Target Gene Identification and sgRNA Design for Waterlogging Tolerance in Foxtail Millet via CRISPR-based Transcriptional Activation

Running Title
Discovery and sgRNA Design of Target Gene for CRISPR-based Activation in Foxtail Millet

Target Gene Identification and sgRNA Design for Waterlogging Tolerance in Foxtail Millet via CRISPR-based Transcriptional Activation

Siti Nor Akmar Abdullah¹ ², Sean Mayes³, Mahdi Moradpour²

¹Faculty of Agriculture, Universiti Putra Malaysia 43400 UPM, Serdang, Selangor, Malaysia
²Institute of Plantation Studies, Universiti Putra Malaysia 43400 UPM, Serdang, Selangor Malaysia
³School of Biosciences, University of Nottingham, Sutton Bonington Campus Loughborough LE12 5RD

*Corresponding Author: Siti Nor Akmar Abdullah

Tel: +603-9769 4895
Fax: +603-9769 1099
Email: snaa@upm.edu.my

Abstract

Background:
CRISPR activation (CRISPRa) uses non-functional Cas9 endonuclease (dCas9) but retains the genome targeting ability through its single guide RNAs (sgRNAs). CRISPRa is widely utilised as a gene activation system exploiting its ability in recruiting various transcriptional activation domains (TADs) for enhancing expression of the target gene(s). Drought tolerant and resource efficient crops like millets have potential for mitigating effects of climate change and for enhancing food security.
Objective:
This study aimed to use the *Setaria italica* (foxtail millet) genome sequence in the identification of a
target gene and the subsequent generation of sgRNAs for use in CRISPRa for conferring water logging
tolerance that will benefit future expansion of its cultivation area.

Methods and Results:
Leveraging on existing RNA-seq data and information on functional studies in model plants and from
other cereal species, maize and barley, have enabled identification of candidate *ERFVII* from the
foxtail millet genome sequence in the attempt to engineer waterlogging tolerance. The study provides
a step-by-step example for using publicly accessible databases and bioinformatics tools from NCBI
and Phytozome for the identification and characterization of the ortholog from *Setaria italica.*
Softberry was used for promoter annotation to obtain the transcription start site (TSS). Subsequently,
CRISP-P 2.0 design tools were employed to generate and select a few efficient sgRNAs for CRISPRa
that minimize potentially deleterious off-target binding.

Conclusion:
The study is a useful example on how to advance in genomics research including the revolutionizing
CRISPR technology in *Setaria italica,* which can be adopted in other plant species, through utilization
of the available genome sequence.

Keywords: CRISPR/dCas9, CRISPRa, sgRNA, *Setaria italica,* water logging tolerance, ERFVII
1. Introduction

The applications of clustered regularly interspaced short palindromic repeat (CRISPR) in genomic research have expanded in recent years and developing this technology would enhance the research capability of many existing laboratories. Nuclease-deficient Cas9 (dCas9) is an inactive mutant of Cas9 deficient in endonuclease activity. The CRISPR/dCas9 system has potential to be applied for 1) genome-wide screening for understanding the gene regulatory network affected by the activation of a selected gene; 2) testing the phenotypic effect as the result of changing the expression of a targeted gene; and 3) precise temporal and spatial regulation of a gene (1). Similar to application with CRISPR/Cas9, in CRISPR/dCas9, the synthetic sgRNA is designed to contain two major regions of importance for the CRISPR system, which is the CRISPR RNA (crRNA) spacer and scaffold (tracrRNA) regions. The nucleotides in the spacer region are complimentary to the sequence of the target gene located adjacent to a protospacer adjacent motif (PAM). Any genes and genomic DNA with a sequence complimentary to the spacer region can become possible targets, providing great flexibility to the CRISPR system (2). The scaffold region has the critical role in forming a complex with dCas9 recruited to the targeted genomic site.

