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

NOOR AL AMIN[1], NAVEED AHMAD[2], WU NAN[1], FU XIUMING[1], WANG NAN[1], BO XIAOXUE[1], MA TONG[1], AND WANG PIWU[1*]

[1]*Plant Biotechnology Center, Agronomy College, Jilin Agricultural University, Jilin, China*
[2]*College of Life Sciences, Jilin Agricultural University, Jilin, China*
[*]*Corresponding author's email: Peiwuw@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 oligosgRNA 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 (Bhaya *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 (Liang *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; Bortesi & 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 Physcomitrella 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 28ºC 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. Vaccum 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 *Physcomitrella 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 CRISP-P2 audience should keep in mind to ensure various factors pre subjecting our target sequence to CRISP-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 (GUUUUAGAGCUAGAAAUAGCAAGUUA A AA UAA GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCAC

CGAGUCGGUGCUUUU), 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).

```
>chromosome:Glycine_max_v2.0:10:50013288:50015760:1

50013288 GAGTGAAATTAATGAACAGTGCATGTGAGTATGTGACTGTGACACAACCCCCGGTTTTCA 50013347
50013348 TATTGCAATGTGCTACTGTGGTGATTAACCTTGCTACACTGTCGTCCTTGTTTGTTTCCT 50013407
50013408 TATGTATATTGATACCATAAATTATTACTAGTATATCATTTTATATTGTCCATACCATTA 50013467
50013468 CGTGTTTATAGTCTCTTTATGACATGTAATTGAATTTTTTAATTATAAAAAATAATAAAA 50013527
50013528 CTTAATTACGTACTATAAAGAGATGCTCTTGACTAGAATTGTGATCTCCTAGTTTCCTAA 50013587
50013588 CCATATACTAATATTTGCTTGTATTGATAGCCCCTCCGTTCCCAAGAGTATAAAACTGCA 50013647
50013648 TCGAATAATACAAGCCACTAGGCATGGTAAATTAAATTGTGCCTGCACCTCGGGATATTT 50013707
50013708 CATGTGGGGTTCATCATATTTGTTGAGGAAAAGAAACTCCCGAAATTGAATTATGCATTT 50013767
50013768 ATATATCCTTTTTCATTTCTAGATTTCCTGAAGGCTTAGGTGTAGGCACCTAGCTAGTAG 50013827
50013828 CTACAATATCAGCACTTCTCTCTATTGATAAACAATTGGCTGTAATGCCGCAGTAGAGGA 50013887
50013888 CGATCACAACATTTCGTGCTGGTTACTTTTTGTTTTATGGTCATGATTTCACTCTCTCTA 50013947
50013948 ATCTCTCCATTCATTTTGTAGTTGTCATTATCTTTAGATTTTTCACTACCTGGTTTAAAA 50014007
50014008 TTGAGGGATTGTAGTTCTGTTGGTACATATTACACATTCAGCAAAACAACTGAAACTCAA 50014067
50014068 CTGAACTTGTTTATACTTTGACACAGGGTCTAGCAAAGGAAACAACAATGGGAGGTAGAG 50014127
50014128 GTCGTGTGGCCAAAGTGGAAGTTCAAGGGAAGAAGCCTCTCTCAAGGGTTCCAAACACAA 50014187
50014188 AGCCACCATTCACTGTTGGCCAACTCAAGAAAGCAATTCCACCACACTGCTTTCAGCGCT 50014247
50014248 CCCTCCTCACTTCATTCTCCTATGTTGTTTATGACCTTTCATTTGCCTTCATTTTCTACA 50014307
50014308 TTGCCACCACCTACTTCCACCTCCTTCCTCAACCCTTTTCCCTCATTGCATGGCCAATCT 50014367
