Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2020) **18**, pp. 721–731 doi: 10.1111/pbi.13239

Generation of a multiplex mutagenesis population via pooled CRISPR-Cas9 in soya bean

Mengyan Bai[1,†], Juehui Yuan[1,†], Huaqin Kuang[2,†], Pingping Gong[2], Suning Li[2,3], Zhihui Zhang[1,2], Bo Liu[1,2], Jiafeng Sun[1,2], Maoxiang Yang[1,2], Lan Yang[4], Dong Wang[4], Shikui Song[2] and Yuefeng Guan[2]*

[1]*College of Resources and Environment, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Fujian Agriculture and Forestry University, Fuzhou, China*

[2]*FAFU-UCR Joint Center for Horticultural Biology and Metabolomics, Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou, China*

[3]*College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, China*

[4]*Key Laboratory of Molecular Biology and Gene Engineering in Jiangxi Province, College of Life Science, Nanchang University, Jiangxi, China*

Received 7 November 2018;
revised 17 July 2019;
accepted 15 August 2019.
**Correspondence* (Tel +8615005002560;
fax 0591-83590961;
email guan@fafu.edu.cn)
†These authors contributed equally to this work

Keywords: CRISPR-Cas9, soya bean, multiplex mutagenesis, CRISPR population, nodulation.

Summary

The output of genetic mutant screenings in soya bean [*Glycine max* (L.) Merr.] has been limited by its paleopolypoid genome. CRISPR-Cas9 can generate multiplex mutants in crops with complex genomes. Nevertheless, the transformation efficiency of soya bean remains low and, hence, remains the major obstacle in the application of CRISPR-Cas9 as a mutant screening tool. Here, we report a pooled CRISPR-Cas9 platform to generate soya bean multiplex mutagenesis populations. We optimized the key steps in the screening protocol, including vector construction, sgRNA assessment, pooled transformation, sgRNA identification and gene editing verification. We constructed 70 CRISPR-Cas9 vectors to target 102 candidate genes and their paralogs which were subjected to pooled transformation in 16 batches. A population consisting of 407 T0 lines was obtained containing all sgRNAs at an average mutagenesis frequency of 59.2%, including 35.6% lines carrying multiplex mutations. The mutation frequency in the T1 progeny could be increased further despite obtaining a transgenic chimera. In this population, we characterized *gmric1/gmric2* double mutants with increased nodule numbers and *gmrdn1-1/1-2/1-3* triple mutant lines with decreased nodulation. Our study provides an advanced strategy for the generation of a targeted multiplex mutant population to overcome the gene redundancy problem in soya bean as well as in other major crops.

Introduction

Soya bean [*Glycine max* (L.) Merr.] is a globally important crop representing a major source of the oil and protein for human diet and feedstock consumption. In addition, as a legume with nodulation, soya bean plants also provide environmental benefits to farming practices. Biological studies on soya bean have, therefore, been driven by its importance worldwide.

Since the announcement of the soya bean draft genome, functional characterization of all 46–56 thousand annotated genes has been the primary goal in soya bean genetic studies (O'Rourke *et al.*, 2017; Schmutz *et al.*, 2010). Genetic mutagenesis, including chemical, radiation (X-rays, gamma rays and fast neutrons) and transformation-induced (T-DNA and transposon insertion) methods (Bolon *et al.*, 2011; Campbell and Stupar, 2017; Cui *et al.*, 2013; Espina *et al.*, 2018; Li *et al.*, 2017; Mathieu *et al.*, 2009; Tsuda *et al.*, 2015), is routinely used as forward genetic strategies. However, the complexity of the soya bean genome limits the success of these approaches. The soya bean genome became paleotetraploid after one polyploidization (2n = 20) and one diploidization (*n* = 20) (Schmutz *et al.*, 2010; Singh and Hymowitz, 1988). Due to multiple genome duplications, approximately 75% of the genes in soya bean are present in multiple copies leading to high genetic redundancy (Schmutz *et al.*, 2010). As a result, the outcomes of forward genetic studies are limited in soya bean compared with other model plant species.

Such challenges can be countered, at least in part, by the clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein (Cas9) (CRISPR-Cas9) system (Cong *et al.*, 2013; Jinek *et al.*, 2012; Li *et al.*, 2015). CRISPR-Cas9 has been successfully adapted in various plants and organisms (Bull *et al.*, 2018; Fauser *et al.*, 2014; Puchta, 2017; Wang *et al.*, 2018). Moreover, multiplex mutations can be generated by CRISPR-Cas9 by either cloning multiple sgRNAs or targeting homologous regions by a single sgRNA. Multiplex mutations have been induced by CRISPR-Cas9 in several crops, such as hexaploid wheat (Gil-Humanes *et al.*, 2017; Sánchez-León *et al.*, 2018), paleopolyploid maize (Char *et al.*, 2017) and allotetraploid rapeseed (Li *et al.*, 2018). In soya bean, the CRISPR-Cas9 system has been successfully utilized to create nonheritable somatic mutations in hairy root cultures and/or whole-plant heritable mutations (Cai *et al.*, 2015, 2018a,b; Curtin *et al.*, 2018; Du *et al.*, 2016; Jacobs *et al.*, 2015; Kanazashi *et al.*, 2018; Li *et al.*, 2015; Michno *et al.*, 2015; Sun *et al.*, 2015). It has been repeatedly shown that a single sgRNA can induce simultaneous mutations on endogenous genes in soya bean (Du *et al.*, 2016; Jacobs *et al.*, 2015; Kanazashi *et al.*, 2018; Sun *et al.*, 2015). The application of CRISPR-Cas9 on two *FLOWERLOCUST* (FT) loci, *GmFT2a* and *GmFT5a*, has successfully modified the flowering time and aided in the expansion of the regional adaptability of soya bean (Cai *et al.*, 2018a,b, 2019; Liu *et al.*, 2018). Thus, the CRISPR-Cas9 tool has a great potential in functional genomic study and molecular breeding in soya bean.

Recent studies have demonstrated that diversity in genetic variation can be generated in plants via pooled transformation of CRISPR-Cas9 libraries (Jacobs *et al*., 2017; Lu *et al*., 2017; Meng *et al*., 2017). Hence, a large-scale CRISPR-Cas9 screening for multiplex mutagenesis would be of great value for soya bean genetic research. Nevertheless, previous protocols of CRISPR-Cas9 library construction in model organisms have largely depended on high-efficiency transformation systems. However, the genetic transformation systems in soya bean have relatively low efficiency, questioning the feasibility of constructing CRISPR libraries as screening tools in this species.

In this study, we developed a pooled CRISPR-Cas9 platform with optimized procedures at multiple steps. Utilizing this platform, we generated a soya bean population with multiplex mutagenesis targeting more than 100 candidate genes. Our findings demonstrate that pooled CRISPR-Cas9 is efficient and feasible as a large-scale mutagenesis tool in soya bean.

Results

Design and construction of sgRNA libraries

To construct a targeted mutant population, we selected 74 candidate genes involved in nodulation and 28 genes with potential functions in seeds (Table S1). We used multiplex sgRNAs to target the paralogs of candidate genes. Overall, we designed 70 sgRNAs to target 102 genes in soya bean, including 26 sgRNAs that were highly specific to a single gene and 44 sgRNAs to target the homologous regions of two or more paralogous genes (Table S2).

A new CRISPR-Cas9 vector, pGES201, was developed for genome editing in soya bean. In this vector, BASTA was used as the transformation selection marker, and soya bean codon-optimized spCas9 was driven by a strong endogenous promoter, *proGmSCREAM M4* (*pM4*) (Figures 1a and S1, S2) (Zhang *et al*., 2015). In a trial experiment, pGES201 vector successfully induced gene editing in stable transgenic plants (Figure S2).

We next developed and optimized a procedure to generate a pooled CRISPR-Cas9 mutagenesis population in soya bean (Figure 1b). We constructed individual CRISPR-Cas9 vectors and then pooled small libraries for transformation (each pool contained 3–5 vectors) (Figure 1b). In total, 70 CRISPR-Cas9 vectors were cloned and pooled into 16 sublibraries (Table S2).