In order to modulate the gene expression at the level of transcription via CRISPR activation (CRISPRa) and CRISPR interference (CRISPRi), the dCas9 fused to the transcriptional effectors is directed to the promoter of a target gene. Transcriptional effectors which include transcriptional activators or repressors are protein domains that assist in the recruitment of RNA polymerase and key cofactors for manipulating the transcription of the target gene(s) (3). However, for regulation via dCas9, the target window is not quite as broad as for gene knockout via Cas9 cutting. For CRISPRa, it is most efficacious to target -200 bp to +1 (TSS) in the upstream region, inclusive of the transcription start site (TSS) while for CRISPRi, it is optimal to target +50 bp to +100 bp downstream of the TSS (4). Thus, about a dozen sgRNAs are generated for a given gene targeting the optimal location. It is important to determine the exact location of the TSS. Different databases annotate the TSS in different ways. PlantProm DB (ppdb) (http://ppdb.agr.gifu-u.ac.jp/ppdb/cgi-bin/index.cgi) is a plant promoter database that provides promoter annotation of the model plants, Arabidopsis and rice. It was also recently shown that the TSSP database, in www.softberry.com which relies on ppdb can help in bioinformatic analysis and in locating the TSS of genes from other plant species (5).
Foxtail (*Setaria italica*) is the most important millet species of Eastern Asia and the second-most widely grown species worldwide after pearl millet. It possesses several desirable features for cultivation as a cash crop such as fast ripening, high photosynthetic efficiency and resistant to pests and diseases. Furthermore, it is nutritious with notable medicinal benefits including for controlling diabetes and hyperlipidemia. It is highly attractive as a model plant for scientists due to the possession of several distinct characteristics which include short stature and life cycle, good production of seeds, self-compatibility, a true diploid nature (*2n = 18*), small genome size and its C4 features which can serve as a model for other C4 crops. A high-quality genome sequence of foxtail millet was completed in 2012. More recently, resequencing of 184 foxtail millet recombinant inbred lines and construction of the high-resolution map was carried out to aid essential research on foxtail millet improvement.

Waterlogging is used to describe the persistent flooding of the plant root system. Many parts of South East Asia including Malaysia experience such situations due to the heavy rainfall at certain periods of the year. One of the effects of climate fluctuation is increases duration of high precipitation which can worsen water logging occurrences. Even though millets perform well under drought, the ability to withstand water logging conditions is considered an important trait to have for domestication in the South East Asian countries including Malaysia. Seeds of four different millet species, *Panicum miliaceum* (proso millet), *Panicum sumatrense* (little millet), *Setaria glauca* (yellow foxtail millet), and *Setaria italica* (foxtail millet) were tested for waterlogging tolerance and the effect of pre- and post-heading waterlogging on growth and grain yield. *P. sumatrense* exhibited waterlogging tolerance through enhancement of root growth and the presence of a high proportion of lysigenous aerenchyma in the crown root. Prolonged effects of water logging leads to severe hypoxia due to poor oxygen availability in cells which adversely impacts plant physiological processes and metabolism. Aerenchyma possesses enlarged gas spaces through the programme death of cells in the root that facilitates the diffusion of gases, notably, oxygen from shoots to roots, and CO₂ and ethylene from roots to shoots.

Plants respond to waterlogging through transcriptional reprogramming that leads to modification of protein and metabolite composition in the root system to overcome hypoxia. Previously, flooding tolerance was extensively investigated at the molecular level in tolerant species, such as *Oryza sativa* L. In rice, several proteins involved in tolerance to hypoxia or avoidance of hypoxia belong to the ethylene response factor (ERF) VII family of transcription factors. ERFVII is well recognized for having activity directly linked to oxygen availability. Analysis of the RNA-seq
data of water logging response in the roots of a tolerant maize inbred line, HKI1105, showed that
ethylene plays a fundamental role in tolerance mechanisms. Furthermore, some members of ERFVII
transcription factor in maize were up-regulated in roots, an observation similar to that reported in
Arabidopsis under hypoxia (18). Water logging stress resulted in induced expression of barley
HvERF2.11 possessing the CMVII-1 motif characteristic of ERFVII in the waterlogging tolerance
lineage and introduction of this gene into Arabidopsis significantly enhanced waterlogging tolerance
(19).

