Mysore J. Agric. Sci., 58 (3) : 226-233 (2024)

Selection and *in vitro* Validation of Highly Efficient Single Guide RNAs Fly, *Zeugodacus cucurbitae* (Coquillett)

M. S. Parvathy[1], C. Pradeep[2], S. N. Nagesha[3], R. Asokan[4] and K. M. Harinikumar[5]
[1,2,3&5]Department of Plant Biotechnology, College of Agriculture, UAS, GKVK, Bengaluru - 560 065
[4]ICAR- Indian Institute of Horticultural Research, Hessaraghatta, Bengaluru - 560 089
e-Mail : parvathyms1995@gmail.com

Authors Contribution

M. S. Parvathy & C. Pradeep : Conceptualization, investigation, original draft preparation, analysis;

S. N. Nagesha, R. Asokan & K. M. Harinikumar : Conceptualization, framed research proposal, draft correction

Corresponding Author :
R. Asokan

Received : April 2024
Accepted : June 2024

Abstract

The melon fly (*Zeugodacus cucurbitae)* is widely distributed throughout the world with more than 81 host species in which fruit losses can range from 30 to 100 per cent. CRISPR-Cas9 mediated gene editing has the potential to bring an area wide management of this devastating pest. The sgRNA employed in a CRISPR/Cas9 experiment has a significant impact on the efficiency of mutagenesis. In this study, scarlet gene of melon fly that encodes ATP-binding cassette (ABC) transporters, which are responsible for the translocation of various eye pigments into the proper eye compartment, was selected as target gene since it can impart a visual phenotype on mutation, which is a shift from wild brick red to mutant yellow eye colour. The scarlet gene coding region of size 2454 bp was amplified, cloned using pTZ57R/T vector followed by sequencing. From the sequences obtained, three sgRNAs ($sgRNA_1$, $sgRNA_2$ and $sgRNA_3$) were designed using CHOP-CHOP and CRISPOR online tools, considering different principles such as the presence of protospacer adjacent motif (PAM) region, NGG; GC content (35-50%) and the secondary structures. The *in vitro* transcribed sgRNAs were incubated with Cas9 protein to form Ribo nucleo protein (RNP) complex. *In vitro* restriction assays were done to establish that the RNP complex effectively recognizes and cleaves the PCR amplified scarlet gene by releasing fragments upon gel analysis. The $sgRNA_2$ and $sgRNA_3$ have shown successful *in vitro* restriction with fragments of 2454bp, 2055bp, 393 bp and 2454 bp, 1958 bp, 496 bp, respectively. The failure of $sgRNA_1$ might be due to the deviation from ideal secondary structure. It is evident that despite the *in silico* predictions, not all sgRNAs exhibit the same cleavage effectiveness. To discover this contradiction, *in vitro* technique offers selection of most efficient sgRNAs that may effectively cause double-stranded break at specific targets prior to delivering genome editing agents into live cells.

Keywords : *In vitro* cleavage assay, CRISPR/Cas9, *Zeugodacus cucurbitae*

Genome editing tools represent a ground breaking arsenal in the field of molecular biology, empowering scientists to precisely alter genetic material with unprecedented accuracy and efficiency. Among these tools, CRISPR/Cas9 stands as a titan, revolutionizing genetic engineering with its simplicity, versatility and remarkable precision (Dong *et al*., 2020). Its widespread adoption has not only democratized genome editing but has also accelerated research across diverse disciplines, from agriculture to medicine, promising transformative advancements in our understanding of genetics and the treatment of genetic diseases (Hsu *et al*., 2014).

CRISPR/Cas9 was initially identified in bacterial cells as an adaptive immune defense system against DNA invasion (Jinek *et al*., 2012). This machinery essentially consists of two parts: the Cas9 endonuclease and a guide RNA (gRNA) that can be altered for precise genome editing. A 20-nucleotide sequence that is unique to a gene is called a gRNA (Gao and Zhao, 2014). The target DNA sequence that each complementary gRNA binds to has the Protospacer Adjacent Motif (PAM), which usually has a 'NGG' nucleotide at the 5' end. The PAM region, which is situated three base pairs downstream of the Cas9 endonuclease's cleavage site, is crucial for Cas9 binding (Ran *et al*., 2013). The 32 end of the 20 nucleotide gRNA is next to an 80 nucleotide gRNA scaffold sequence which is necessary for Cas9 binding. When the gRNA-Cas9 complex forms, Cas9 causes a double-strand break precisely 3 bp before the PAM region (Jiang *et al*., 2013). To fix the break site, which is typically prone to errors and causes insertion or deletion (indel) modifications at the cut site, non-homologous end joining (NHEJ) is utilized. Frame-shift mutations, which affect protein translation and thus jeopardize gene function are commonly caused by such indel mutations.

