AN EFFICIENT TRANSIENT ASSAY FOR CRISPR CAS9 SYSTEM DELIVERING TARGETED MUTATION USING SYNTHETIC OLIGO SgRNA IN SOYBEAN (GLYCINE MAX)

NOOR AL AMIN¹, NAVEED AHMAD², WU NAN¹, FU XIUMING¹, WANG NAN¹, BO XIAOXUE¹, MA TONG², AND WANG PIWU²

¹Plant Biotechnology Center, Agromomy College, Jilin Agricultural University, Jilin, China
²College of Life Sciences, Jilin Agricultural University, Jilin, China
*Corresponding author’s email: Peiww@163.com; Tel: +0431-84532908

Abstract

In recent years CRISPR Cas9 system has been proved as a versatile genetic tool for efficient target mutation in a variety of plants. (CRISPR)-associated Cas9 Nuclease system can efficiently incorporate high frequency targeted mutation wherein Cas9 act as endonuclease coupled with an oligo-sgRNA sequence. This RNA guided nuclease complex causes a Cas9-directed nick production in nuclear DNA in the form of double-strand breaks (DSBs). Importantly, our findings revealed various factors affecting the specificity and activity of small guide RNAs (sgRNAs), casting a new light on flexibility and efficiency of sgRNA encountering with targeted mutagenesis in various agronomical crops. We demonstrated an efficient protocol and online resources used for the optimization of best possible sgRNA designing and it’s (PAM) Protospacer Adjacent Motif to achieve targeted mutation in FAD2-1A. Collectively our findings confirmed an efficient transient expression assay of FAD2-1A gene in isolated protoplasts via PEG mediated transformation. High throughput protein modeling revealed the successful integration of targeted mutation induced by Small guide RNA induced CRISPR Cas9 system. Preferably, these findings can be replicated in designing multiple novel genome models to facilitate efficient target mutagenesis and gene expression approaches in several other higher plants.

Key words: Targeted mutagenesis, CRISPR Cas9 system, PEG-mediated transformation, SgRNA,

Introduction

Soybean (Glycine max) is an important legume crop and is well known for providing high quality oil that is extremely rich in unsaturated fatty acid and other dietary minerals. Soybean oil is regarded as beneficial for human health because it contains significantly less amount of desirable fatty acids (24%) compared to other competing oils (Jinek et al., 2012; Zhang et al., 2017). In recent times with the advance genetic engineering tools and targeted mutation soybean research have been marked essential to inquire new advancement in soybean research. The CRISPR induced Cas9 system has considerably been proved as an ideal approach to unfolding the mechanism of targeted mutagenesis in highly duplicated genomes including soybean. This system allows high frequency mutation within gene of interest without interrupting the rest of genome. The principle of CRISPR Cas9 system had been adapted from the adaptive immune system found in prokaryotes and archaea bacteria (Bhay et al., 2011).

Type II CRISPR Cas9 system inherited from Streptococcus pyogenes has been reported most generally for site-directed mutagenesis in a variety of plants. Cas9 endonuclease in association with a small guide RNA (sgRNA) forms a complex of RNA guide Nuclease harboring a DSB within the specified genomic sequence. Small guide RNA consists of a 19 base pair short sequence complementary to its target DNA with a PAM, (3-NGG-5 motif) at downstream of the target site which acts as a screening agent for target cleavage. Once the target sequence is recognized by sgRNA motifs the two domains of Cas9 endonuclease activates and causes a nick production inside target genome. The process of integrating targeted cleavage was previously conducted in a variety of plants and mammals by introducing hairpin loop precisely at 3’end of sgRNA occurred by activating its Cas9 nuclease partner for site-specific mutagenesis (Jinek et al., 2012). The occurrence of DSB in target genome induced by CRISPR Cas9 system can simultaneously trigger two well-known mechanisms of DNA repair pathways: non homologous end joining (NHEJ) and homologous recombination (HR) (Puchta & Fauser, 2014). Among these two independent pathways homologous direct recombination (HDR) proved to precisely repair the DSB caused by CRISPR Cas9 system by providing exogenous homologous donor DNA as template. However, non homologous end joining (NHEJ) mechanism can lead to a small chromosomal loss in the form of deletion and insertion which manipulate gene expression and regulation of the target gene (Fan et al., 2015; Sun et al, 2015).