Millet, like maize and barley is highly sensitive to water logging. In order to produce
waterlogging tolerant millet through CRISPRa, it is critical to look for a target gene whose
transcriptional activation will enhance water logging response mechanisms that protect the plant. This
project aims to perform bioinformatics analysis for designing sgRNA sequence targeting the promoter
of the most highly homologous gene to the maize ERFVII in foxtail millet for future research to
enhance its transcriptional activity through the application of CRISPR/dCas9 technology for increasing
tolerance to waterlogging.

Materials and methods

1.1 Identification of potential CRISPR targets

Information about the nucleotide and amino acid sequences of the maize gene
(GrMZM2G018398) encoding an ERFVII that was highly up-regulated under waterlogging was
obtained from the RNA-seq data in NCBI. The steps involved in identifying the foxtail millet
ortholog and the design of sgRNAs targeting it using CRISPR-P 2.0 program through to the
production of PCR primers to generate DNA template for in vitro transcription are given in Figure
1 and the details of all the steps are provided below.
Figure 1. Steps involved in sgRNA design targeting foxtail millet gene orthologous to maize ERFVII

BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=ProteinsBLASTp) using the encoded amino acid sequence was used to search for the gene ortholog from foxtail millet. The most strongly homologous gene (hereafter referred to as SiERF1.1) was identified. Comparison of the protein functional domains between the maize ERFVII and the foxtail millet SiERF1.1 was performed to determine the presence of the expected ERFVII signature domains. The identified SiERF1.1 (XP_012698581.1) sequence was blasted in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using non-redundant protein sequence database within Setaria italica (taxid:4555) organism by blastp (protein-protein BLAST). All ERF that produced significant alignments were selected and downloaded in FASTA (complete sequence) format. ERF1 genes possessing the amino acid sequences of the highly conserved 6-bp MCGGAI/L (signature for ERFVII) and the 60-70 bp AP2 domains were selected. The obtained sequences were aligned by MEGA X (20) using Clustal alignment and then the phylogenetic tree of the ERF genes was constructed and estimated using the neighbor-joining method (with 1000 replicates) based on deduced amino acid sequences. The reliability of a phylogenetic tree was also estimated by the bootstrap method.
The promoter sequence of the *SiERF1.1* was retrieved in the *Setaria italica* v2.2 genome. BLAST using the nucleotide sequence of *SiERF1.1* obtained from NCBI as input was used to search the reference *Setaria italica* genome (*Setaria italica* v2.2 genome) in Phytozone 12.1.6 (https://phytozome.jgi.doe.gov/pz/portal.html#) for the nucleotide sequence 500 bp upstream of the start codon (ATG) was performed. The transcript sequence and sequence information found upstream was obtained by specifying the size of sequence information that is required for walking 5' from the 5'-UTR. For obtaining the expected locations of the TSS and TATA box, the 500bp upstream sequence information including the ATG was then used as an input in the promoter prediction program for plant genes (TSSP) in Softberry (www.softberry.com).

1.2 Design of optimized single guide RNAs

The CRISPR-P program version 2.0 (http://crispr.hzau.edu.cn/CRISPR2/) was employed to design sgRNAs with *Setaria italica* v2.2 as the target genome. After selecting the target genome in CRISPR-P 2.0, a gene locus, chromosome position or sequence of the targeted DNA region for search could also be selected. In our design, the promoter region of *SiERF1.1* 150 bp upstream of the TSS including the TATA box was targeted for gene activation using dCas9-activators, and used as the input sequence in CRISPR-P 2.0.