The melon fruit fly, scientifically known as *Z. cucurbitae* (Diptera: Tephritidae), exhibits a broad distribution across temperate, tropical and subtropical regions worldwide (Fletcher, 1987). Within India, it stands as the sole *Tephritid* species with a consistent and widespread presence, infesting a diverse range of cucurbit fruits. With a host range spanning over 81 species, this pest is responsible for significant fruit losses, varying from 30 to 100 per cent (Dhillon *et al*., 2005). Among its favored hosts, cucumber (*Cucumis sativus* L.; Cucurbitales: Cucurbitaceae) holds a prominent position.

Different control measures like cultural control, pheromone-based control, biological control, hot water treatment and lure based approaches have been established for pest control of fruit flies (Subhash *et al*., 2018). However, each one of these control measures is not fully effective in controlling the flies. Furthermore, insecticidal-based control is less effective against fruit flies as the actively feeding larval stage develops inside the fruit, where it is protected from insecticide applications (Hsu *et al*., 2012). The sterile insect technique (SIT) is a species-specific and environment friendly technique for controlling fruit flies by irradiation. Even though SIT is successful in several pest species, there are many obstacles, including expense, difficulty in mass production, mating competitiveness of mass-reared sterile versus wild type insects and damage in somatic tissues after entire body irradiation indicating off-target effects.

To effectively reduce these Tephritids, deploying a Precision Guided Sterile Insect Technique (pgSIT) for area wide pest management mediated by CRISPR/Cas9 can be an attractive alternative (Kandul *et al*., 2019). However, in the process of establishing a pgSIT system, it is required to create two transgenic lines *i.e.,* a sgRNA line and a Cas9 line and on the crossing, the progeny will be non-transgenic and edited at the genomic site specified by the sgRNA (Kandul *et al*., 2019). Further, it can be released into the environment following the biosafety guidelines to achieve either population suppression or population replacement. This requires validation of individual target genes such as sex determination genes, spermatogenesis related genes, *etc.* Validating the target genes *via.,* DNA demands the creation of transgenic lines and it requires more effort, time and higher biosafety levels. Thus, a non-transgenic approach, employing Ribo nucleo protein (RNP) complex consisting of cognate sgRNA and the Cas9 protein (sgRNA+Cas9) delivered to the G_0 embryo is preferred. Validating the marker genes that exhibit the noticeable symptoms is vital before modifying the genes, which are essential to developing the pgSIT system. The eye colour gene scarlet (Ewart and Howells, 1998) was employed as a marker gene in this regard.

The precision and effectiveness of CRISPR/Cas9 mediated genome editing hinge largely upon the sequence and structural characteristics of the single guide RNA (sgRNA). Consequently, prior to conducting RNP-mediated editing in insects, it is crucial to assess the efficacy of different sgRNAs

through *in vitro* validation processes. In this study, we systematically assessed the effectiveness of multiple sgRNAs targeting the scarlet gene in *Z. cucurbitae* using an *in vitro* cleavage assay.

Material and Methods

Rearing of Melon Fruit Fly

The low cost rearing protocol for the mango fruit fly *Z. cucurbitae* was carried out. The nucleus culture of *Z. cucurbitae* was collected from Biopesticide lab, Division of Basic Sciences, ICAR-IIHR and was reared at 27±1°C and 75±1°C per cent relative humidity with 14h: 10h photoperiod (Pradhan *et al*., 2023) in wooden rearing cages. The Elakki variety of banana was provided as an oviposition source and larval food. The pre-pupal wandering larvae on banana fruit were transferred to rectangular plastic boxes (25 x 10 cm) containing sieved soil for pupation. The pupae were collected after 12-14 days by sieving the soil. The collected pupae were placed on a petri dish and transferred to a wooden rearing cage for adult emergence. The adult flies were provided with sugar, yeast extract and water as a food source. Multivitamin asta xanthin syrup was provided twice in a week diluted in water (Jayanthi and Varghese, 2002).