Recent practical developments have revealed that CRISPR Cas9 system has been successfully characterized for genome site specific mutation in various plants such as maize (Li et al., 2014), rice (Miao et al., 2013; Zhang et al, 2014), Arabidopsis (Feng et al., 2014), wheat (Wang et al., 2014), tomato (Lin et al., 2016), Soybean (Li et al., 2015) and tobacco. Unlike predecessors genome editing technology including ZFN, and TALENs, the CRISPR Cas9 is more simple, easier to design, cost-effective and efficient genome editing tool with rare off target mutagenesis occurrence (Xu et al., 2015). Hence it is recommended for incorporating site directed mutation in order to facilitate novel checkpoints in genome engineering approaches (Rinaldo & Ayliffe 2015; Bortes & Fischer 2015).

In our study, we have demonstrated an easy accessed downloadable online tool CRISPR P2.0 for the designing and optimization of sgRNA. Multiple sets of (19-nt) CRISPR sgRNA were designed for the FAD2-1A locus.
However it is essential to keep in mind that there are several factors affecting CRISPR Cas9 activity, we suggest while generating sgRNA for a target genome one can consider critically the mismatch scores, RNA scaffold, PAM type, sgRNA sequence size, and snoRNA promoter which may influence your targeted mutation (Cradick et al., 2013; Hsu et al., 2013). Furthermore, A single binary vector was successfully constructed after integrating sgRNA and Cas9 system and was successfully transformed in soybean protoplast through PEG transformation resulting in targeted disruption of FAD2-1A gene in soybean. There is no such other report published likewise on the designing of an efficient sgRNA for FAD2-1A gene using CRISPR Cas9 system in soybean aim to be rich in high oleic acid content.

Material and Methods

CRISPR-P: A widely known web tool is recruited for sgRNA designing of CRISPR system Cas9 in accordance with our target genome (FAD2-1A in soybean). It can be accessed through any Mac OS system and windows system. FAD21-A searching mode has been set efficiently in order to investigate both on target and off-target consequences within coding regions. First of all, we need to extract the complete genomic sequence of FAD2-1A (FASTA file) which is openly available on several genomic databases. [FAD2-1A whole genomic sequence: GenBank: accession number NC_016097. Sequence homology and possible target sites in exonic regions were determined at [Ensembl: GTF format] followed by a BLASTN against Phycomitrella patens which have been well studied as a model plant for targeted gene knockouts approaches and other physiological characters. Percentage identity was shown between coding regions with best possible E. value.

Afterward, ECMA (European Computer Manufacturers Association) standardized program named SMS [Sequence Manipulation Suite] was applied to filter any user-provided sequence by eliminating non-DNA characters and gaps. The principle of SMS program works on the basis of input provided by the user, after reading and screening the sequence resulted in a filtered outputs facilitating sequence smarting for further potential uses. SMS can also be used as a significant tool during conversion of formats as well as Protein Isoelectric Point analysis, EMBL Trans Extractor analysis and Pair wise Alignment.

Finally, the most important step while designing a routine sgRNA sequence for a CRISPR Cas9 system was determined with [CRISPR P2] where the assembly of all possible sgRNA sequences of FAD2-1A target sequence occurred. The selection of an appropriate filtered genomic region (30-5,000 bp), along with other parameters likewise RNA scaffold, type of PAM, length of guide sequence (spacer) and snoRNA promoter should be kept in mind. After submitting this thoroughly optimized query to the system, a visual of complete sequence map was created for FAD2-1A sequence. Out of all possible sgRNA sequences select the one comprising maximal score. Other factors such as GC content, potential sited for endonucleases and off-targeting scores must be kept in mind before designing small guide RNA sequence. Lastly, select 19 bp short sgRNA sequences in the coding region excluding PAM motifs and integrate it with the CRISPR construction kit Site (BGK041-Soybean) where the designing of oligos were generated. We ordered the aforementioned oligos from the company which will then correspond to our target genome. The ligation of synthetic oligos into CRISPR (BGK041) vector was followed by manufactures protocol. We efficiently transformed CRISPR (BGK041) vector into Agrobacterium tumafeciens (EHA105 strain) by heat and shock method followed by overnight incubation at 28°C and positive clones were observed on YEP+K plates and confirmed with half colony method.

Transient gene expression assay: Protoplast isolation was performed from the fresh and young leaves of soybean according to Yoo et al., (2007) method with slight modifications. Approximately 50 fresh leaves were subjected to small pieces which were subsequently transferred into digestion solution. Vacuum infiltration was carried out in full dark condition with a help of vacuum pump with 6 hours digestion and 30 rpm agitation at 16−20mmHg. Finally, the soybean protoplasts were successfully resuspended in a solution named MGG solution for conducting the transformation of bacterial plasmids containing SgRNA-FAD2-1A-CRISPR Cas9 vector. The efficient transformation of plasmid was carried out in the aforementioned protoplasts induced by PEG method as suggested by (Yoo et al., 2007). The confirmation was done with PCR using FAD2-1A gene primers (Table 2). The positive bands were cloned in the T1 vector for sequence analysis. After sequencing high throughput bioinformatics analysis such as Protein modeling via (https://web.expasy.org/translate/) reveals further validation of the targeted mutations.