Generation of a multiplex mutagenesis population in soya bean

The sublibraries within 16 batches were transformed, BASTA-resistant seedlings were selected, and a total of 407 plants were PCR validated as transgene-positive. We then genotyped each T0 transgenic plant to determine sgRNA identity. The identification of sgRNAs in the individual plants was performed via sgRNA-specific PCRs (SSP) (Figure 1b and S3; Experimental Procedures). SSP results showed that all 70 sgRNAs could be identified in the T0 population, and most of the sgRNAs (65 out of 70) were present in more than 3 transgenic lines (Table S3 and Figure 2a, b). Among all the transformants, single sgRNA was found in 222 T0 plants. Examination of multiple sgRNA integration lines revealed that 126 plants contained two sgRNAs, 33 plants contained three sgRNAs, 21 plants contained four sgRNAs, and 5 plants contained all the five gRNAs. Therefore, approximately half of the T0 transgenic plants contained multiple sgRNAs (Table S3 and Figure 2a). These results suggest that a pooled transformation strategy can uniformly and randomly cover the CRISPR sublibraries in transgenic soya bean populations.

Next, we examined the extent of gene editing in the T0 populations using PCR and Sanger sequencing (Figure S4 and Table S3). We were able to characterize double or triple mutants in the pooled mutagenesis population via a single multiplex sgRNA or combinations of multiple sgRNAs. Therefore, a suite of combinatorial mutant plants was generated from the resulting CRISPR-Cas9 population. Among 407 transgenic plants, 96 lines contained mutations at one target locus and 145 lines carried mutations in at least two target loci (Table S3 and Figure 2a,c).

During the early stages of this study, 43 vectors were constructed (sublibraries A01 to B04) and used directly for stable transformations. At later stages, we optimized the procedure such that highly efficient sgRNAs in 27 vectors were selected by hairy root assay before stable transformation (Figure 1b) (B05 to C05 in 6 sublibraries). In comparison with the editing rate of 41.51% for sgRNAs without prescreening in hairy roots, the introduction of sgRNA efficiency assessment improved the editing frequency of the pipeline to 61.41%. Overall, the editing frequency across all sgRNAs averaged at 49.15% (Table S3 and Figure 2d). Similar results were obtained in terms of mutation frequency per targeted gene (Table S3 and Figure 2e). This result indicated that the hairy root assessment procedure could improve the efficiency of mutagenesis in the CRISPR-Cas9 population. Taken together, these results show that our pooled CRISPR-Cas9 procedure can produce a high frequency of target mutations in T0 transgenic soya bean plants.

Heritability of T-DNA and mutations in T1 progeny

A portion of the transgenic plants were propagated in the T1 generation, and we selected sublibrary A02, which consisted of 5 sgRNAs (sgRNA-006 to sgRNA-010) and 20 transgenic lines, for the detailed analysis of heritable population structure and editing efficiency (Figure 3 and Table S4). We first examined the inheritance of T-DNA in T1 progeny via SSP. For each transgenic line, 6–9 T1 plants were assessed. If a sgRNA was present in a T0 plant but not in any T1 plant, it was deemed noninheritable (Figure S5). If a sgRNA was detected in any T1 plant, it was considered heritable by genetic segregation (Figure S5). The SSP results showed that a significant portion of transgenes were not transmitted to the T1 generation (Figure 3 and Table S4). T-DNAs were absent in approximately half (11 of 20) of the lines (Figure 3 and Table S4), which is consistent with the high frequency of mosaicism via the cotyledonary node transformation method (Parrott *et al*., 1989). Nevertheless, most lines with multiple sgRNA integration in the T0 generation still contained T-DNAs in the T1 generation (Figure 3 and Table S4).

Next, we checked for gene editing in all the potential target loci in the T1 progeny. We found that most noninheritable sgRNAs failed to induce mutations in both T0 and T1 plants, with the exception of two lines. In line A02-14, a T0 mutation induced by sgRNA-010 was absent in T1 progeny along with the noninheritance of T-DNA (Figure 3 and Table S4). In contrast, a T0 mutation induced by sgRNA-007 was still present in line A02-02 T1 plants, despite the loss of T-DNA (Figure 3 and Table S4). In case of heritable sgRNAs, six T0 mutations were transmitted to the T1 generation. Moreover, inherited Cas9 and sgRNA expression induced new mutations in 6 lines (Figure 3 and Table S4). Overall, the mutation frequencies were 8 in 20 lines and 14 in 34 sgRNA events, which show an increase from 6 in 20 lines and 8 in

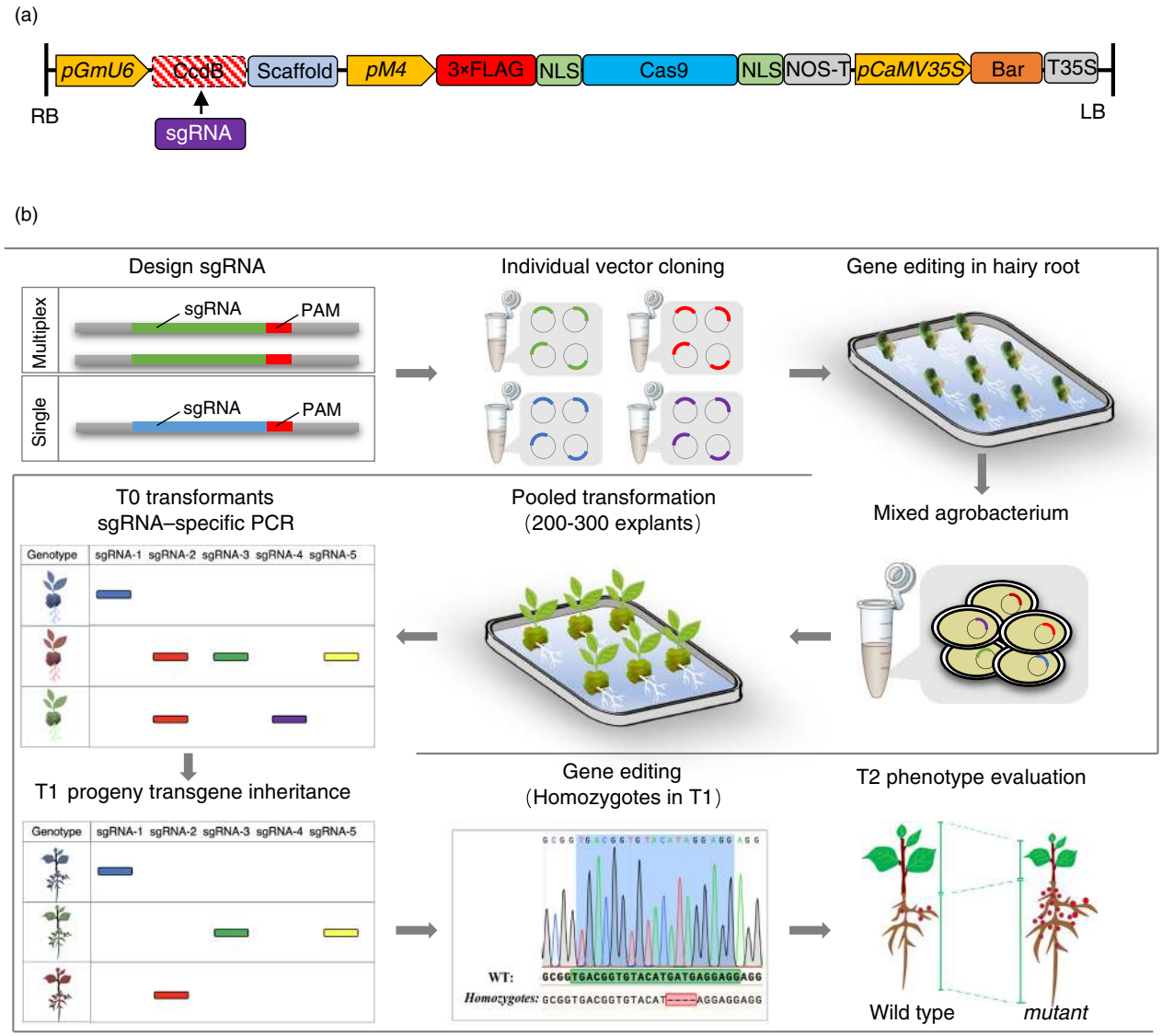

Figure 1 A pooled CRISPR-Cas9 mutagenesis procedure in soya bean. (a) sgRNAs are designed to target multiple loci or specific single loci, and CRISPR vectors are cloned individually. In the optimized procedure, the gene editing efficiency of each sgRNA is assessed in hairy roots. Highly efficient vectors are pooled into a sublibrary in *A. tumefaciens*, which is then transformed into soya bean. The T0 and T1 plants are genotyped by SSP (the coloured bars indicate sgRNA positive), and gene editing is verified by PCR and Sanger sequencing. Desired homozygous mutants will be obtained in the T1 or T2 generation and then subjected to phenotyping. (b) Architecture of the pGES201 vector. Type IIS BsaI restriction sites were used for sgRNA cloning between the U6 promoter and sgRNA scaffold. In between two BsaI restriction sites is a toxin gene, CcdB.pM4 was used to drive spCas9. 35S-driven BASTA is the selection marker.