50014368 ATTGGGTTCTCCAAGGTTGCCTTCTCACTGGTGTGTGGGTGATTGCTCACGAGTGTGGTC 50014427
50014428 ACCATGCCTTCAGCAAGTACCAATGGGTTGATGATGTTGTGGGTTTGACCCTTCACTCAA 50014487
50014488 CACTTTTAGTCCCTTATTTCTCATGGAAAATAAGCCATCGCCGCCATCACTCCAACACAG 50014547
50014548 GTTCCCTTGACCGTGATGAAGTGTTTGTCCCAAAACCAAAATCCAAAGTTGCATGGTTTT 50014607
50014608 CCAAGTACTTAAACAACCCTCTAGGAAGGGCTGTTTCTCTTCTCGTCACACTCACAATAG 50014667
50014668 GGTGGCCTATGTATTTAGCCTTCAATGTCTCTGGTAGACCCTATGATAGTTTTGCAAGCC 50014727
50014728 ACTACCACCCTTATGCTCCCATATATTCTAACCGTGAGAGGCTTCTGATCTATGTCTCTG 50014787
50014788 ATGTTGCTTTGTTTTCTGTGACTTACTCTCTACCGTGTTGCAACCCTGAAAGGGTTGG 50014847
50014848 TTTGGCTGCTATGTGTTTATGGGGTGCCTTTGCTCATTGTGAACGGTTTTCTTGTGACTA 50014907
50014908 TCACATATTTGCAGCACACACACTTTGCCTTGCCTCATTACGATTCATCAGAATGGGACT 50014967
50014968 GGCTGAAGGGAGCTTTGGCAACTATGGACAGAGATTATGGGATTCTGAACAAGGTGTTTC 50015027
50015028 ATCACATAACTGATACTCATGTGGCTCACCATCTCTTCTCTACAATGCCACATTACCATG 50015087
50015088 CAATGGAGGCAACCAATGCAATCAAGCCAATATTGGGTGAGTACTACCAATTTGATGACA 50015147
50015148 CACCATTTTACAAGGCACTGTGGAGAGAAGCGAGAGAGTGCCTCTATGTGGAGCCAGATG 50015207
50015208 AAGGAACATCCGAGAAGGGCGTGTATTGGTACAGGAACAAGTATTGATGGAGCAACCAAT 50015267
50015268 GGGCCATAGTGGGAGTTATGGAAGTTTTGTCATGTATTAGTACATAATTAGTAGAATGTT 50015327
50015328 ATAAATAAGTGGATTTGCCGCGTAATGACTTTGTGTGTATTGTGAAACAGCTTGTTGCGA 50015387
50015388 TCATGGTTATAATGTAAAAATAATTCTGGTATTAATTACATGTGGAAAGTGTTCTGCTTA 50015447
50015448 TAGCTTTCTGCCTAAAAATGCACGCTGCACGGGACAATATCATTGGTAATTTTTTTAAAAT 50015507
50015508 CTGAATTGAGGCTACTCATAATACTATCCATAGGACATCAAAGACATGTTGCATTGACTT 50015567
50015568 TAAGCAGAGGTTCATCTAGAGGATTACTGCATAGGCTTGAACTACAAGTAATTTAAGGGA 50015627
50015628 CGAGAGCAACTTTAGCTCTACCACGTCGTTTTACAAGGTTATTAAAATCAAATTGATCTT 50015687
50015688 ATTAAAACTGAAAATTTGTAATAAAATGCTATTGAAAAATTAAAATATAGCAAACACCTA 50015747
50015748 AATTGGACTGATT                                    50015760
```

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 ProtospacerAdjacent Motif (5'NGG3'),SnoRNA promoter (U6), RNA scaffold (GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG CACCGAGUCGGUGCUUUU) 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 19 bp 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. (19 bp short oligo does not include PAM sequence) (Fig. 4). Low oligo (Reverse) is the exact complementary copy of up oligo (Forward) sequence (5'-ATTGGTTGCTCCATCAATAC-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 (95ºC) and shock (20ºC) 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 tumafeciens (EHA105 strain) by heat and shock method followed by overnight incubation at 28ºC. 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: 95ºC/45s, 55ºC/35s and 72ºC/1m.