Isolation and Cloning of Scarlet Gene

Total RNA was isolated from 5 days old flies using Trizol method. To verify the quality of RNA agarose electrophoresis (2%) was carried out followed by quantification by Nanodrop (Thermo Scientific, USA). The cDNA synthesis was performed using RevertAid First Strand cDNA synthesis kit (Thermo Fisher Scientific, USA) using 1000ng total RNA as template and the cDNA was stored in -20°C until further usage. The PCR amplification of the scarlet gene was set up for an amplicon size of 2454 bp using gene specific forward and reverse primers (Table 1). The PCR conditions for scarlet gene amplification is shown in Table 2.

The amplified gene product was analyzed by agarose electrophoresis (1.5%) using PCR clean-up kit (Macherey-Nagel, Germany) and further ligated to pTZ57R/T cloning vector (Thermo Scientific, USA). The ligation product was transformed into chemical competent *Escherichia coli* strain DH5α cells. Afterwards, the transformants were detected on an LB plate with ampicillin (100 mg/mL) using blue/white selection. Plasmids were isolated using Thermo Scientific's Plasmid isolation kit and clones were run on 1.2 per cent agarose gel and validated by differential mobility assay. The cloned plasmids were sequenced in triplicates by Sanger sequencing and the identity of sequences were further confirmed by Nucleotide BLAST database of NCBI (https://www.ncbi.nlm.nih.gov). Using the draft genome database available in NCBI, the target genes were characterized for intron/exon structure and annotated using the mRNA sequence. The SWISS-MODEL software (https://swissmodel.expasy.org/) was

Table 1
Primer and sgRNA details

Primer	Sequence	
	PCR primers	Amplicon size (bp)
Scarlet_F	ATGTCAACAGATAATCTCATCTCG	2454
Scarlet_R	GGCACATTAGAGGCAGTTC	
	sgRNA primers	PAM Position
sgRNA1	CTACAGCAAGTGGTCGCCCA	237
sgRNA2	CGCATCATTAACAACTCGAC	396
sgRNA3	TGCCGGCGCCACCACTCACC	469

Table 2
PCR conditions for scarlet gene amplification

Step	Temp. (°C)	Time	No. of Cycles
Initial Denaturation	95	2 min	1
Denaturation	98	10 sec	35
Annealing	56	30 sec	
Extension	72	2 min 50 sec	
Final Extension	72	10 min	1

employed to predict the protein structure using the obtained sequences.

Designing of sgRNA

The exon sequences of target genes were used as an input to design the sgRNA. The target sites were selected based on the criteria: 52 -GGN-18nts-NGG-3'. The sgRNAs used minimized and on-target scored sgRNAs were selected (Table 1). The secondary structure of selected single stranded RNA was verified using RNA fold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi).

sgRNA Synthesis

Cloning of sgRNA

Primers were designed to amplify the Bbs1 site as well as the sgRNA scaffold and terminator from pEn-C1.1 plasmid, the restriction sites of *Sac* I and *Hind* III were tagged at 52 ends in forward and reverse primer, respectively. The pBluescript plasmid was pre-digested with SacI and *Hind* III restriction sites and purified using a PCR clean-up kit (Macherey-Nagel, Germany). PCR was carried out with sgRNA Sac + Bbs1Forward (AGAAGAGCTCATTGGGG TCT TCGAGAAGA) and sgRNA Scaff + *Hind* III Reverse sequences. These PCR product was ligated with pre-digested with *Sac* I and *Hind* III pBluescript plasmid. These were cloned and confirmed by sequencing. The confirmed plasmid was digested with the *Bbs* I restriction site to ligate the sgRNAs. The linker sequence of ATTG and CCCG was added at 52 regions in forward and reverse primer, respectively. Forward and reverse oligonucleotides were hybridized and ligated to the *Bbs* I pre-digested pBluescript plasmid. The ligated products were cloned and the presence of sgRNA was confirmed by sequencing. The confirmed clones were linearized with *Hind* III restriction site, purified using PCR clean-up kit (Macherey-Nagel, Germany) and used as a template for *in vitro* transcription. sgRNA that got transcribed was purified using Monarch® Total RNA Miniprep Kit and was DNase digested, with DNAse I to remove genomic DNA contamination.

