated targeted mutagenesis of GmLHY genes alters plant
 421 height and internode length in soybean. *BMC Plant Biol* **19**: 562
- 422 **Choudhury D, Singh S, Seah JSH, Yeo DCL, Tan LP** Commercialization of Plant-
 423 Based Meat Alternatives. *Trends in Plant Science*
- 424 **Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove**
 425 **AJ, Voytas DF** (2010) Targeting DNA double-strand breaks with TAL
 426 effector nucleases. *Genetics* **186**: 757-761

- 427 **Cigan M, Gao H, Liu Z-B, Mutti J, Podlich D, Scelonge C** (2016) Generation Of
428 Site-specific-integration Sites For Complex Trait Loci In Corn And Soybean,
429 And Methods Of Use. *In*. 2016/040030 A1
430 WO
- 431 **Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W,**
432 **Marraffini LA, Zhang F** (2013) Multiplex genome engineering using
433 CRISPR/Cas systems. *Science* **339**: 819-823
- 434 **Di R, Purcell V, Collins GB, Ghabrial SA** (1996) Production of transgenic soybean
435 lines expressing the bean pod mottle virus coat protein precursor gene. *Plant*
436 *Cell Rep* **15**: 746-750
- 437 **Di YH, Sun XJ, Hu Z, Jiang QY, Song GH, Zhang B, Zhao SS, Zhang H** (2019)
438 Enhancing the CRISPR/Cas9 system based on multiple GmU6 promoters in
439 soybean. *Biochem Biophys Res Commun* **519**: 819-823
- 440 **Do PT, Nguyen CX, Bui HT, Tran LTN, Stacey G, Gillman JD, Zhang ZJ,**
441 **Stacey MG** (2019) Demonstration of highly efficient dual gRNA
442 CRISPR/Cas9 editing of the homeologous GmFAD2-1A and GmFAD2-1B
443 genes to yield a high oleic, low linoleic and alpha-linolenic acid phenotype in
444 soybean. *BMC Plant Biol* **19**: 311
- 445 **Epinat JC, Arnould S, Chames P, Rochaix P, Desfontaines D, Puzin C, Patin A,**
446 **Zanghellini A, Paques F, Lacroix E** (2003) A novel engineered
447 meganuclease induces homologous recombination in yeast and mammalian
448 cells. *Nucleic Acids Res* **31**: 2952-2962
- 449 **Gasiunas G, Barrangou R, Horvath P, Siksnys V** (2012) Cas9-crRNA
450 ribonucleoprotein complex mediates specific DNA cleavage for adaptive
451 immunity in bacteria. *Proc Natl Acad Sci U S A* **109**: E2579-2586

- 452 **Graham PH, Vance CP** (2003) Legumes: Importance and Constraints to Greater
453 Use. *Plant Physiology* **131**: 872-877
- 454 **Han J, Guo B, Guo Y, Zhang B, Wang X, Qiu LJ** (2019) Creation of Early
455 Flowering Germplasm of Soybean by CRISPR/Cas9 Technology. *Front Plant*
456 *Sci* **10**: 1446
- 457 **Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E** (2012) A
458 programmable dual-RNA-guided DNA Endonuclease in adaptive bacterial
459 immunity. *Science*. **337**
- 460 **Kim YG, Cha J, Chandrasegaran S** (1996) Hybrid restriction enzymes: zinc finger
461 fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**: 1156-1160
- 462 **Li C, Nguyen V, Liu J, Fu W, Chen C, Yu K, Cui Y** (2019) Mutagenesis of seed
463 storage protein genes in Soybean using CRISPR/Cas9. *BMC Res Notes* **12**:
464 176
- 465 **Li M, Li X, Zhou Z, Wu P, Fang M, Pan X, Lin Q, Luo W, Wu G, Li H** (2016)
466 Reassessment of the Four Yield-related Genes *Gn1a*, *DEP1*, *GS3*, and *IPA1* in
467 Rice Using a CRISPR/Cas9 System. *Front Plant Sci* **7**: 377
- 468 **Li S, Gao F, Xie K, Zeng X, Cao Y, Zeng J, He Z, Ren Y, Li W, Deng Q, Wang S,**
469 **Zheng A, Zhu J, Liu H, Wang L, Li P** (2016) The *OsmiR396c-OsGRF4-*
470 *OsGIF1* regulatory module determines grain size and yield in rice. *Plant*
471 *Biotechnol J* **14**: 2134-2146
- 472 **Li X, Wang Y, Chen S, Tian H, Fu D, Zhu B, Luo Y, Zhu H** (2018) Lycopene Is
473 Enriched in Tomato Fruit by CRISPR/Cas9-Mediated Multiplex Genome
474 Editing. *Front Plant Sci* **9**: 559