Results and Discussion

Glycine max FAD2-1A sequence has been extracted from GenBank (NC_016097; FASTA format) covering 1873 bp length. A query of BLASTN against ENSEMBL PLANTS which in results showed maximum sequence homology (100%) within exonic regions at position 10: +49418203. The whole sequence of Soybean FAD2-1A (1873bp) was subjected to BLASTN with Phycomitrella patens genome where sequence similarities were conducted within coding regions of FAD2-1A (Fig. 1). The extracted sequence was further processed via SMS which was an ECMA regulated program in order to filter user provided sequence by removing non-DNA codes, gaps, dots and numerical from the aligned sequence. It is extremely essential to process the target sequence with SMS before designing of the sgRNA, by ignoring this step the final synthetic oligos may lead to unwanted and nonspecific outcomes in sgRNA sequence. The filtered sequence was further entered into CRISPR P2 design for sgRNA synthesis.
Designing of potential sgRNA in a given genome: CRISPR-P 2 design consists of 49 genome databases of plants which are widely used in CRISPR Cas9 induced genetic mutation. Before starting with CRISPR-P2 audience should keep in mind to ensure various factors pre subjecting our target sequence to CRISPR-P2 that can influence sgRNA activity and specificity. In order to carry out all potential sites of sgRNA for FAD2-1A gene of soybean CRISPR-P followed the two steps method.

Step 1: Provide the filtered BLASTN target sequence in sequence box, selection of the target genome (Glycine max L.) and other factors such as PAM (NGG), RNA scaffold (GUUUAGACGCUAAAUAGCAAGUUA A AA UAA GGUUAGUCCGGUUAUCAACUUUGAAAGUUGGAC CGAGUGCGCGUAAA), sequence length (19bp), snoRNA promoter (U6) must be kept in mind which can cause divergence in the route of our expected sgRNA sequence (Fig. 2). We can also design a more specific sgRNA sequence for any target sequence by providing gene locus, or desired chromosome position but these are not mandatory only complete genomic sequence is enough to design a potential sgRNA sequence.

Step 2: Post submission of step 1 a complete map of the desired genome and all possible sgRNA sequences with sequence length, GC content and target score were generated on the screen (Fig. 3A). We choose a 19 bp short sequence of sgRNA excluding PAM sequence within coding regions for final oligos synthesis (Fig. 3B).

Fig. 1. Sequence homology BLASTN query of FAD2-1A aligned with Physcomitrella patens genome. A total of 1873 bp score were recorded at Glycine_max v2.0:10:50013288:50015760:1. Similarities were shown as 100% with 0 E value in coding regions. The percentage of Exons represented in Red, Short target sequence (19bp) within target sequence indicated in green, multiple alignments of additional scores indicated with blue and grey.
Fig. 2. Various factors affecting the activity of sgRNA including Protospacer Adjacent Motif (5'NGG3'), SnoRNA promoter (U6), RNA scaffold (GUUUAGACCUAGAAGCAAGAUUUAAUAGGCUGUCUAUCAACUUGAAAAAGUGGCACCCAGUGCCGUUUU) and length of Short guide sequence and Sequence (19bp) of the targeted genome. Choose your Target genome however locus Taq and Position can be automatically optimized after selection target genome. The consideration of these parameters can drastically change the activity of sgRNA during designing of CRISPR Cas9 system.

CRISPR vector construction kit suite for oligos synthesis: Once the tailoring of sgRNA for FAD2-1A gene of soybean is completed, their integration in CRISPR Cas9 (BGK041) plasmid requires a forward and reverse oligos which correspond to our target sequence. The resubmission of the previously guided sequence was performed using BGK041-CRISPR/Cas9 Vector construction kit suite. Up oligos (Forward) constituting 19bp short sequence (5'-GTATTGATGGAGCAACCAAT-3') was obtained. This short sgRNA forward oligo should be linked just after 3'AGG (PAM) located at the end of target regions. (19bp short oligo does not include PAM sequence) (Fig. 4). Low oligo (Reverse) is the exact complementary copy of up oligo (Forward) sequence (5'-ATTGGTAGCCATCAATAC-3'). In order to minimize the chances of off-target mutations, it is pragmatic to analyze approximately three mismatches in other related sequences preferably located at PAM site or nearly adjacent to PAM site (Hsu et al., 2013).