34 sgRNA events in the T0 generation, respectively (Figure 3 and Table S4).

In A02-02 and A02-65 lines, we identified Cas9-free plants carrying targeted mutations, in which the T-DNA was lost due to mosaicism or segregation while edited genes were transmitted (Figure 3 and Table S4). In A02-14 and A02-40 lines, the T-DNA of sgRNA-009 was segregated but targeted mutations were identified. Nevertheless, these lines still contained other sgRNAs and were not yet 'Cas9-free' (Figure 3 and Table S4). These results showed that Cas9-free plants could be obtained in T1 progenies.

Identification of *gmric1/gmric2* double homozygous mutants with increased nodule number

Glyma.13G292300 (*GmRIC1*) and Glyma.06G284100 (*GmRIC2*) are paralogous genes that encode two nodule-enhanced *CLAVATA3/EMBRYO SURROUNDING REGION-RELATED* (CLE) peptides, which have been proposed to be the root-derived signals in autoregulation of nodulation (AON) (Reid *et al.*, 2011). However, previous genetic evidence regarding the roles of CLE peptides has been based mostly on ectopic expression in hairy roots, synthetic peptide treatment or RNAi assays (Ferguson *et al.*, 2014; Lim *et al.*, 2011; Okamoto *et al.*, 2013; Reid *et al.*, 2011). The lack of stable genetic materials has hindered the functional characterization of GmRIC1/GmRIC2 in soya bean.

In this study, we designed sgRNA-009 and sgRNA-010 specifically to target *GmRIC1* and *GmRIC2*, respectively. In the T2 generation, two double homozygous *gmric1/gmric2* mutant plants with different types of mutations were identified, which caused frameshifts in the protein sequence (A2-40-6-1: *gmric1* + 1 bp, *gmric2* −5 bp; A2-40-6-55: *gmric1* −4 bp and

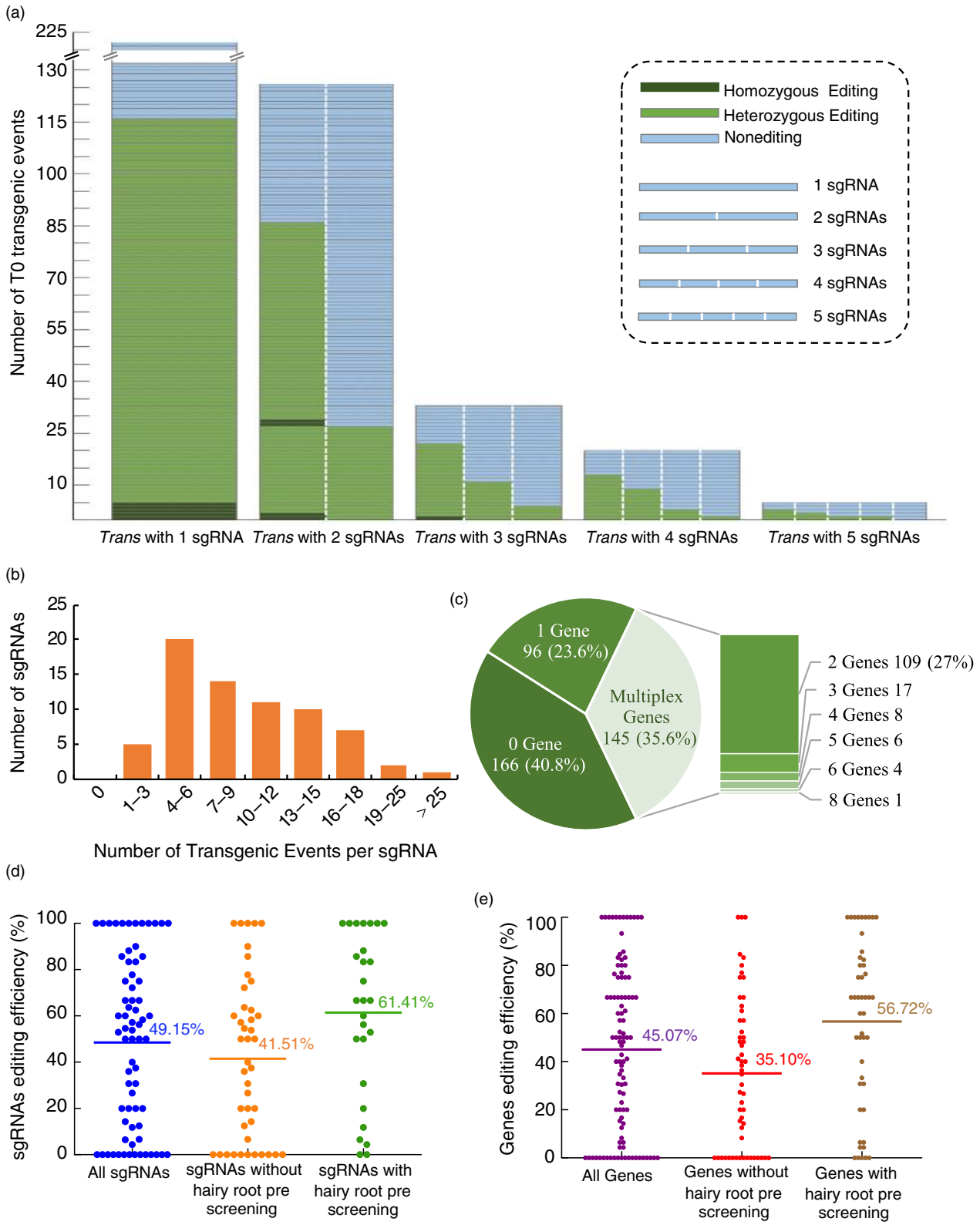

Figure 2 sgRNA distribution and gene editing efficiency in the T0 population. (a) sgRNA and gene editing frequencies in the population. Each line in the bars indicates a transgenic plant. The divided bars indicate the presence of multiple sgRNAs. The dark green colour indicates homozygous mutations of (one) target loci of corresponding sgRNA. The light green colour indicates a heterozygous or biallelic mutation. The blue colour indicates that no mutation was detected. (b) Distribution of transgenic events per sgRNA in the population. (c) Pie chart of number of characterized single or multiple mutants in T0 population. (d) Mutant frequency per sgRNA in the population. (e) Mutant frequency per plant in the population.