The target sequence of *SiERF1.1* promoter was mapped to its genome, and all possible sgRNAs were screened and shown in a graphical genome model. On-target scores to assess the on-target efficiency of sgRNAs were also obtained from CRISPR-P 2.0. Potential sgRNAs were then identified, their efficiencies were calculated and the predicted results were listed and scored. The sgRNAs DNA template sequences were designed after identifying the target sequence in the promoter region of *SiERF1.1* upstream of the TSS. The template sequence was composed of the T7 promoter sequence, the sequence of the target-specific sgRNAs, and the fixed sequence of the tracrRNA. In Figure 2, the T7 Promoter sequence is shown in blue. Transcription begins at and includes the bold G from the T7 promoter sequence. The non-variable tracrRNA of 80 nucleotides in length is shown in green (Figure 2).
5’TAATACGACTCATACTTATAGGNNNNNNNNNNNNNNNNNNAGTCTAGAGCTAGAAATAGCTTTAAATAAAGGCTAGTCCG
TTATCAACTTGAAAAGTAGGCGAGTGCTGCTTTT-3’

Figure 2. The sgRNA DNA template sequence.

1.3 Design of forward and reverse oligonucleotides for PCR assembly

After identifying the final target sequences, the forward and reverse oligonucleotides were designed to be PCR assembled with the Tracr Fragment + T7 Primer Mix to generate the sgRNAs DNA template. The Tracr Fragment + T7 Primer Mix contains the universal forward and reverse amplification primers and the 80-nt tracrRNA region. Two 34- to 38-bp oligonucleotides were required to assemble the synthetic sgRNA template: a Target F1 forward primer harboring the T7 promoter sequence and a Target R1 reverse primer that harbours the 5’ end of the tracrRNA constant sequence as shown in Figure 3A which will be used for assembly of sgRNA DNA template as shown in Figure 3B. Shortening of the oligonucleotide lengths (≤40 bases) is favoured for the target primers to prevent synthesis mistakes, which occur at higher probability with long oligonucleotides. Forward and reverse target primer sequences that are 34-nt long are produced by the GeneArt™ CRISPR Search and Design tool by default.
Figure 3. PCR assembly of sgRNA DNA template. A) Sequences of the Target F1 forward and Target R1 reverse oligonucleotides required for synthetic sgRNA template assembly. B) Schematic diagram demonstrating the amplified region using the Target F1 forward and Target R1 reverse oligonucleotides to produce the DNA template for in vitro transcription to produce the sgRNA.
2. Results and discussion

2.1 Identification of a target gene in foxtail millet and sgRNA design

BLASTp search showed that the foxtail millet ERF1 (XP_012698581.1) here referred to SiERF1.1 has the highest homology to one of the most highly upregulated ERFVII under hypoxia in maize (21) hence this millet gene was chosen as the target gene for the CRISPR/dCas9 transcriptional activation in our study. Comparing the protein structures of the maize ERFVII, barley HvERF2.11 and foxtail millet SiERF1.1 showed the presence of a conserved N terminal sequence motif MCGGAI/L and the AP2 domain of 60 to 70 amino acids. The presence of one AP2 domain is a common feature that characterised all ERF transcription factor from the AP2 superfamily (22). The N-terminal MCGGAI/L designated as the CMVII-1 motif is the signature motif for ERFVII (23) hence the SiERF1.1 may be categorised as an ERFVII. Our BLASTp search using the MCGGAI/L domain motif discovered at least ten other ERFVII in foxtail millet besides SiERF1.1. The multiple sequence alignment and phylogenetic analysis of the different foxtail millet ERFVII are shown in Figure 4. The phylogenetic tree shows that the foxtail millet ERFVII can be separated into three separate groups based on their protein sequences. XP 004956913.1, which we refer to as SiERF1.2, has the highest amino acid sequence homology with barley HvERF2.11 (19) and is the most closely related member to the foxtail millet SiERF1.1 used as our target gene and they both belong to the same group in the phylogenetic tree, thus increasing the confidence in choosing SiERF1.1 as the target gene for CRISPRa. The effects of targeting both SiERF1.1 and SiERF1.2 through CRISPRa may also be explored in the future through a multiplexing approach (1).
Figure 4. Sequence alignment and phylogenetic analysis of foxtail millet ERF1.1 (XP_012698581.1) with other ERF members from foxtail millet having the MCGGAI/L signature motif identified through BLASTp in NCBI. A) Nucleotide and predicted amino acid sequence of SiERF1.1. B) The sequences of Setaria italica ERF with accession numbers XP 004956913.1, XP 004958676.1, XP 004962330.1, XP 004964607.1, XP 004967520.1, XP 004985469.1, XP 004985472.1, XP 012698500.1, XP 012698581.1, XP 012698773.1, XP 012267988.1 used for constructing the phylogenetic tree using the neighbourhood joining method. The numbers on the nodes indicate bootstrap values from 1000 replicates. C) Multiple sequence alignment of SiERF1.1 and other ERF family members having the N-terminal MCGGAI/L. The same sequences were used in developing the phylogenetic tree.