- 475 **Li Z, Liu ZB, Xing A, Moon BP, Koellhoffer JP, Huang L, Ward RT, Clifton E,**
 476 **Falco SC, Cigan AM** (2015) Cas9-Guide RNA Directed Genome Editing in
 477 Soybean. *Plant Physiol* **169**: 960-970
- 478 **Liu J, Chen J, Zheng X, Wu F, Lin Q, Heng Y, Tian P, Cheng Z, Yu X, Zhou K,**
 479 **Zhang X, Guo X, Wang J, Wang H, Wan J** (2017) GW5 acts in the
 480 brassinosteroid signalling pathway to regulate grain width and weight in rice.
 481 *Nat Plants* **3**: 17043
- 482 **Liu S, Zhang M, Feng F, Tian Z** (2020) Toward a "Green Revolution" for Soybean.
 483 *Mol Plant* **13**: 688-697
- 484 **Mao Y, Botella JR, Liu Y, Zhu J-K** (2019) Gene editing in plants: progress and
 485 challenges. *National Science Review* **6**: 421-437
- 486 **Michno JM, Virdi K, Stec AO, Liu J, Wang X, Xiong Y, Stupar RM** (2020)
 487 Integration, abundance, and transmission of mutations and transgenes in a
 488 series of CRISPR/Cas9 soybean lines. *BMC Biotechnol* **20**: 10
- 489 **Nonaka S, Arai C, Takayama M, Matsukura C, Ezura H** (2017) Efficient increase
 490 of -aminobutyric acid (GABA) content in tomato fruits by targeted
 491 mutagenesis. *Sci Rep* **7**: 7057
- 492 **Orel N, Kyryk A, Puchta H** (2003) Different pathways of homologous
 493 recombination are used for the repair of double-strand breaks within tandemly
 494 arranged sequences in the plant genome. *Plant J* **35**: 604-612
- 495 **Paz MM, Martinez JC, Kalvig AB, Fonger TM, Wang K** (2006) Improved
 496 cotyledonary node method using an alternative explant derived from mature
 497 seed for efficient *Agrobacterium*-mediated soybean transformation. *Plant Cell*
 498 *Rep* **25**: 206-213

499 **Peterson D, Barone P, Lenderts B, Schwartz C, Feigenbutz L, St Clair G, Jones**
500 **S, Svitashv S** (2021) Advances in Agrobacterium transformation and vector
501 design result in high frequency targeted gene insertion in maize. *Plant*
502 *Biotechnol J*

503 **Puchta H** (1999) Double-strand break-induced recombination between ectopic
504 homologous sequences in somatic plant cells. *Genetics* **152**: 1173-1181

505 **Puchta H** (2017) Applying CRISPR/Cas for genome engineering in plants: the best is
506 yet to come. *Curr Opin Plant Biol* **36**: 1-8

507 **Ray DK, Mueller ND, West PC, Foley JA** (2013) Yield Trends Are Insufficient to
508 Double Global Crop Production by 2050. *PLoS One* **8**: e66428

509 **Rozov SM, Permyakova NV, Deineko EV** (2019) The Problem of the Low Rates of
510 CRISPR/Cas9-Mediated Knock-ins in Plants: Approaches and Solutions.
511 *International Journal of Molecular Sciences* **20**: 3371

512 **Salomon S, Puchta H** (1998) Capture of genomic and T-DNA sequences during
513 double-strand break repair in somatic plant cells. *EMBO J* **17**: 6086-6095

514 **Shi J, Gao H, Wang H, Lafitte HR, Archibald RL, Yang M, Hakimi SM, Mo H,**
515 **Habben JE** (2017) ARGOS8 variants generated by CRISPR-Cas9 improve
516 maize grain yield under field drought stress conditions. *Plant Biotechnol J* **15**:
517 207-216