ORG: Glycine max (V1.0), **Position:** 10:49416243..49418715, **Length:** 2473

Fig. 3a
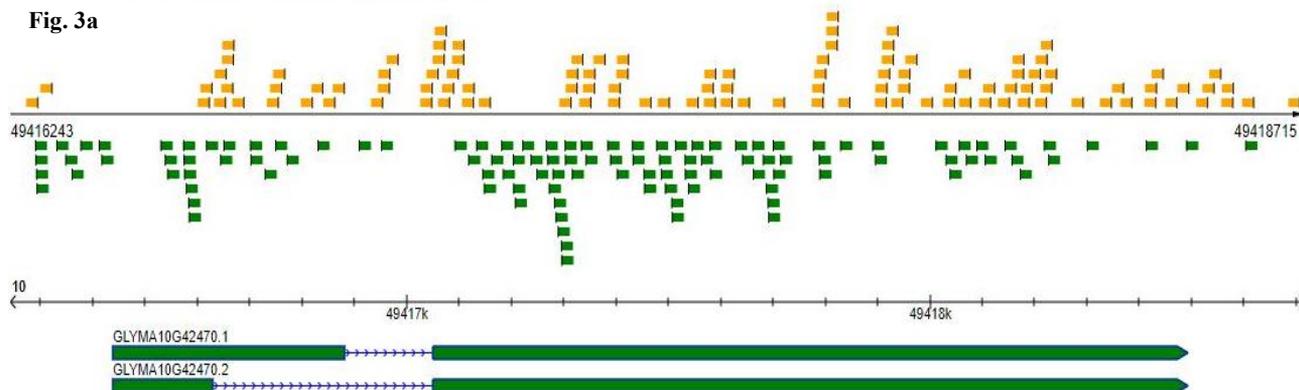
Fig. 3b

	On-score ⑦	Sequence	Region	%GC	
guide1	0.9208	ATACACACAAAGTCATTACGCGG	utr	35%	☑
guide2	0.8959	ACCCCACATGAAATATCCCGAGG	utr	50%	☑
guide3	0.8581	GTATTGATGGAGCAACCAATGGG	exon	40%	☑
guide4	0.8122	GCAAAGGAAACAACAATGGGAGG	exon	45%	☑
guide5	0.7900	ACCCTTTCAGGGTTGCAACACGG	exon	50%	☑
guide6	0.7188	GTTGTGATCGTCCTCTACTGCGG	utr	50%	☑
guide7	0.7093	AATAACCTTGTAAAACGACGTGG	Intergenic	35%	☑
guide8	0.6981	CTACTCATAATACTATCCATAGG	utr	30%	☑
guide9	0.6980	GTTGCATTGACTTTAAGCAGAGG	Intergenic	40%	☑
guide10	0.6768	CATTGCAATATGAAAACCGGGGG	Intergenic	40%	☑
guide11	0.6758	TGCCTTCAGCAAGTACCAATGGG	exon	45%	☑
guide12	0.6193	GATGAAGGAACATCCGAGAAGGG	exon	45%	☑
guide13	0.6088	ACATTGCAATATGAAAACCGGGG	Intergenic	35%	☑
guide14	0.5641	TTAAGCAGAGGTTCATCTAGAGG	Intergenic	40%	☑
guide15	0.5632	GTGCCTTTGCTCATTGTGAACGG	exon	45%	☑
guide16	0.5573	TGTGACTGTGACACAACCCCCGG	Intergenic	55%	☑
guide17	0.5471	TTTATACTCTTGGGAACGGAGGG	utr	40%	☑
guide18	0.5450	TTATACTCTTGGGAACGGAGGGG	utr	45%	☑
guide19	0.5290	AAGAGACTATAAACACGTAATGG	Intergenic	30%	☑
guide20	0.5061	ACCGTGTTGCAACCCTGAAAGGG	exon	50%	☑