In the CRISPRa mechanism for transcriptional activation, the transcriptional activation domain (TAD) recruited by the dCas9 needs to be positioned in the promoter region within 200 bp upstream of TSS (4). In order to identify the region upstream of the promoter, it was essential to determine the position of the TSS. Figure 5 shows the nucleotide 500 bp upstream of the ATG of the SiERF1.1 with the A at position 389 as the TSS and the TATA box at 352 as determined by Softberry. This also indicates that the SiERF1.1 belongs to the TATA-containing genes. TATA-box is an important core promoter element involved in transcription initiation of eukaryotic genes (24).

Figure 5. Genomic sequence of SiERF1.1 500 bp upstream of the start codon. Red boxes show the start codon (ATG), the transcription start site (TSS) and the TATA box of the promoter region that was used to design the sgRNAs. Softberry (www.softberry.com) was used in determining the positions of the TSS and TATA box.
Since sgRNA-dCas9 complex could still bind target DNA that is not a perfect match, the off-target effect of CRISPR/dCas9 system is a great concern among researchers. The on-target efficiency scores only supported those sgRNAs containing 5'-NGG-3' PAM in *Streptococcus pyogenes*, thus the binding specificity/capability depends on the PAM-proximal sequence (25). Targeting the promoter region in CRISPRa through CRISPR/dCas9 may produce fewer occurrences of off-targets binding compared to targeting the coding region which may be affected by homologous regions found in gene family members (26). Nevertheless, it is still very important to use a platform for designing the sgRNA which is equipped with the ability to evaluate the characteristics of the sgRNA especially the binding position in the genome as well as its GC content and secondary structure, criteria that influence the functional properties of the sgRNA. CRISPR-P 2.0 ([http://crispr.hzau.edu.cn/CRI</p>
### 2.2 Advanced selection of sgRNAs

The CRISPR-P 2.0 design tool employs a scoring module to evaluate the sgRNAs based on sequence features of sgRNAs, which leads to improvement of on-target efficiency and the construct a predictive model to design critically active sgRNAs (27, 28). The choice of the targeting site is the most critical step in CRISPR/dCas9 technology. Genome-wide specificity analysis included in CRISPR-P 2.0 helps overcome or reduce off-target effects (30). In this study, about 26 sgRNAs were generated when mapped to the genome of foxtail millet targeting \( \text{SiERF1.1} \) promoter region. The results showed that off-target potential among these 26 sgRNAs varies from 0.051 to 0.9. In general, optimum sgRNA should have high on-target scores and less off-target score sites (31). It is important to optimise the on-target location (intergenic for \( \text{SiERF1.1} \)) of the sgRNA through analysing on-target and off-target scores. Six sgRNAs with higher on-target of above 0.4 were selected. All six sgRNA have higher score for on-target compared to off-target. GC content (%) of sgRNAs is also important for the efficiency of CRISPR/dCas9 systems (28). Our results showed that GC content among the six selected gRNAs was high ranging between 50% to 70% and is within the expected range of 30% to 80% for plant sgRNAs (28) as those sgRNAs having exceptionally high or low GC content may be less active (27). Table 1 shows the results for the on-targets, the microhomology score and features of the secondary structure that aid in choosing efficient sgRNAs.