518 **Subedi U, Jayawardhane KN, Pan X, Ozga J, Chen G, Foroud NA, Singer SD**
519 (2020) The Potential of Genome Editing for Improving Seed Oil Content and
520 Fatty Acid Composition in Oilseed Crops. *Lipids* **55**: 495-512

521 **Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A,**
522 **Westphal W, Heck AJ, Boekema EJ, Dickman MJ, Doudna JA** (2011)
523 RNA-guided complex from a bacterial immune system enhances target

524 recognition through seed sequence interactions. *Proc Natl Acad Sci U S A*
525 **108**: 10092-10097

526 **Wood RD** (1996) DNA repair in eukaryotes. *Annu Rev Biochem* **65**: 135-167

527 **Wu E, Lenderts B, Glassman K, Berezowska-Kaniewska M, Christensen H,**
528 **Asmus T, Zhen S, Chu U, Cho MJ, Zhao ZY** (2014) Optimized
529 *Agrobacterium*-mediated sorghum transformation protocol and molecular data
530 of transgenic sorghum plants. *In Vitro Cell Dev Biol Plant* **50**: 9-18

531 **Xing HL, Dong L, Wang ZP, Zhang HY, Han CY, Liu B, Wang XC, Chen QJ**
532 (2014) A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC*
533 *Plant Biol* **14**: 327

534 **Yu QH, Wang B, Li N, Tang Y, Yang S, Yang T, Xu J, Guo C, Yan P, Wang Q,**
535 **Asmutola P** (2017) CRISPR/Cas9-induced Targeted Mutagenesis and Gene
536 Replacement to Generate Long-shelf Life Tomato Lines. *Sci Rep* **7**: 11874

537 **Zastrow-Hayes GM, Lin H, Sigmund AL, Hoffman JL, Alarcon CM, Hayes KR,**
538 **Richmond TA, Jeddeloh JA, May GD, Beatty MK** (2015) Southern-by-
539 Sequencing: A Robust Screening Approach for Molecular Characterization of
540 Genetically Modified Crops. *The Plant Genome* **8**:
541 plantgenome2014.2008.0037

542 **Zheng N, Li T, Dittman JD, Su J, Li R, Gassmann W, Peng D, Whitham SA, Liu**
543 **S, Yang B** (2020) CRISPR/Cas9-Based Gene Editing Using Egg Cell-Specific
544 Promoters in Arabidopsis and Soybean. *Frontiers in Plant Science* **11**
545
546

Figure legends

Figure 1. Schematic description of the constructs pDD38 (A) and pDD51 (B)

containing donor template for gene targeting at soybeanDD38 and DD51

genomic site, respectively. Both constructs contain guide RNA (gRNA) and

Cas9 expression cassettes for genome editing, and *spectinomycin* (*spcN*) cassette as

plant selectable marker. The construct A for DD38 site also contained *DsRED* color

marker. Donor DNA repair template contained *Hygromycin* (*hptII*, A) or *YFP* (B)

gene cassette flanked by homology arms (HR1 & HR2) and Cas9 cut sites (shown as

scissors) matching DD38 or DD51 genomic site. Following Cas9 and gRNA

expression the donor template will be released and used as template for repair at

DD38 or DD51 target site for gene targeting (Bottom panel). HR1 and HR2

diagnostic PCRs were done to detect gene targeting as described in Figure 2. RB &

LB: Right and left T-DNA border, respectively. The component sizes are not to the

scale.

Figure 2. Quick PCR screening to detect gene targeting. HR1 (Upper panel) and

HR2 PCR (Lower panel) of leaf samples 466 T0 shoots regenerated for DD38 GT

experiment. Positive PCR products of expected sizes obtained both for HR1 and HR2

junctions are shown in black frames. PCR products amplified only for HR1 or HR2

are indicated in red frames. The PCR results are summarized in Table 1.