Validation of sgRNAs through *in vitro* Restriction Assay

For obtaining efficient cleavage activity it is essential to keep the molar ratio of Cas9 and sgRNA per target site at 10:10:1 or higher. *In vitro* digestion was done following the guidelines provided by the manufactures of *in vitro* digestion kit New England Biolab Inco. The components 30nM Cas9 Nuclease, *S. pyogenes*, 30nM sgRNA were incubated in 10x NE 3.1 Buffer in a reaction volume of 17 µl at 25°C for 30 minutes. The scarlet PCR product (30nM) was added to this mixture followed by incubation at 25°C for 1 hr. The reaction was terminated by addition of 1µl of Proteinase K to each sample and incubated at 40°C for 10 minutes. Analysis of the enzymatic digestion was performed by electrophoresis on a 2 per cent agarose gel.

Results and Discussion

Isolation and Molecular Cloning of Scarlet Gene

The total RNA of high quality was extracted (Fig. 1) and amplified using gene specific primers have given a product of size 2454 bp (Fig. 2). The amplicon after cloning was sequenced and analyzed. The SWISS-MODEL software predicted the protein structure to have a transmembrane segment and modeling was based on the template hypothetical protein Alpha

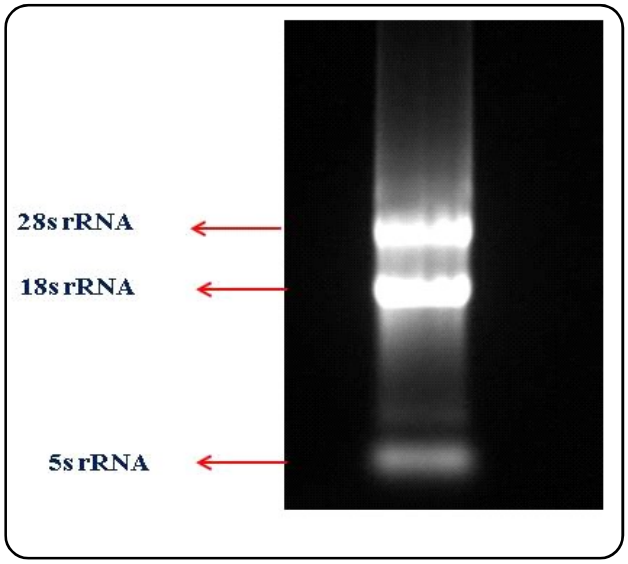

Fig. 1 : Gel image of total RNA isolated from *Z. cucurbitae*

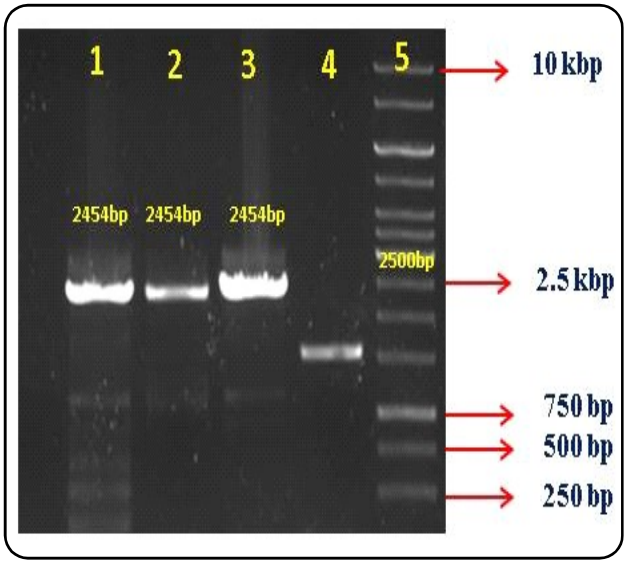

Fig. 2 : Gel image of scarlet gene amplification by PCR. Lane 1,2,3 - scarlet gene Lane 5- 1 kb ladder