Table 1. Results of sgRNAs designed to target \( \text{SiERF1.1} \) promoter region generated by CRISPR-P 2.0.

<table>
<thead>
<tr>
<th>Guide</th>
<th>On-score</th>
<th>Micro-Score</th>
<th>Secondary structure</th>
<th>Structure features</th>
</tr>
</thead>
<tbody>
<tr>
<td>guide1</td>
<td>0.677</td>
<td>77.64</td>
<td><img src="image" alt="Secondary structure" /></td>
<td>TBP; IBP</td>
</tr>
</tbody>
</table>

\( \text{guide1} \)
The on-target efficiency scores only support gRNAs with 5'-NGG-3' PAM for utilization with *Streptococcus pyogenes* dCas9. Consequently, PAM sequences were considered in our advanced selection of sgRNAs, which profiles secondary structure. The function of the sgRNA relies on the interaction of its secondary structure with the Cas9 protein in vivo. For CRISPR/Cas9 system, the secondary structure of sgRNA can interfere with the editing efficiency as a link between secondary structure and editing efficiency of sgRNAs has been suggested (28, 29). Further selection of sgRNAs was done based on the recommended criteria for selection of efficient sgRNAs as follows:

The total base pairs between guide sequence and the other sequence (TBP) should not be higher than 12, consecutive base pairs (CBP) not higher than 7, while internal base pairs in the guide sequence (IBP) should not be greater than 6. Four out of the six guide sgRNA that met the criteria were selected for generation of the secondary structures as shown in Figure 7.

<table>
<thead>
<tr>
<th>guide</th>
<th>score</th>
<th>TBP</th>
<th>CBP</th>
<th>IBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>67.62</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>69.43</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>73.22</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>95.49</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 7. Schematic representation of the SiERF1.1 sgRNAs secondary structures. The secondary structure of sgRNA 2, 3, 6 and 7 to target promoter region of SiERF1.1.

The secondary structure of sgRNAs showed that two of the designed sgRNAs, Guide 2 and Guide 7 have intact secondary structures including stem loop RAR, stem-loop one, stem-loop two and stem-loop three. The repeat and anti-repeat region (stem loop RAR) could trigger precursor CRISPR RNA (pre-crRNA) processing by the enzyme RNase III and subsequently activates crRNA-guided DNA cleavage (binding for dCas9). The stem-loop one is essential for the function of dCas9-sgRNA-DNA complex. The stem-loop two and three meanwhile, promote formation of a stable complex. Clearly, all three stem-loop structures are required for successful application of CRISPR (30).

There are other online tools that can be used for sgRNA design besides CRISPR-P 2.0. For example, CRISPR/Cas9 target online predictor (CCTop) (https://cctop.cos.uni heidelberg.de:8043/index.html) determines empirically the off-target scores for each sequence, while the CRISPRater score is used to predict the efficiency of sgRNAs (32, 33). E-CRISP (http://www.e-crisp.org/E-CRISP/) is equipped with its own SAE (Specificity, Annotation, Efficacy) score to evaluate the quality of each sgRNA (34). CRISPOR (http://crispor.tefor.net/) provides a versatile platform that can rank the gRNAs according to different scores for evaluating potential off-targets in the specified genome, and for predicting on–target activity (35). A large number of CRISPR/Cas-derived RNA-guided endonucleases (RGENs) have been identified or modified to improve the cutting efficiency and the editing range. Some tools enable the design of gRNAs for RGENs. For example, Cas-Designer (http://www.rgenome.net/cas-designer/) allows
users to choose 20 PAM types from different RGENs (36), while CRISPOR also offers various
PAMs from a defined list. An important criterion to be considered by biologists in exploring these
web-based tools is user-friendliness as this can expedite the process of designing efficient sgRNA
with minimum occurrence of off-targets as demonstrated by CRISPR-P 2.0.