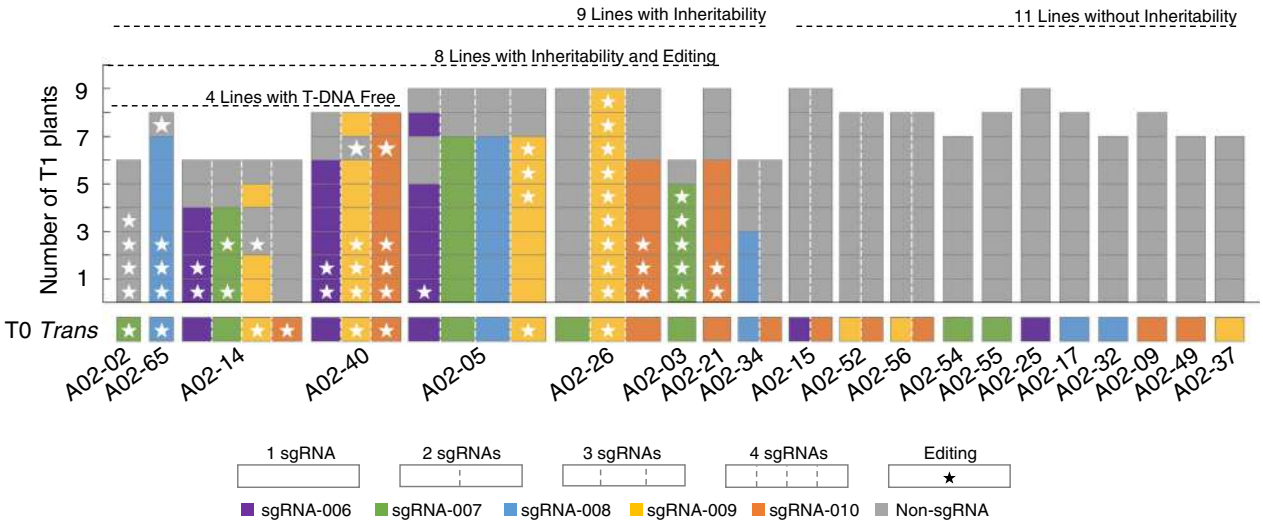

Figure 3 sgRNA inheritance and mutant frequency in T0 and T1 progenies of sublibrary A02. *Y*-axis indicates the number of examined T1 plants. Small blocks indicate examined plants. Purple (sgRNA-006), green (sgRNA-007), blue (sgRNA-008), yellow (sgRNA-009) and orange (sgRNA-010) colours indicate the presence of represented sgRNA in the plant. Grey colour indicates the absence of sgRNA. Stars indicate the presence of gene editing in the target loci.

gmric2 −5 bp) (Figure 4a,b and S6). In addition, all the sgRNAs were segregated in the T2 plants and could be considered 'Cas9 free'. Although sgRNA-006 was also identified in T0 progeny of A2-40 line, it was segregated in A2-40-6-1 and A2-40-6-55 without editing the targeted locus (Glyma.08G246100) (Figure S7). Besides gene editing at the DNA level, the expression of *GmRIC1 and GmRIC2* was significantly repressed in A2-40-6-1 and A2-40-6-55 double mutants, which is apparently due to the action of cellular non-sense-mediated mRNA decay (NMD) machinery (Figure 4c). No mutations were found in the top 10 potential off-target sites for sgRNA-009 and sgRNA-010 (Figure S8).

Hydroponically grown soya bean plants were inoculated, and nodule number was assessed at 30 days after inoculation. An average of 228 nodules were counted on WT plants, compared with 381 (*gmric1/gmric2*$^{+1/-5}$) and 348 (*gmric1/gmric2*$^{-4/-5}$) on *gmric1/gmric2* mutant lines ($P < 0.001$, Student's *t*-test) (Figure 4d,e). Therefore, simultaneous knockout of *GmRIC1* and *GmRIC2* resulted in an increase in nodule number. Nevertheless, the increase in nodule number in *gmric1/gmric2* mutant was moderate, in comparison with the supernodulation phenotype for mutation of *NARK* (nodule autoregulation receptor kinase) gene which encodes for the receptor-like kinase for AON signals (Lin *et al*., 2010). These results indicate that other redundant *CLE* genes might also be involved in AON signalling besides *GmRIC1* and *GmRIC2*.

Identification of *gmrdn1-1/1–2/1–3* triple homozygous mutants with decreased nodule number

In *Medicago truncatula*, *ROOT DETERMINED NODULATION1* (*RDN1*) encodes an unknown protein that may be involved in processing AON signals and negative regulation of nodule numbers in the roots (Schnabel *et al*., 2011). It remains elusive whether *RDN*-like genes in soya bean play similar roles in nodulation. Glyma.02G279600 (*GmRDN1-1*), Glyma.14G035100 (*GmRDN1-2*) and Glyma.20G040500 (*GmRDN1-3*) are soya bean homologs of *RDN1* (Schnabel *et al*., 2011). sgRNA-012 was designed to simultaneously target these three genes. In sublibrary A03, 10 T0 plants were identified to carry sgRNA-012. T1 progeny had 2 lines that carried mis-sense mutations at all loci and 2 independent triple homozygous plants from the A3-8 line (A3-8-6-54: *rdn1-1* + 1 bp, *rdn1-2* −7 bp, *rdn1-3* + 1 bp; A3-8-2-20: *rdn1-1* −14 bp, *rdn1-2* −29 bp and *rdn1-3* −2 bp) were identified in the T2 generation (Figure 5a,b and S9). No other sgRNAs were detected in T1 plants of the A3-8 line. Like in *gmric1/gmric2* mutants, the expression of all three target genes was down-regulated in the *gmrdn1-1/1-2/1-3* plants (Figure 5c), and no off-target editing was observed (Figure S10).

In *M. truncatula*, *rdn1* mutant exhibited a supernodulation phenotype (Schnabel *et al*., 2011). However, *gmrdn1-1/1-2/1-3* triple mutants exhibited decreased nodulation, together with defective vegetative growth (Figure 5d,e and S11). The nodule numbers in *gmrdn1-1/1-2/1-3* triple mutant lines were only 91 (*gmrdn1-2/1-2/1-3*$^{+1/-7/+1}$) and 100 (*gmrdn1-2/1-2/1-3*$^{-14-/29/-2}$) as compared to the wild-type (WT) plants ($P < 0.001$, Student's *t*-test) (Figure 5d,e). The shoot and root length and dry weight of triple mutants showed significant reduction in comparison with the WT plants (Figure S11). These results indicate that GmRDNs are required for normal growth and nodulation in soya bean, which is distinct from their *M. truncatula* homologs.

Discussion

In this study, we developed a pooled CRISPR platform for the generation of a targeted mutagenized soya bean population. The BsaI-mediated one-step cloning system allows easy cloning of sgRNAs, which reduced both labour inputs and the cost associated with the CRISPR library construction. The optimized CRISPR-Cas9 vector and transformation procedure induce a high frequency of mutations. From the CRISPR mutagenesis population consisting of 407 transgenic lines, we identified single or multiplex mutants that covered 102 target genes, and demonstrated the robustness of this platform.

The feasibility of pooled CRISPR-Cas9 mutagenesis screening in soya bean

Many of the crops have complex genome structures, such as polyploidy (wheat, cotton and rapeseed) and paleopolyploidy (maize and soya bean). Genome complexity leads to high gene