SgRNA-Oligos and CRISPR-Cas9 cloning: The preparation of synthetic oligos and its ligation into BGK041-CRISPR Cas9 vector requires a thermocycler which can be set for a heat and cold shock cycle followed by an initial step of 3 minutes heat (95°C) and shock (20°C) respectively. It is necessary to dissolve the fresh oligos in 10µM nuclease-free water before ligation in BGK041-CRISPR Cas9. The previous experiment was followed according to given manufacturer's protocol. The binary vector was constructed using Agrobacterium tumefaciens (EHA105 strain) by heat and shock method followed by overnight incubation at 28°C. Positive colonies were detected on YEP kanamycin resistant (Table 1) and further confirmation was performed with PCR using BGK041-CRISPR Cas9 primers (Table 2; Fig. 5). The PCR System includes 30 cycles with following conditions: 95°C/45s, 55°C/35s and 72°C/1m.
CRISPR Cas9 NUCLEASE ACTIVITY DRIVEN BY SMALL GUIDE RNA

Fig. 3A. A Complete genome map of all possible sgRNA sequences in our target FAD2-1A gene using CRISPR P tool. Yellow color described the presence of maximum results i.e. s > 0.7 score and green color represents the minimum scoring map for FAD2-1A gene. (3B) showed the list of all in one predicted sgRNA sequences against FAD2-1A, in our case we select the third guide sequence in the exonic region with a maximum on target score (0.8581).

Fig. 4. The designing of short sgRNA oligos required the input of previously selected guide sequence (GTATTGATGGAGCAACCAATGGG) into CRISPR vector construction kit suite. The last GGG (PAM) sequence is excluded from the selected sequence. Select CRISPR BGK41 (Soybean) in order to construct forward and reverse oligos for FAD2-1A gene. Order the aforementioned oligos for your CRISPR Cas9 vector, keep in mind that the red color sequences (GGGTT) are the PAM and one must exclude them before ordering these oligos.
NOOR AL AMIN ET AL.,

Fig. 5. The detection of positive Agrobacterium tumafaciens (EHA105 strains) carrying sgRNA-CRISPR Cas9 vector through PCR using CRISPR vector-specific primers. M: 2000marker, Lane 1 described as a negative control and 2-5 represents different positive clones.

Fig. 6. The amplification of FAD2-1A genes fragments using FAD2-1A specific primers from the genomic DNA of mutant protoplasts induced by PEG mediated transformation. The positive and clear bands were cloned successfully and subjected to sequence analysis afterward. M: 2000marker, Lane 1-6 represents differently transformed protoplast with FAD2-1A gene.

PEG induced transient gene expression assay: The introduction of foreign DNA into plant protoplasts simplifies the phenomenon of a transient assay for rapid gene expression analysis and producing high frequency transgenic lines. Polyethylene glycol (PEG)-induced transformation is commonly applied to a variety of systems which allows to yield a rapid and efficient population of transformed cell lines in no time with a rate of maximal survival and multiplication (Potrykus 1991). We have successfully carried out the activity of soybean protoplasts by transforming SgRNA-FAD2-1A-CRISPR Cas9 system through Polyethylene glycol (PEG) induced method. The genomic DNA extraction from the transformed protoplasts was performed after 46 hours of incubation in a complete dark condition. PEG induced transformation was confirmed through PCR using FAD2-1A using gene specific primers (Table 1, Fig. 6). The PCR products of FAD2-1A gene obtained from mutant protoplasts were successfully cloned in T1 cloning vector and sequenced to detect target mutation. In order to predict the structure of protein, we translated both mutant and wild-type sequence into amino acid sequences using [https://web.expasy.org/translate/](https://web.expasy.org/translate/). The SWISS-MODEL template library (SMTL version 2018-05-23, PDB release 2018-05-18) was searched with BLAST (Camacho et al., 2009) and HHblits (Remmers et al., 2012) for evolutionary related homology matching the target sequence in (Table 3).

Finally, the target sequence was searched with BLAST against the primary amino acid sequence contained in the SMTL. An initial HHblits profile was generated using Remmers et al. (2012) method, followed by 1 iteration of HHblits against NR20. The obtained profile was further screened against all profiles of the SMTL. Collectively 11 templates were found. For each identified template, the template's quality was estimated from the features of the target/template alignment. The templates with the maximum quality were selected for final model building using ProMod3 model. The conserved coordinates between target and template are copied from the template and inserted into the model. Insertions and deletions are remodeled using a fragment library. Side chains were rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. Despite ProMod3, an alternative model was built using PROMOD-II in order to verify the output (Guex et al., 2009). Based on the protein modeling result, the predicted mutant protein structure showed slight changes to that of wild-type meaning that the mutation did occur. The mutant structure contained 3 full loops and single smooth ligands with no additional threads, while the wild-type contained an extra thread attached to its ligand along 2 loops at the down corner (Fig. 7).
Table 1. Composition of the YEP+Kanamycine medium used for Agrobacterium incubation.