Figure 3. Southern-by-Sequencing (SbS) for high resolution molecular

characterization of PCR positive gene targeted T1 soybean plants. A. Sequence

coverage graphs of targeted plants and controls mapped to expected schematic of

precise donor insertion at DD38 genomic site. I. Schematic representation of the

Selection-free gene targeting in soybean

572 precisely targeted soybean DD38 site, which is mapped to sequence coverage graph
573 of wild type plant as negative control (**II**), representative precisely targeted P2 T1
574 plant (**III**), P2 T1 plant observed with imprecise insertion (**IV**) due to missing part of
575 the donor (indicted as dotted red boxes), P3 and P4 T1 plants (**V and VI**,
576 respectively) showing imprecise insertion due to rearrangements at targeted locus
577 (indicted as red boxes), WT plant DNA with spiked in plasmid DNA used for gene
578 targeting as positive SBS control (**VII**). **B.** Sequence coverage graphs of targeted
579 plants and controls mapped to schematic of transgene used. **I.** Schematic
580 representation of plasmid map aligned to sequence coverage graph of plants and
581 controls described in **A. I-VII** above. Red box indicates presence of plasmid DNA,
582 dotted box shows missing part of donor. Red font in the maps indicate soybean
583 genomic sequences
584

Table 1: Summary of gene targeting and mutation analyses at soybean DD38 genomic site

Plants analyzed	Gene Targeting			Mutation		
	<i>HR1</i>	<i>HR2</i>	<i>HR1 & HR2</i>	<i>Mutated plants</i>	<i>Mono-allelic</i>	<i>Bi-allelic</i>
466	48 (10%)	43 (9.2%)	16 (3.4%)	435 (93.3%)	401 (86%)	34 (7.3%)

*qPCR read from 0-0.099: bi-allelic ; 0.1-0.7: mono-allelic; 0.7-1.0: no mutation (WT).

Table 2: Summary of gene targeting and mutation analyses at soybean DD51 genomic site

Plants analyzed	Gene Targeting			Mutation		
	<i>HR1</i>	<i>HR2</i>	<i>HR1 & HR2</i>	<i>Mutated plants</i>	<i>Mono-allelic</i>	<i>Bi-allelic</i>
690	22 (3.2%)	19 (2.8%)	14 (2%)	614 (89%)	548 (78.5%)	66 (9.5%)

*qPCR read from 0-0.099: bi-allelic ; 0.1-0.7: mono-allelic; 0.7-1.0: no mutation (WT).

Table 3: Inheritance and Segregation of GT in T1 Generation

<i>T0 plant</i>	<i>Total T1 plants</i>	T-DNA				Gene targeting		
		<i>Nulls</i>	<i>Homo</i>	<i>Hemi</i>	<i>Copy# (PSB1)</i>	<i>HR1/2</i>	<i>Positive T1 plants</i>	<i>SBS*</i>
						<i>Total</i>	<i>Trangene free</i>	
P1	91	31	21	39	3	7	1 ^ϕ	Deletion
P2	93	19	16	58	1	71	1	Clean
P3	86	19	26	41	2	67	16	Rearrangement
P4	95	26	20	49	2	71	22	Clean

^ϕHR2 only

*Maximum 5 plants analyzed

Figure 1.

Figure 1. Schematic description of the constructs pDD38 (A) and pDD51 (B) containing donor template for gene targeting at soybean DD38 and DD51 genomic site, respectively. Both constructs contain guide RNA (gRNA) and Cas9 expression cassettes for genome editing, and *spectinomycin* (*spcN*) cassette as plant selectable marker. The construct A for DD38 site also contained *DsRED* color marker. Donor DNA repair template contained *Hygromycin* (*hptII*, A) or *YFP* (B) gene cassette flanked by homology arms (HR1 & HR2) and Cas9 cut sites (shown as scissors) matching DD38 or DD51 genomic site. Following Cas9 and gRNA expression the donor template will be released and used as template for repair at DD38 or DD51 target site for gene targeting (Bottom panel). HR1 and HR2 diagnostic PCRs were done to detect gene targeting as described in Figure 2. RB & LB: Right and left T-DNA border, respectively. The component sizes are not to the scale.

Figure 2. Quick PCR screening to detect gene targeting. HR1 (Upper panel) and HR2 PCR (Lower panel) of leaf samples 466 T0 shoots regenerated for DD38 GT experiment. Positive PCR products of expected sizes obtained both for HR1 and HR2 junctions are shown in black frames. PCR products amplified only for HR1 or HR2 are indicated in red frames. The PCR results are summarized in Table 1.