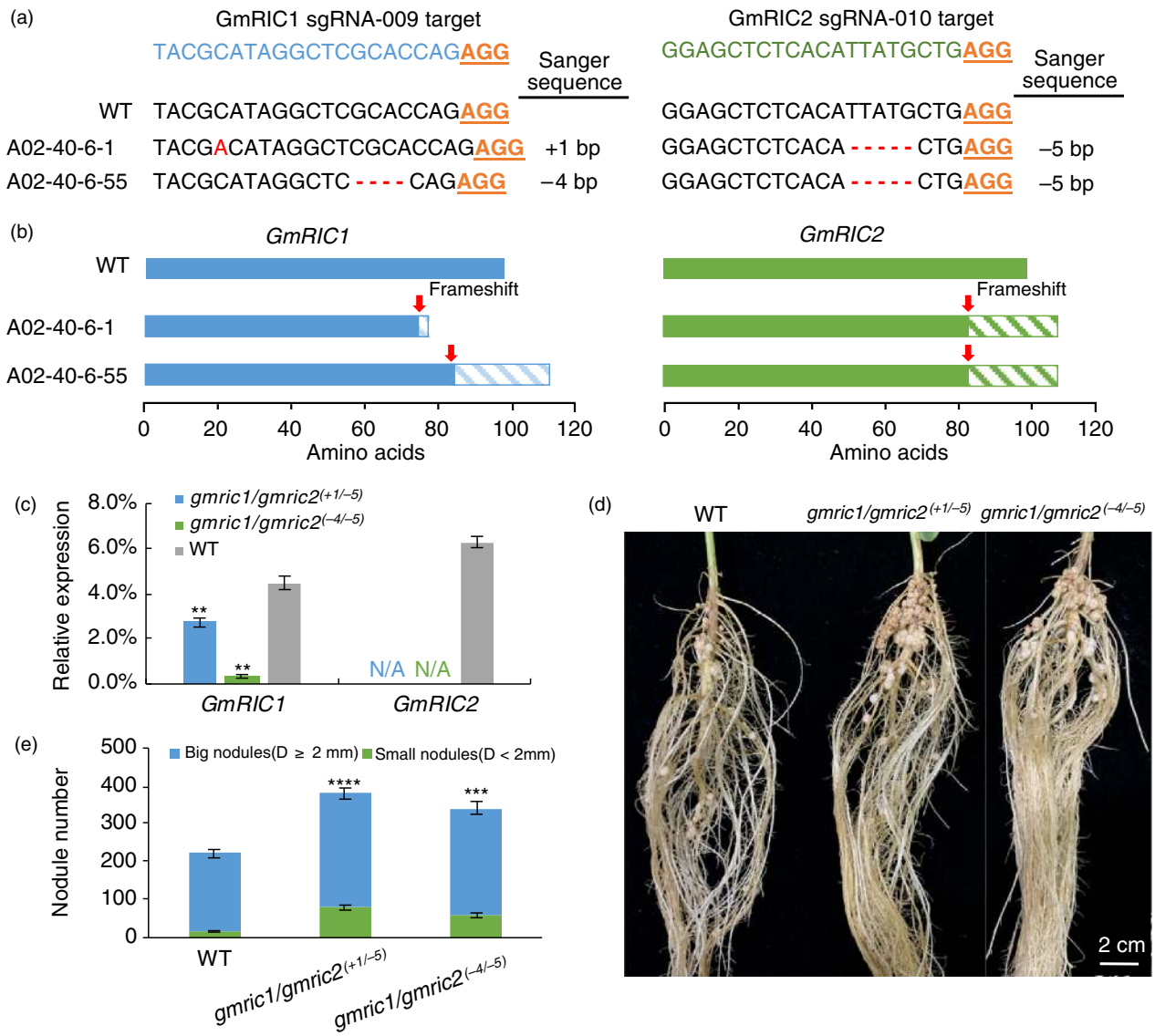

Figure 4 Characterization of *gmric1/gmric2* double mutants with increased nodulation. (a) DNA sequences of A02-40-5 and A02-40-6 at target loci. (b) Predicted protein structures of *gmric1/gmric2* double mutants. The positions of frameshift mutations are indicated by red arrows, and hatched areas represent potentially translated regions until the next predicted termination codon. (c) Expression of *GmRIC1* and *GmRIC2* in the double mutants. (d) Root architectures and (e) nodule numbers at 30 days after inoculation of wild-type and *gmric1/gmric2* T3 plants.

redundancy in these crops which limits the output of mutant screening by forward genetics approaches. A large-scale CRISPR-Cas9 mutagenesis population would be of great value for crop research. Recent studies have generated mutant populations in plants via pooled transformation of CRISPR-Cas9 libraries (Jacobs *et al*., 2017; Lu *et al*., 2017; Meng *et al*., 2017). In tomato, two pooled CRISPR-Cas9 plasmid libraries were constructed and transformed (Jacobs *et al*., 2017); and heritable mutations were obtained in 15 out of 54 genes by single sgRNA constructs and in 15 of 18 genes by triple sgRNA constructs (Jacobs *et al*., 2017). In rice, genome-scale sgRNA libraries covering thousands of genes can be constructed using array synthesized primers and can be used for pooled transformation to create large sizes of mutant populations (Lu *et al*., 2017; Meng *et al*., 2017). The identity of sgRNAs in individual transgenic lines was characterized by barcoded PCR and next-generation sequencing (Lu *et al*., 2017). These technical routes offer significant advantages, such as high coverage of genomic loci, low cost of library construction and high throughput in mutant screening.

In gene editing practices, obtaining multiple independent transgenic events at the target locus is necessary to determine whether the observed phenotype is indeed correlated with the targeted mutation(s) or is caused by off-target editing or spontaneous mutations (Char *et al*., 2017; Puchta, 2017). However, due to the relatively low transformation efficiency of soya bean, the generation of a genome-scale CRISPR population with multiple alleles of target genes is highly impracticable. So far, the transformation efficiency of reported transformation methods could not exceed 8%, which means a single transformation of 300 explants would eventually produce 20–40 T0 transgenic plants (Chen *et al*., 2018). We chose to generate batches of mutant collections using small pooled CRISPR libraries (5–10 sgRNAs). At such a scale, creating small sublibraries is practical with one pooled transformation where multiple alleles per sgRNA

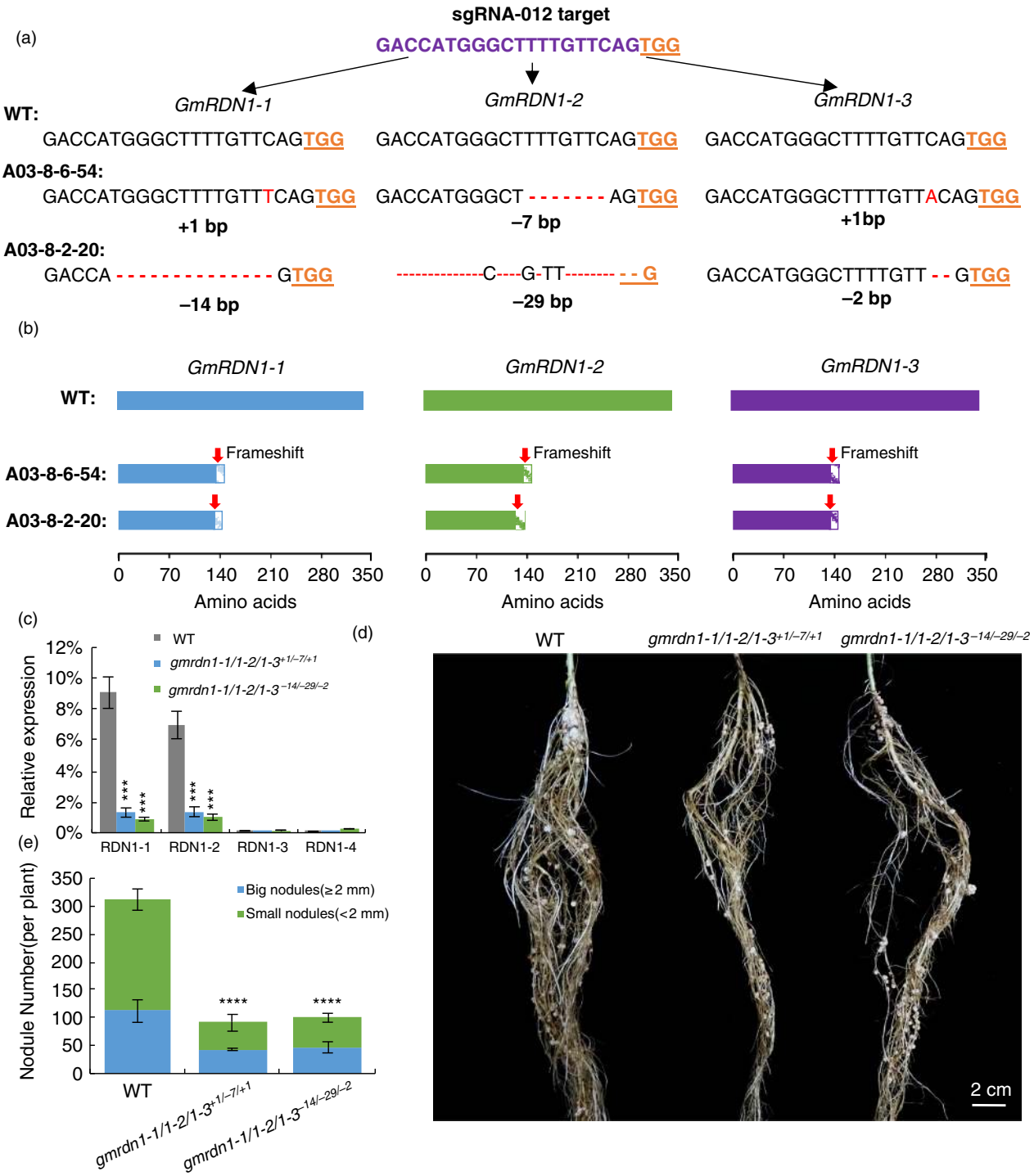

Figure 5 Characterization of *gmrdn1-1/1-2/1-3* triple mutants. (a) DNA sequences of sgRNA-012 at target loci. (b) Predicted protein structures of *gmrdn1-1/1-2/1-3* mutants. The positions of frameshift mutations are indicated by red arrows, and hatched areas represent potentially translated regions until the next predicted termination codon. (c) Expression of *GmRDN1 genes* in the mutants. (d) Root architectures and (e) nodule numbers at 30 days after inoculation.

are obtained to facilitate the genetic manipulation of target genes.