<table>
<thead>
<tr>
<th>Chemicals</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast extract medium</td>
<td>Yeast Extract 10g/L, Tryptone 10g/L, NaCl 5g/L, and agar 15g/L. The beaker was placed on a shaker and total volume was adjusted with double distilled autoclaved water. pH: 7</td>
</tr>
</tbody>
</table>

Table 2. Composition of digestion solution and MGG solution used in Protoplast induced Transformation.

<table>
<thead>
<tr>
<th>Solutions</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digestion solution</td>
<td>Cellulose R10 (0.5%), Macerozyme R10 (0.5%), Pectolase Y23 (0.1%), Mannitol, (0.6 M), MES (4-morpholineethanesulfonic Acid) 10mM, KCl (20mM), CaCl2 (10mM), BSA (0.1%) pH: 5.7</td>
</tr>
<tr>
<td>MGG solution</td>
<td>MES (4mM), Mannitol, (0.4 M), MgCl2 (15mM). pH: 5.7</td>
</tr>
</tbody>
</table>

Table 3. Primary amino acid sequences of FAD2-IA predicted templates were searched and models were built accordingly.

<table>
<thead>
<tr>
<th>Type</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>NSTELVYTLTQGLAKETMGGGRVRVAKVEVGKPLSRVNPNTKPPFVTGQLKKAIPPHEFQRSSLTSFSFYVYVYDLSFAFIFIYIAATTYHFHLPQPFSLIAWPIYWVLQCGCLLTGVWVIAHECGHHAFLSKYQWVDDVVGTLHSTLLVPPYFSWKSIIHRHHISNTGSLDREVFVPKPKSVWFSKLYNLNPLGRAVSSLVTITGMPMYLAFNVSGRPYDSFASHYHPAPIYSNRERLIIY VSDVAFSVTSTSLYRATKLGVWLLCQVGYVPALLLVINGFLVTITYLQHHTFHALPHYSDSWDNLKLATMDRDGYILNKVFHHITDTHVAAHILFSMPHYTEAMATNAKIPILGEYYQFDDTPFYKALWREARECLYVEPDEGTSEKGVYWYRNYK</td>
</tr>
<tr>
<td>Predicted mutant</td>
<td>MGGRGGRVAKVEVGKPLSRVNPNTKPPFVTGQLKKAIPPHEFQRSSLTSFSFYVYDLSFAFIFIYIAATTYHFHLPQPFSLIAWPIYWVLQCGCLLTGVWVIAHECGHHAFLSKYQWVDDVVGTLHSTLLVPPYFSWKSIIHRHHISNTGSLDREVFVPKPKSVWFSKLYNLNPLGRAVSSLVTITGMPMYLAFNVSGRPYDSFASHYHPAPIYSNRERLIIY VSDVAFSVTSTSLYRATKLGVWLLCQVGYVPALLLVINGFLVTITYLQHHTFHALPHYSDSWDNLKLATMDRDGYILNKVFHHITDTHVAAHILFSMPHYTEAMATNAKIPILGEYYQFDDTPFYKALWREARECLYVEPDEGTSEKGVYWYRNYK</td>
</tr>
</tbody>
</table>

Conclusion

The phenomenon of targeted mutation is considered as one of the most advanced tools in order to identify gene function and crop improvements. CRISPR Cas9 system can integrate high frequency targeted mutagenesis most efficiently and aptly. Our findings concluded an efficient transient expression assay system mediated by CRISPR Cas9 machinery in soybean. Besides this, we have demonstrated the overall online tools used to design the sgRNA component of CRISPR Cas9 system before targeting a selected genome with maximum efficacy. After constructing a sgRNA-CRISPR Cas9 vector for FAD2-IA gene we efficiently transformed it into soybean protoplasts followed by PEG induced transient expression assay to verify the expected results. The Sequencing and protein modeling results confirmed the targeted mutagenesis of FAD2-IA gene at a very specific gene locus. Keeping in mind our findings, this system can be applied in designing other important genome models in future to study similar site-directed mutagenesis and gene expression approaches.

Acknowledgments

This project was funded by the Natural Science Foundation of China under grant number (31771868).

References


(Received for publication 9 December 2017)