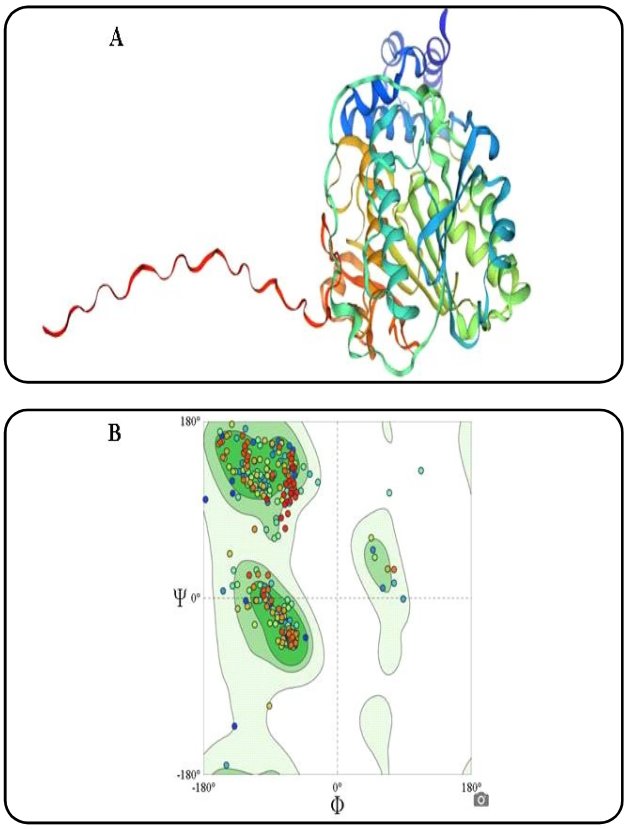

Fig. 3 : Homology modeling of scarlet protein (A) The protein structure of scarlet protein using online software SWISS MODEL (https://swissmodel.expasy.org) (B) The Ramachandran plot analysis revealing 92.31 per cent of amino acid residues were falling under allowed regions

Fold DB model of A0A811V542_CERCA [gene: A0A811V542_CERCA, organism: *C. capitata* (Mediterranean fruit fly) (*Tephritis capitata*)] (A0A811V542.1.A) with a sequence identity of 90.20 per cent for 100 per cent coverage an GMQE value of 0.77 (Fig. 3A). Moreover, 92.31 per cent amino acids were confined to Ramachandran favored region (Fig. 3B).

Designing of sgRNA

The gene sequences obtained were structurally annotated to have seven exons and six introns (Fig. 4). The sequence result obtained was subjected to CHOP CHOP software and three potential guide RNAs were selected from exon 1 of the gene. Koidou *et al.* (2020) have reported that the screening of designed sgRNA targeting scarlet gene in the 5' end of exon one has given better cleavage efficiency in the *in vitro* cleavage assay and was proceeded for CRISPR/Cas9 induced mutagenesis. The secondary structure prediction of these sgRNAs were produced using RNAfold software and the obtained self-folding value/Minimum free energy (ΔG) of $sgRNA_1$, $sgRNA_2$, $sgRNA_3$ were found to be -1.70, -0.20 and 0.00 kcal/mol (Fig. 5) (Table 3).

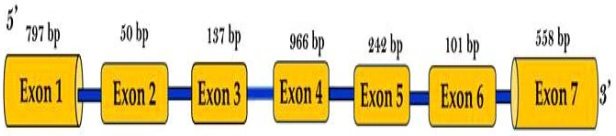

Fig. 4 : Gene structure of scarletgene. It consists of 7 exons (yellow boxes) and 6 introns

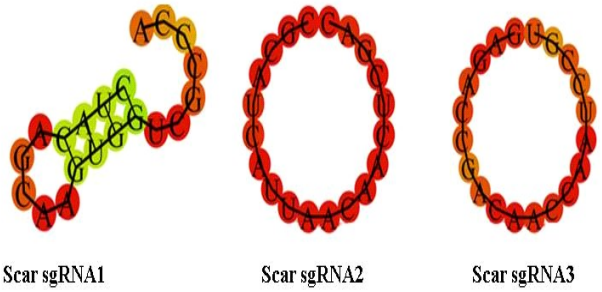

Fig. 5 : Predicted secondary structure of sgRNAs designed for scarlet gene. Legend : Scar-scarlet

TABLE 3

Self-folding value/Minimum free energy (ΔG) of selected single guide RNAs (sgRNAs)

sgRNA name	Free energy (ΔG) value (kcal/mol)
Scarlet sgRNA1	-1.70
Scarlet sgRNA2	-0.20
Scarlet sgRNA3	0.00

These three guides without off-targets were synthesized *in vitro* (Table 1). The *in vitro* transcription of sgRNA PCR templates were quantified using fragment analysis (Fig. 6) and Nanodrop (approx. 2 µg). *In vitro* cleavage assay was performed using the sgRNAs, Cas9 and the scarlet gene product and the cleavage efficiency was confirmed by 2 per cent agarose gel analysis. Upon *in vitro* cleavage process, the scarlet gene of size 2454bp cleaved with $sgRNA_2$-Cas9 RNP gave 3 fragments of sizes 2454 bp, 2055 bp, 393 bp and while $sgRNA_3$-Cas9 RNP gave fragments of 2454 bp, 1958 bp and 496 bp (Fig. 7B). But $sgRNA_1$-Cas9 RNP failed to cleave the gene (Fig. 7A).