Regardless of the transformation throughput, a small-scale, rather than genome-scale, CRISPR library is preferred for designing multiplex mutant collections. With respect to co-transformation with multiple sgRNAs, different combinations of CRISPR-Cas9 T-DNA that target different genomic loci may be integrated into one transgenic plant. If a small CRISPR library is designed to target a specific set of genes (e.g. a gene clade with high homology or genes with coexpression patterns), multiplex mutations could provide useful genetic diversity to observe additional or synergistic effects. For instance, we obtained transgenic plants

that harboured sgRNAs targeting both, *GmRIC1* and *GmRIC2*, which led to the identification of *gmric1/gmric2* double mutants.

Optimization of CRISPR-Cas9 tools in soya bean and future perspectives

To increase the mutation frequency in the population, high efficiency of gene editing is desired via a high-throughput CRISPR-Cas9 system. The promoters driving the spatiotemporal expression of Cas9 greatly impact gene editing efficiency (Bortesi *et al*., 2016). Here, the *M4* promoter-driven Cas9 exhibited reliable gene editing ability and thus was used as a routine CRISPR-Cas9 tool specialized for soya bean and other leguminous plant species. The design of sgRNA is also critical to the efficiency of CRISPR-Cas9 mutagenesis. Although computational models have been developed to predict sgRNA efficiency based on sequence features, the actual performance of the sgRNAs varies case by case (Bortesi *et al*., 2016; Xu *et al*., 2015). Stable transformation of soya bean is time-consuming, costly and technically demanding. Therefore, sgRNA efficiency assessment and selection are preferred before stable transformation. With respect to our platform, we incorporated an early identification of efficient sgRNAs in the hairy root culture system before stable transformation. Although this incorporation added an extra month to the whole procedure, this step increases the overall performance of the sgRNAs in the mutagenesis population.

Although our current platform works well for pooled CRISPR-Cas9, additional optimizations are still needed. For instance, Cas9 driven by embryonic stage-specific promoters has shown robust editing ability with rare chimeric mutations in several species (Bortesi *et al*., 2016). In maize, compared with the 35S or ZmUBI1 promoter-controlled CRISPR-Cas9 systems, the meiosis-specific dmc1 promoter-controlled (DPC) system is more efficient (Feng *et al*., 2018). In *Arabidopsis thaliana*, Cas9 driven by the anther-specific DD45 promoter (Mao *et al*., 2016), the egg cell-specific EC1.2 promoter (Wang *et al*., 2015) or the meristem-enriched YAO promoter (Yan *et al*., 2015) has been routinely used for high-frequency heritable gene editing. To the best of our knowledge, such promoters have not yet been utilized in soya bean. The characterization of embryonic stage-enriched promoters is, therefore, needed to further increase the editing ability of soya bean CRISPR-Cas9 vectors.

In functional genomic studies, CRISPR-Cas9 target sites are selected based upon the experimental goals. As emphasized in this report, multiplex editing is desired in soya bean to overcome gene duplications. Nevertheless, single gene editing is also needed to validate whether the gene paralogs are functionally redundant or specialized. However, in certain scenarios, NGG PAM targets may not be available in a highly specific genomic region, or the sgRNAs may be ineffective. Vectors consisting of Cas9 variants with expanded PAMs can be developed as alternative tools in such cases (Hu *et al*., 2016, 2018; Kleinstiver *et al*., 2015).

The prospect of transgene-free gene editing of soya bean plants would facilitate field trials and breeding applications. In the current procedure, the identification of transgene cassettes requires BASTA selection and PCR verification, which is time-consuming and labour-intensive. If fluorescent markers could be introduced into the vector, the workload for transgene-free plant screening could be greatly reduced (Gao *et al*., 2016).

Conclusions

In this study, we developed and tested a CRISPR-Cas9 library screening platform in soya bean. Our study demonstrated that targeted single and multiplex mutations could be produced efficiently. The optimized CRISPR-Cas9 platform can be adapted as a screening tool in crops with complex genomes and low transformation efficiency, to empower genetic studies and molecular breeding in these species.

Experimental procedures

Plant material and growth conditions

The soya bean cultivar Huachun6 was used for hairy root and whole-plant transformations. With respect to hairy root transformation, plasmids were individually transformed into *Agrobacterium rhizogenes* strain K599. The hypocotyl and cotyledonary node of WT soya bean were used in accordance with a previously described protocol (Kereszt *et al*., 2007). With respect to whole-plant transformation, plasmids were first individually transformed into *Agrobacterium tumefaciens* strain GV3101. Before plant transformation, the bacterial solutions containing each vector were pooled together at similar amounts. Cotyledonary node transformation was then performed according to a previously reported protocol (Song *et al*., 2013). The transgenic soya bean plants were grown outdoors in the summer and in a growth chamber in the winter under the following conditions: light intensity of 400 μmol photons/m^2/s, a regimen of 13-h/26 °C day, and 11-h/24 °C night, and a relative humidity of 65%.

For hydroponic culture and inoculation, soya bean plants were grown in a growth chamber at the following conditions: light intensity of 400 μmol photons/m^2/s^1, 13-h/26 °C day, and 11-h/24 °C night regime, and 65% humidity. Seven-day-old seedlings were inoculated with the rhizobium strain *Bradyrhizobium* sp. *BXYD3* and transferred to hydroponic culture in N-free medium as previously described (Li *et al*., 2012). Plant growth and nodulation were measured at 30 days after culture, and more than 3 biological replicates were used for each measurement.

Construction of pGES201 vectors

The binary vector backbone was a pCAMBIA1300 (Cambia, GPO Box 3200, Canberra, Australia) vector. HygR was initially replaced with the bar marker gene through restriction digestion and ligation between the PspXI and AatII sites. The CRISPR-Cas9 expression cassettes were assembled with overlapping PCR and NEBuilder HiFi DNA Assembly Master Mix (NEB, Ipswich, MA). The endogenous promoter sequences (*pM4*, *pM8* and *GmU6*) were cloned from the genomic DNA of Huachun6. The gRNA scaffold and Cas9 were fused to a nuclear localization signal, and 3xFLAG was amplified from the BGK03 vector. To enable the use of BsaI sites in sgRNA assembly, an extra BsaI site in the pCAMBIA1300 backbone was disrupted by introducing a sense mutation by overlapping PCR (Figure S1).

Searching for single and multiplex sgRNAs

For the sgRNA design of target genes, the guide RNA spacer sequences with NGG PAM were computationally identified based on Wm82.a2 genomic sequences. The candidate sgRNAs were then filtered using the following criteria; first, sgRNA sequences should localize to exon regions of target genes; second, the GC content of each sgRNA should be between 25% and 80%; and third, the sgRNAs should target the region near the 5′ half of the genes. Last, single gene-specific sgRNAs should contain at least one mismatch at the 3′ end from potential off-targets, while multiplex sgRNAs should match perfectly with at least two target genes. Off-target sites were predicted by the Cas-OFFinder online tool (http://www.rgenome.net/cas-offinder/).

sgRNA cloning

With respect to vector cloning, sgRNA fragments were produced by annealing complementary oligonucleotides. To make the annealed products compatible with BsaI-digested vectors, additional GGATT nucleotides were added to the 5′ end of the forward primers, and AAAC and A were added both to the 5′ and 3′ ends of the reverse primers, respectively (Table S5). The synthesized primers of each sgRNA were mixed together [50 pM of forward primer, 50 pM of reverse primer, 50 mM NaCl and 10 mM Tris (pH 7.5–8.0) in a 50 μL reaction] and then annealed by ramping the temperature from 94 °C at 0.1 °C/s to 16 °C. Ligation was performed using 5 ng of annealed oligomers and 5 ng of BsaI-digested pGES201 plasmids with a T4 DNA ligation kit (Takara, Kyoto, Japan) according to the manufacturer's instructions. After *E. coli* transformation, positive clones were identified by colony PCR and Sanger sequencing of the extracted plasmids.