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 sgRNAoligos required the input of previously selected guide sequence (GTATTGATGGAGCAACCAAT**GGG**) 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.
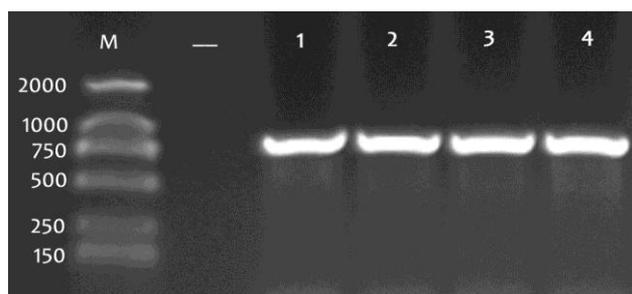
Fig. 5. The detection of positive Agrobacterium tumafaciens (EHA105 strains) carrying sgRNA-CRISPR Cas9 vector through PCR using CRISPR vector-specific primers. M: 2000marker, Lane 1 described as a negative control and 2-5 represents different positive clones.
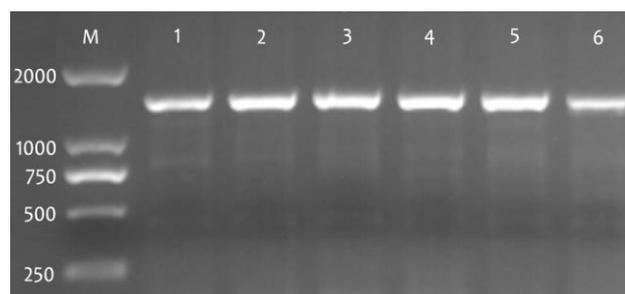
Fig. 6. The amplification of *FAD2-1A* genes fragments using *FAD2-1A* specific primers from the genomic DNA of mutant protoplasts induced by PEG mediated transformation. The positive and clear bands were cloned successfully and subjected to sequence analysis afterward.
M: 2000marker, Lane 1-6 represents differently transformed protoplast with *FAD2-1A* gene.
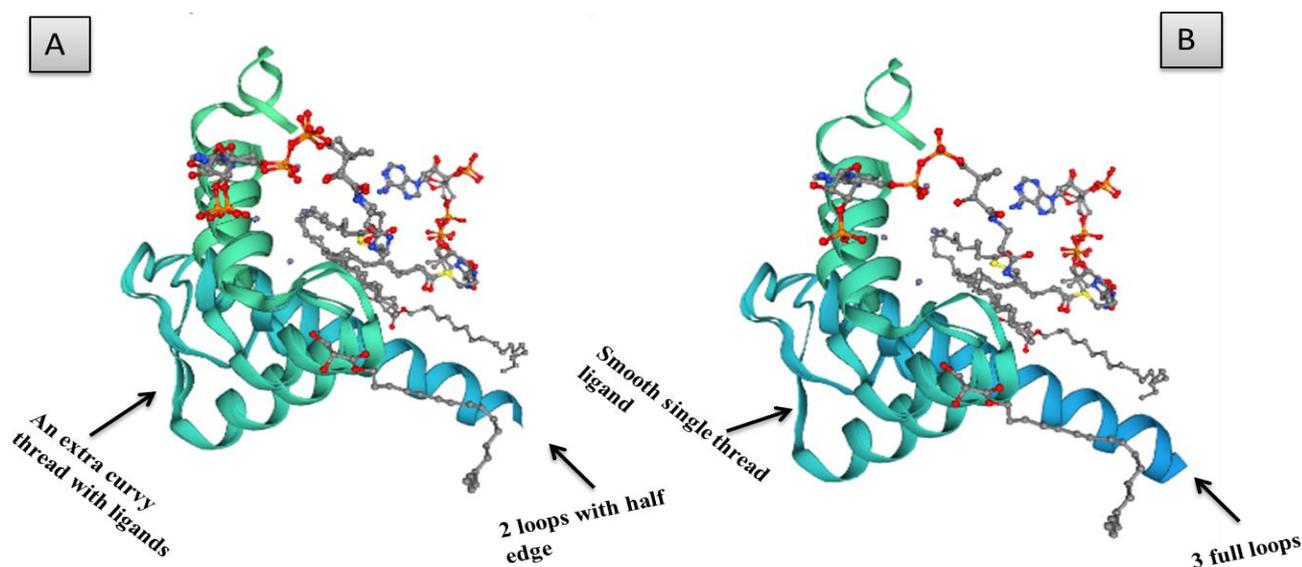
Fig. 7. Protein modeling of (A) wild-type (B) Predicted mutant.