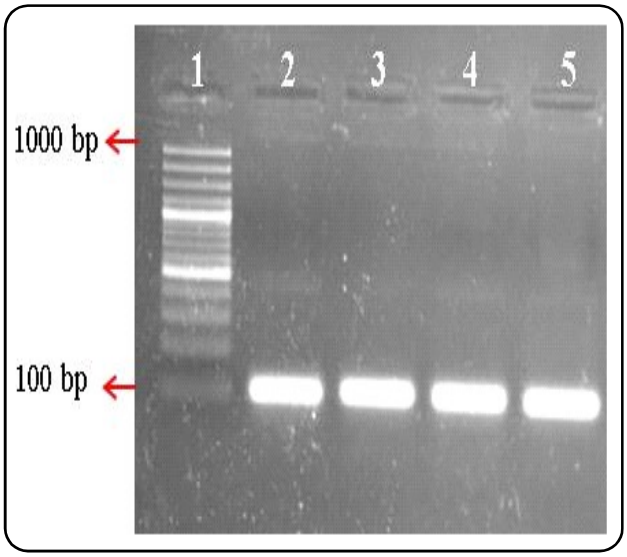

Fig. 6 : *In vitro* transcription of scarlet sgRNAs; M- 100bp DNA ladder, Lane1- scar $sgRNA_1$, Lane2- scar $sgRNA_2$, Lane3- scar $sgRNA_3$

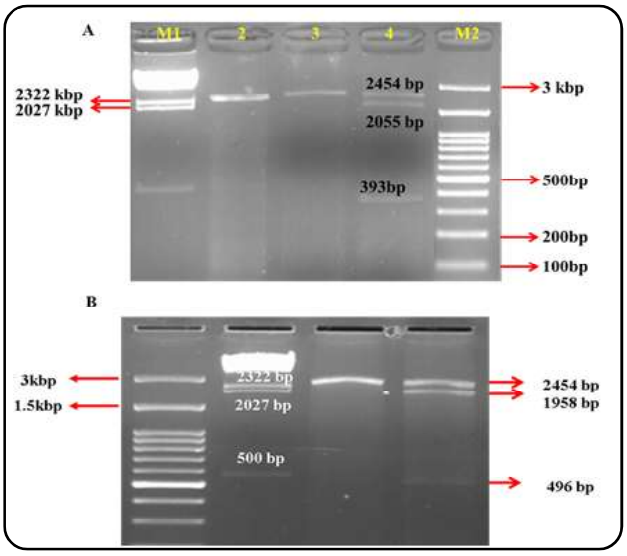

Fig.7 : *In vitro* cleavage of scarlet (scar) gene. (A) *In vitro* cleavage of scarlet (scar) gene with $sgRNA_1$. M1:Δ Hind III Ladder, 2: Control scarlet gene, 3: *In vitro* cleavage reaction mixture of $sgRNA_1$, 4: *In vitro* cleavage reaction mixture of $sgRNA_2$ (B) *In vitro* cleavage of scarlet (scar) gene with $sgRNA_3$. Lane 1-100 bp plus DNA Ladder, 2-Δ Hind III Ladder, 3- Control scarlet gene, 4- IVC $sgRNA_3$

The *in vitro* cleavage assay shows that even though bioinformatically selected, some sgRNAs are not able to cleave the target DNA. The reasons behind this can be several and the secondary structure of the sgRNA is a major one among them. The main structural feature that has been found associated with the functional sgRNA is the presence of unpaired seed sequence of sgRNA employing the target recognition (Hassan *et al*., 2021). The seed region of $sgRNA_1$ is paired and is not free for the target recognition on the scarlet gene template according to it's predicted secondary structure (Fig. 5). Moreover, the self-folding value of $sgRNA_1$ is -1.70 kcal/mol, which is higher than the other two sgRNAs selected. Hence these factors might have contributed to the failure of $sgRNA_1$.