Examination of CRISPR-Cas9-induced mutations in plants

With respect to genomic DNA extraction, individual hairy roots longer than 5 cm or leaf discs of transgenic plants were collected. A PCR-based screening was performed for the verification of sgRNA-mediated editing. With respect to the detection of each sgRNA, a sgRNA-specific primer (forward primer used for sgRNA annealing) was used along with a universal reverse primer named SSP-R (targeting the plasmid backbone) for PCR (Table S5). PCR for each sgRNA was performed individually, and positive PCR results indicated the presence of sgRNA in the transformants. After verification of the sgRNA integration, gene editing in the corresponding target regions was examined by both PCR and sequencing. The PCR primers were designed specifically to amplify the target region instead of the paralogous region. After PCR, the 300–1000 bp products were purified for Sanger sequencing; the sequencing results were subjected to BLAST analysis to detect potential mutations. Assessment of the exact genotype can be difficult in heterozygous and biallelic mutants, especially when transgenic chimera occurs in hairy roots or in stable transgenic T0 plants. Given the genotype complexity in the PCR products, we considered the following cases as representatives of the occurrence of gene editing. (i) Compared to the references sequence, the sequences containing indel mutations were considered homozygous mutants. (ii) Sequences called 'N' due to two predominant peaks starting at the sgRNA target region were considered heterozygous or biallelic mutants. (iii) In some cases, sequence reads were similar to WT; however, the chromatograms exhibited strong noise starting in the sgRNA target region. If the noise peaks were higher than 1/4th of the WT peaks, they were also considered heterozygous mutations in the PCR products (Figure S4). For plants containing multiple sgRNAs, all possible target loci were examined. For multiplex sgRNAs, if one of the target loci was edited, the sgRNA was considered active.

Raw sequencing and SSP data reported in this work have been deposited to Zenodo database (https://doi.org/10.5281/zenodo.3228783).

Acknowledgements

This work was supported by funds from the National Key Research and Development Program of China (2016YFD0100700) and the National Science Foundation of China (31772379) to Y.G. M.B. and J.Y. were supported by the K+S scholarship from the International Magnesium Institute. The authors are grateful to Dr. Yongqing Yang for providing Huachun6-seeds and Donghai Zheng for greenhouse support and plant maintenance.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Author contributions

Y.G. and M.B. designed the studies and workflow. M.B. designed and constructed the CRISPR libraries. J.Y. analysed the genotypes of the T1 progeny and characterized the multiplex mutants. H.K. and S.S. performed the soya bean transformations. P.G. and D.W. carried out pGES201 vector construction and optimization. M.B. and J.Y. analysed genotypes of the hairy roots and the T0 population. D.W. searched for the sgRNAs in the soybean genome. All the authors contributed to the writing of the manuscript.

References

Bolon, Y.-T., Haun, W.J., Xu, W.W., Grant, D., Stacey, M.G., Nelson, R.T., Gerhardt, D.J. *et al*. (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiol.* **156**, 240–253.

Bortesi, L., Zhu, C., Zischewski, J., Perez, L., Bassié, L., Nadi, R., Forni, G. *et al*. (2016) Patterns of CRISPR/Cas9 activity in plants, animals and microbes. *Plant Biotechnol. J.* **14**, 2203–2216.

Bull, S.E., Seung, D., Chanez, C., Mehta, D., Kuon, J.E., Truernit, E., Hochmuth, A. *et al*. (2018) Accelerated ex situ breeding of GBSS- and PTST1-edited cassava for modified starch. *Sci. Adv.* **4**, eaat6086.

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

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

Cai, Y., Chen, L., Sun, S., Wu, C., Yao, W., Jiang, B., Han, T. *et al*. (2018b) CRISPR/Cas9-mediated deletion of large genomic fragments in soybean. *Int. J. Mol. Sci.* **19**, 3835.

Cai, Y., Wang, L., Chen, L., Wu, T., Liu, L., Sun, S., Wu, C. *et al*. (2019) Mutagenesis of GmFT2a and GmFT5a mediated by CRISPR/Cas9 contributes for expanding the regional adaptability of soybean. *Plant Biotechnol. J.* https://doi.org/10.1111/pbi.13199

Campbell, B.W. and Stupar, R.M. (2017) Soybean (Glycine max) mutant and germplasm resources: current status and future prospects. *Curr. Protoc. Plant Biol.* **1**, 307–327.

Char, S.N., Neelakandan, A.K., Nahampun, H., Frame, B., Main, M., Spalding, M.H., Becraft, P.W. *et al*. (2017) An Agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. *Plant Biotechnol. J.* **15**, 257–268.

Chen, L., Cai, Y., Liu, X., Yao, W., Guo, C., Sun, S., Wu, C. *et al*. (2018) Improvement of soybean agrobacterium-mediated transformation efficiency by adding glutamine and asparagine into the culture media. *Int. J. Mol. Sci.* **19**, 3039.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D. *et al*. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.

Cui, Y., Barampuram, S., Stacey, M.G., Hancock, C.N., Findley, S., Mathieu, M., Zhang, Z. *et al*. (2013) Tnt1 retrotransposon mutagenesis: a tool for soybean functional genomics. *Plant Physiol.* **161**, 36–47.

Curtin, S.J., Xiong, Y., Michno, J.M., Campbell, B.W., Stec, A.O., Čermák, T., Starker, C. *et al*. (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.

Du, H., Zeng, X., Zhao, M., Cui, X., Wang, Q., Yang, H., Cheng, H. *et al*. (2016) Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *J. Biotechnol.* **217**, 90–97.

Espina, M.J., Ahmed, C.M.S., Bernardini, A., Adeleke, E., Yadegari, Z., Arelli, P., Pantalone, V. *et al*. (2018) Development and phenotypic screening of an ethyl methane sulfonate mutant population in soybean. *Front. Plant Sci.* **9**, 394.

Fauser, F., Schiml, S. and 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.

Feng, C., Su, H., Bai, H., Wang, R., Liu, Y., Guo, X., Liu, C. *et al*. (2018) High-efficiency genome editing using a dmc1 promoter-controlled CRISPR/Cas9 system in maize. *Plant Biotechnol. J.* **16**, 1848–1857.

Ferguson, B.J., Li, D., Hastwell, A.H., Reid, D.E., Li, Y., Jackson, S.A. and Gresshoff, P.M. (2014) The soybean (Glycine max) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (Phaseolus vulgaris), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnol. J.* **12**, 1085–1097.

Gao, X., Chen, J., Dai, X., Zhang, D. and Zhao, Y. (2016) An effective strategy for reliably isolating heritable and Cas9-Free arabidopsis mutants generated by CRISPR/Cas9-mediated genome editing. *Plant Physiol.* **171**, 1794–1800.

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C.V., Sánchez-León, S., Baltes, N.J. *et al*. (2017) High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J.* **89**, 1251–1262.

Hu, X., Wang, C., Fu, Y., Liu, Q., Jiao, X. and Wang, K. (2016) Expanding the range of CRISPR/Cas9 genome editing in rice. *Mol. Plant*, **9**, 943–945.

Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M. *et al*. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, **556**, 57–63.

Jacobs, T.B., LaFayette, P.R., Schmitz, R.J. and Parrott, W.A. (2015) Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* **15**, 16.

Jacobs, T.B., Zhang, N., Patel, D. and Martin, G.B. (2017) Generation of a collection of mutant tomato lines using pooled CRISPR libraries. *Plant Physiol.* **174**, 2023–2037.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.