2.3 sgRNA DNA template design

The sgRNAs DNA template sequences were designed after identifying the target sequences in
the promoter region of SiERF1.1 upstream of the TSS. The NNNNs in Figure 8 were replaced with
the target sequences in the selected sgRNAs. The target region represented by the Ns can be up to
20 bases in length. It was noted that the use of only 18 bases (deleting the first two bases from the
5’ end) improves the specificity in binding to the target (29).

5' TAATACGACTCATATAGNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAG
TTAAAATAAGGCTAGTCCGGTTATCAACCTGGAAAAAGTGCCACCAGTCGGTGCTTTT-3'

Guide 2

5' TAATACGACTCATATAGGCTTACGTTTTCCGCGGGTTTGTAGAGCTAGAAATAGCA
AGTAAAAATAAGGCTAGTCCGGTTATCAACCTGGAAAAAGTGCCACCAGTCGGTGCTTTT-3'

Guide 7

5' TAATACGACTCATATAGCCCCACGTCAGCCGAACGGGGTTTTGTAGCTAGAAATAGC
AAGTAAAAATAAGGCTAGTCCGGTTATCAACCTGGAAAAAGTGCCACCAGTCGGTGCTTTT-3'

Figure 8. The sgRNAs-DNA template for SiERF1.11 sequence. The target sequence is in red.

Having at least one G at the start of the transcript improves sgRNA yield from the in vitro transcription
(IVT) reaction. A 5’ G was added to the target sequence at the T7 forward primer in the Tracr Fragment
+ T7 Primer Mix used for the sgRNA template assembly. Target regions with the added 5’ Gs longer
than 21 bases can have a significant affect the on-target activity (37). As transcription starts
immediately after the TATA of the T7 promoter sequence, we may select a target sequence that adds
one to two 5’ Gs within the 20 base sequence naturally or use the T7 promoter sequence in order to
have a single G at the 5’ end of the target sequence because it is found to enhance promoter activation by boosting the transcription initiation of sgRNA (38).

3. Conclusion

Mining the foxtail millet genome sequence using RNA-seq data and information from functional studies in model plants performed earlier and in closely related species, maize and barley, respectively, has enabled identification of candidate foxtail millet ERFVII to explore engineering waterlogging tolerance. This could provide great benefits not only in terms of research advancement but also major saving in time and financial investment of not having to repeat similar gene discovery research in each species. Subsequent use of the CRISP-P 2.0 design tools produced efficient sgRNAs for the foxtail millet ERFVII, minimizing potentially deleterious off-target binding. The sgRNAs will be valuable in future functional studies through CRISPRa to understand the gene regulatory network involved in conferring waterlogging tolerance in foxtail millet. This will benefit future expansion in the cultivation of this crop which naturally grows in arid regions through to wetter parts of the world.

4. Ethics Approval and Consent to Participate

Not applicable.

5. Human and Animal Rights

Not applicable.

6. Availability of Data and Materials

The data supporting the findings of the article is available in the NCBI Sequence Read Archives under the project PRJNA377604, reference no [21]

7. Funding
The authors acknowledge the financial support from Universiti Putra Malaysia for Prof. Siti Nor Akmar Abdullah’s sabbatical leave and this research is part of the outputs of her study during that period.

8. Consent for Publication
   Not applicable

9. Conflict of Interest.
   There is no conflict of interest.

10. Acknowledgements
    The research plan was based on the discussion between SNAA and SM. SNAA carried out the research and wrote the manuscript with the assistance of MM.
11. References