Optimization and designing of highly efficient sgRNA is important for the success of gene editing experiments. However, functionality of sgRNA can be affected due to various reasons *viz.,* position of sgRNA within the gene, off-target score, secondary structure and free energy of sgRNA etc. Thus, before

selecting the sgRNA, it is important to make sure that sgRNA should be preferably in beginning of first exon, with less off-target score and less negative value of free energy for sgRNA. Research has shown that when it comes to self-folding potential or the capacity to form internal structure, non-functional gRNA exhibits a larger negative value of ΔG than functional gRNA (Wong *et al*., 2015 and Jansen *et al*., 2017). Therefore, a gRNA with a more negative ΔG value will have a higher probability of producing secondary structure in a secondary structure investigation. Because of this, base pairing between the gRNA and the target sequence will be less possible.

References

Dong, O. X., Yu, S., Jain, R., Zhang, N., Duong, P. Q., Butler, C., Li, Y., Lipzen, A., Martin, J. A., Barry, K. W. and Schmutz, J., 2020, Marker-free carotenoid-enriched rice generated through targeted gene insertion using CRISPR-Cas9. *Nat. Commun*., **11** (1) : 1178.

Ewart, G. D. and Howells, A. J., 1998, ABC transporters involved in transport of eye pigment precursors in *Drosophila melanogaster*. *Meth. Enzymol*., **292** : 213 - 224.

Fletcher, B. S., 1987, The biology of dacine fruit flies. *Annu. Rev. Entomol.*, **32** (1) : 115 - 144.

Gao, Y. and Zhao, Y., 2014, Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. *J. Integr. Plant Biol*., **56** : 343 - 349.

Hassan, M. M., Chowdhury, A. K. and Islam, T., 2021, *In silico* analysis of gRNA secondary structure to predict its efficacy for plant genome editing. *CRISPR-Cas Methods,* **2** : 15 - 22.

Hsu, J. C., Chien, T. Y., Hu, C. C., Chen, M. J. M., Wu, W. J., Feng, H. T., Haymer, D. S. and Chen, C. Y., 2012, Discovery of genes related to insecticide resistance in *Bactrocera dorsalis* by functional genomic analysis of a de-novo assembled transcriptome. *PLoS ONE*, **7** (8) : e40950.

Hsu, P. D., Lander, E. S. and Zhang, F., 2014, Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157** (6) : 1262 - 1278.

Jayanthi, P. K. and Verghese, A., 2002, A simple and cost-effective mass rearing technique for the tephritid fruit fly, *Bactrocera dorsalis* (Hendel). *Curr. Sci*., **82** (3) : 266 - 268.

Jensen, K. T., Floe, L., Petersen, T. S., Huang, J., Xu, F., Bolund, L., Luo, Y. and Lin, L., 2017, Chromatin accessibility and guide sequence secondary structure affect CRISPR/Cas9 gene editing efficiency. *FEBS Lett.*, **591** (13) : 1892 - 1901.

Jiang, F. and Doudna, J. A., 2017, CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys*., **46** : 505 - 529.

Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L. A., 2013, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol*., **31** : 233 - 239.

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

Kandul, N. P., Liu, J., Sanchez C, H. M., Wu, S. L., Marshall, J. M. and Akbari, O. S., 2019, Transforming insect population control with precision guided sterile males with demonstration in flies. *Nat. Commun.*, **10** (1) : 84.

Koidou, V., Denecke, S., Ioannidis, P., Vlatakis, I., Livadaras, I. and Vontas, J., 2020, Efficient genome editing in the olive fruit fly, *Bactrocera oleae*. *Insect Mol. Biol.,* **29** (4) : 363 - 372.

Pradhan, S. K., Asokan, R. and Shivanna, B., 2023, Cloning of White Gene of the Melon Fly, *Zeugodacus cucurbitae* (Coquillett) (Diptera: Tephritidae) and *in vitro* restriction analysis of different single guide RNAs (sgRNAs). *Mysore J. Agric. Sci.,* **57** (2) : 317 - 326.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. and Zhang, F., 2013, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc*., **8** : 2281 - 2308.

SUBHASH, S., CHAKRAVARTHY, A. K. AND KAMALA JAYANTHI, P. D., 2018, Olfactory responses of gravid melon fly females, *Bactrocera cucurbitae* (Coquillett) to selected cucurbitaceous fruit volatiles. *Mysore J. Agric. Sci*., **52** (2) : 197 - 203.

TYCKO, J., MYER, V. E. AND HSU, P. D., 2016, Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. cell*, **63** (3) : 355 - 370.

WONG, N., LIU, W. AND WANG, X., 2015, WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.*, **16** (1) : 1 - 8.