Kanazashi, Y., Hirose, A., Takahashi, I., Mikami, M., Endo, M., Hirose, S., Toki, S. *et al*. (2018) Simultaneous site-directed mutagenesis of duplicated loci in soybean using a single guide RNA. *Plant Cell Rep.* **37**, 553–563.

Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C.D., Nontachaiyapoom, S., Kinkema, M. and Gresshoff, P.M. (2007) Agrobacterium rhizogenes-mediated transformation of soybean to study root biology. *Nat. Protoc.* **2**, 948–952.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P. *et al*. (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, **523**, 481–485.

Li, C., Gui, S., Yang, T., Walk, T., Wang, X. and Liao, H. (2012) Identification of soybean purple acid phosphatase genes and their expression responses to phosphorus availability and symbiosis. *Ann Bot* **109**, 275–285.

Li, Z., Liu, Z.B., Xing, A., Moon, B.P., Koellhoffer, J.P., Huang, L., Ward, R.T. *et al*. (2015) Cas9-guide RNA directed genome editing in soybean. *Plant Physiol.* **169**, 960–970.

Li, Z., Jiang, L., Ma, Y., Wei, Z., Hong, H., Liu, Z., Lei, J. *et al*. (2017) Development and utilization of a new chemically-induced soybean library with a high mutation density. *J. Integr. Plant Biol.* **59**, 60–74.

Li, C., Hao, M., Wang, W., Wang, H., Chen, F., Chu, W., Zhang, B. *et al*. (2018) An efficient CRISPR/Cas9 platform for rapidly generating simultaneous mutagenesis of multiple gene homoeologs in allotetraploid oilseed rape. *Front. Plant Sci.* **9**, 442.

Lim, C.W., Lee, Y.W. and Hwang, C.H. (2011) Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. *Plant Cell Physiol.* **52**, 1613–1627.

Lin, Y.H., Ferguson, B.J., Kereszt, A. and Gresshoff, P.M. (2010) Suppression of hypernodulation in soybean by a leaf-extracted, NARK- and Nod factor-dependent, low molecular mass fraction. *New Phytol.* **185**, 1074–1086.

Liu, W., Jiang, B., Ma, L., Zhang, S., Zhai, H., Xu, X., Hou, W. *et al*. (2018) Functional diversification of Flowering Locus T homologs in soybean: GmFT1a and GmFT2a/5a have opposite roles in controlling flowering and maturation. *New Phytol.* **217**, 1335–1345.

Lu, Y., Ye, X., Guo, R., Huang, J., Wang, W., Tang, J., Tan, L. *et al*. (2017) Genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. *Mol. Plant*, **10**, 1242–1245.

Mao, Y., Zhang, Z., Feng, Z., Wei, P., Zhang, H., Botella, J.R. and Zhu, J.-K. (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.

Mathieu, M., Winters, E.K., Kong, F., Wan, J., Wang, S., Eckert, H., Luth, D. *et al*. (2009) Establishment of a soybean (*Glycine max Merr. L*) transposon-based mutagenesis repository. *Planta*, **229**, 279–289.

Meng, X., Yu, H., Zhang, Y., Zhuang, F., Song, X., Gao, S., Gao, C. *et al*. (2017) Construction of a genome-wide mutant library in rice using CRISPR/Cas9. *Mol. Plant*, **10**, 1238–1241.

Michno, J.-M., Wang, X., Liu, J., Curtin, S.J., Kono, T.J.Y. 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.

Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y. and Kawaguchi, M. (2013) Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nat. Commun.* **4**, 2191.

O'Rourke, J.A., Graham, M.A. and Whitham, S.A. (2017) Soybean functional genomics: bridging the genotype-to-phenotype gap. In *The Soybean Genome* (Nguyen, H.T. and Bhattacharyya, M.K., eds), pp. 151–170. Cham: Springer International Publishing.

Parrott, W.A., Hoffman, L.M., Hildebrand, D.F., Williams, E.G. and Collins, G.B. (1989) Recovery of primary transformants of soybean. *Plant Cell Rep.* **7**, 615–617.

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

Reid, D.E., Ferguson, B.J. and Gresshoff, P.M. (2011) Inoculation- and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Mol. Plant Microbe Interact.* **24**, 606–618.

Sánchez-León, S., Gil-Humanes, J., Ozuna, C.V., Giménez, M.J., Sousa, C., Voytas, D.F. and Barro, F. (2018) Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol. J.* **16**, 902–910.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L. *et al*. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Schnabel, E.L., Kassaw, T.K., Smith, L.S., Marsh, J.F., Oldroyd, G.E., Long, S.R. and Frugoli, J.A. (2011) The ROOT DETERMINED NODULATION1 gene regulates nodule number in roots of medicago truncatula and defines a highly conserved, uncharacterized plant gene family. *Plant Physiol.* **157**, 328–340.

Singh, R.J. and Hymowitz, T. (1988) The genomic relationship between *Glycine max (L.) Merr.* and *G. soja Sieb.* and *Zucc.* as revealed by pachytene chromosome analysis. *TAG. Theoretical and applied genetics. Theor. Appl. Genet.* **76**, 705–711.

Song, S., Hou, W., Godo, I., Wu, C., Yu, Y., Matityahu, I., Hacham, Y. *et al*. (2013) Soybean seeds expressing feedback-insensitive cystathionine gamma-synthase exhibit a higher content of methionine. *J. Exp. Bot.* **64**, 1917–1926.

Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H. and Xi, Y. (2015) Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Sci. Rep.* **5**, 10342.

Tsuda, M., Kaga, A., Anai, T., Shimizu, T., Sayama, T., Takagi, K., Machita, K. *et al*. (2015) Construction of a high-density mutant library in soybean and development of a mutant retrieval method using amplicon sequencing. *BMC Genom.* **16**, 1014.

Wang, Z.-P., Xing, H.-L., Dong, L., Zhang, H.-Y., Han, C.-Y., Wang, X.-C. and Chen, Q.-J. (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.

Wang, P., Zhang, J., Sun, L., Ma, Y., Xu, J., Liang, S., Deng, J. *et al*. (2018) High efficient multisites genome editing in allotetraploid cotton (Gossypium hirsutum) using CRISPR/Cas9 system. *Plant Biotechnol. J.* **16**, 137–150.

Xu, H., Xiao, T., Chen, C.-H., Li, W., Meyer, C.A., Wu, Q., Wu, D. *et al*. (2015) Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147–1157.

Yan, L., Wei, S., Wu, Y., Hu, R., Li, H., Yang, W. and Xie, Q. (2015) High-efficiency genome editing in *Arabidopsis* using *YAO* promoter-driven CRISPR/Cas9 system. *Mol. Plant*, **8**, 1820–1823.

Zhang, N., McHale, L.K. and Finer, J.J. (2015) Isolation and characterization of "GmScream" promoters that regulate highly expressing soybean (Glycine max Merr.) genes. *Plant Sci.* **241**, 189–198.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 The sequence of pGES201 vector.
Figure S2 Comparison of mutation frequencies using Cas9 driven by pM4 and pM8 promoters.
Figure S3 Principle of sgRNA specific PCR (SSP).
Figure S4 Representative chromatograms for gene editing.
Figure S5 Example of SSP result indicating inheritable and non-inheritable sgRNAs.
Figure S6 Predicted protein sequences of GmRIC1 and GmRIC2 in *gmric1/gmric2* double mutants.
Figure S7 Segregation of sgRNA006 and sgRNA009 from *gmric1/gmric2* double mutants.
Figure S8 Off-target examination in *gmric1/gmric2* double mutants.
Figure S9 Predicted protein sequences of GmRDN1-1. GmRDN1-2 and GmRDN1-3 in *gmrdn1-1/1-2/1-3* triple mutants.
Figure S10 Off-target examination in *gmrdn1-1/1-2/1-3* triple mutant.
Figure S11 The vegetative growth phenotype in *gmrdn1-1/1-2/1-3* triple mutant.
Table S1 List of targeted genes.
Table S2 List of all sub-libraries, sgRNAs with its targeted genes and Annotation of the targeted genes.
Table S3 sgRNA distribution and gene editing in T0 transgenic lines.
Table S4 Gene editing in individual T1 plants of sub-library A02.
Table S5 List of primers used in this study.