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/). 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.

Chemicals	Composition
Yeast extract medium	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

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

Solutions	Composition
Digestion solution	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
MGG solution	MES (4mM), Mannitol, (0.4 M), MgCl2 (15mM). pH: 5.7

Table 3. Primary amino acid sequences of *FAD2-1A* predicted templates were searched and models were built accordingly.

Type	Sequences
Wild-type	NSTELVYTLTQGLAKETTMGGRGRVAKVEVQGKKPLSRVPNTKPPFTVGQLKKAIPPHC FQRSLLTSFSYVVYDLSFAFIFYIATTYFHLLPQPFSLIAWPIYWVLQGCLLTGVWVIAHEC GHHAFSKYQWVDDVVGLTLHSTLLVPYFSWKISHRRHHSNTGSLDRDEVFVPKPKSKVA WFSKYLNNPLGRAVSLLVTLTIGWPMYLAFNVSGRPYDSFASHYHPYAPIYSNRERLLIY VSDVALFSVTYSLYRVATLKGLVWLLCVYGVPLLIVNGFLVTITYLQHTHFALPHYDSSE WDWLKGALATMDRDYGILNKVFHHITDTHVAHHLFSTMPHYHAMEATNAIKPILGEYY QFDDTPFYKALWREARECLYVEPDEGTSEKGVYWYRNKY
Predicted mutant	MGGRGRVAKVEVQGKKPLSRVPNTKPPFTVGQLKKAIPPHCFQRSLLTSFSYVVYDLSFA FIFYIATTYFHLLPQPFSLIAWPIYWVLQGCLLTGVWVIAHECGHHAFSKYQWVDDVVGL TLHSTLLVPYFSWKISHRRHHSNTGSLDRDEVFVPKPKSKVAWFSKYLNNPLGRAVSLLV TLTIGWPMYLAFNVSGRPYDSFASHYHPYAPIYSNRERLLIYVSDVALFSVTYSLYRVATL KGLVWLLCVYGVPLLIVNGFLVTITYLQHTHFALPHYDSSEWDWLKGALATMDRDYGIL NKVFHHITDTHVAHHLFSTMPHYHAMEATNAIKPILGEYYQFDDTPFYKALWREARECL YVEPDEGTSEKGVYWYRNKY

Conclusion

The phenomenon of targeted mutation is considered most advanced tool 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 *FAD2-1A* 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 *FAD2-1A* 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

References

Bhaya, D., M. Davison and R. Barrangou. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Ann. Rev. Genet.*, 45: 273-297.

Bortesi, L. and R. Fischer. 2015. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.*, 33: 41-52.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T.L. Madden. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, 10: 421-430.

Cradick, T.J., E.J. Fine, C.J. Antico and G. Bao. 2013. CRISPR/Cas9 systems targeting beta-globin and CCR5genes have substantial off-target activity, *Nucleic Acids Res*,doi:10.1093/nar/gkt714.

Fan, D., L. Tingting, L.J.B. Chaofeng, L. Shuang, H. Yishu and L. Keming. 2015. Efficient CRISPR/Cas9-mediated Targeted Mutagenesis in Populus in the First Generation. *Sci. Rep.*, 5: 12217.

Feng, Z., Y. Mao, N. Xu, B. Zhang, P. Wei, D.L. Yang, Z. Wang, Z. Zhang, R. Zheng, L. Yang, L. Zeng, X. Liu and J.K. Zhu. 2014. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proceedings of the National Academy of Sciences*, 111: 4632-4637.

Guex, N., M.C. Peitsch and T. Schwede. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, 30: S162-S173.

Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, E.J. Fine, X. Wu, O. Shalem, T.J. Cradick, L.A. Marraffini, G. Bao and F. Zhang. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases, *Nat Biotechnol*, 31: 827-832.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna and E. Charpentier. 2012. A programmable Double-RNA Guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337: 816-821.doi:10.1126/science.1225829.

Li, Z., B. Zhan, L.X. Aiqiu, P.M. Bryan, P.K. Jessica, H. Lingxia, R.T. Ward, S.C. Elizabeth, C. Falco and A.M. Cigan. 2015. Cas9-Guide RNA Directed Genome Editing in Soybean. *Plant Physiol.*, 169: 960-970.

Liang, Z., K. Zhang, K. Chen and C. Gao,. 2014. Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *J. Genet. Genomics*, 41: 63-68. doi: 10.1016/j.jgg.2013.12.001.

Lin, D., Y. Xiang, Z. Xian and Z. Li. 2016. Ectopic expression of SlAGO7 alters leaf pattern and inflorescence architecture and increases fruit yield in tomato. *Physiol. Plant*, 157: 490-506. doi: 10.1111/ppl.12425.

Miao, J., D. Guo, J. Zhang, Q. Huang, G. Qin, X. Zhang J. Wan, H. Gu and L.J. Qu. 2013. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.*, 23: 1233-1236. doi: 10.1038/cr.2013.123.

Potrykus, I. 1991. Gene transfer to plants assessment of pubhshed approaches and results. *Ann Rev. Plant Physiol. Plant Mol. Blol.,* 42: 205-255.

Puchta, H. and F. Fauser. 2014. Synthetic nucleases for genome engineering in plants: prospects for a bright future. *Plant J.*, 78: 727-41.

Remmert, M., A. Biegert, A. Hauser and J. Söding. 2012. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*, 9: 173-175.

Rinaldo, A.R. and M. Ayliffe. 2015. Gene targeting and editing in crop plants: a new era of precision opportunities. *Mol. Breeding,* 35: 40. https://doi.org/10.1007/s11032-015-0210-z.

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

Wang, Y., X. Cheng, Q. Shan, Y. Zhang, J. Liu, C. Gao and J.L. Qiu. 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.*, 32: 947-951. doi: 10.1038/nbt.2969.

Xu, R.F., H. Li, R.Y. Qin, J. Li, C.H. Qiu, Y.C. Yang, H. Ma, L. Li, P.C. Wei and J.B. Yang. 2015. Generation of inheritable and "transgene clean" targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Sci. Rep.*, 5: 11491. doi: 10.1038/srep11491.

Yoo, S. D., Y.H. Cho and J. Sheen. 2007. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.,* 2: 1565-72.

Zhang, J., F. nasir, Y. Kong, L. Tian, A. Batool, A. Bahadur, X. Li and C. Tian. 2017. Drought stress shapes the root-associated bacterial and fungal community structure in soybean genotypes. *Pak. J. Bot.*, 49(5): 1933-1942.

Zhang, H., J. Zhang, P. Wei, B. Zhang, F. Gou, Z. Feng, Y. Mao, L. Yang, H. Zhang, N. Xu and J.K. Zhu. 2014. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.,* 12: 797-807.