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 Abdullah[1,2], Sean Mayes[3], Mahdi Moradpour[2]

[1]Faculty of Agriculture, Universiti Putra Malaysia 43400 UPM, Serdang, Selangor, Malaysia

[2]Institute of Plantation Studies, Universiti Putra Malaysia 43400 UPM, Serdang, Selangor Malaysia

[3] School 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 ~ +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 (6) with notable medicinal benefits including for controlling diabetes (7) and hyperlipidemia (8). 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 (9). 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 (10).

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 (11). 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 (12). Prolonged effects of water logging leads to severe hypoxia due to poor oxygen availability in cells which adversely impacts plant physiological processes and metabolism (13). 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 CO_2 and ethylene from roots to shoots (14).

Plants respond to waterlogging through transcriptional reprogramming that leads to modification of protein and metabolite composition in the root system to overcome hypoxia (15, 16). 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 (17). 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.

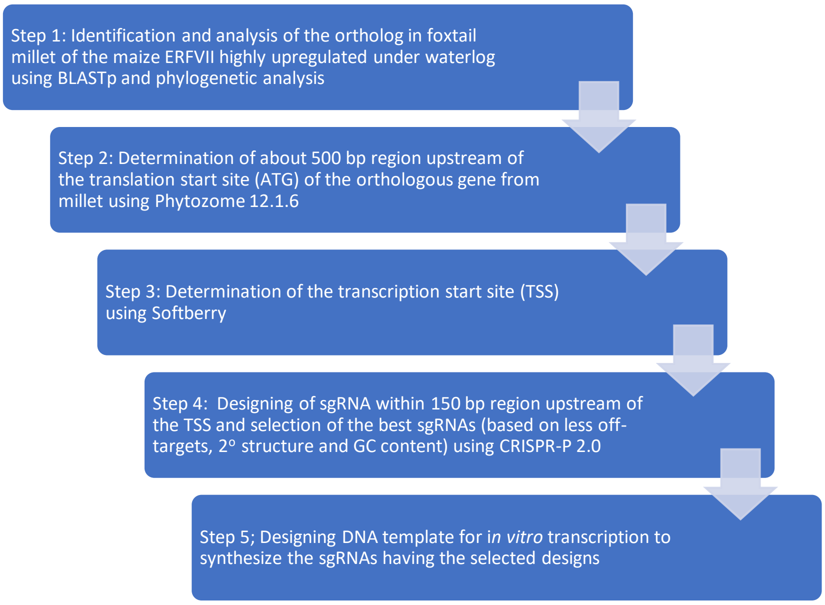

Step 1: Identification and analysis of the ortholog in foxtail millet of the maize ERFVII highly upregulated under waterlog using BLASTp and phylogenetic analysis
Step 2: Determination of about 500 bp region upstream of the translation start site (ATG) of the orthologous gene from millet using Phytozome 12.1.6
Step 3: Determination of the transcription start site (TSS) using Softberry
Step 4: Designing of sgRNA within 150 bp region upstream of the TSS and selection of the best sgRNAs (based on less off-targets, 2° structure and GC content) using CRISPR-P 2.0
Step 5; Designing DNA template for *in vitro* transcription to synthesize the sgRNAs having the selected designs

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 (here after 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 *SiEREF1.1 was retrieved* in the *Setaria italica* v2.2 genome. BLAST using the nucleotide sequence of *SiEREF1.1* obtained from NCBI as input was used to search the reference *Setaria italica* genome (*Setaria italica* v2.2 genome) in Phytozome 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 nueclotides in length is shown in green (Figure 2).

5’TAATACGACTCACTATA**GNNNNNNNNNNNNNNNNNNNN**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG
TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT-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.

A.

Target F1: **TAATACGACTCACTATAG**+first 16–20 nt of the target sequence.
Target R1: **TTCTAGCTCTAAAAC**+first 19–20 nt of the target sequence reverse complement

B.

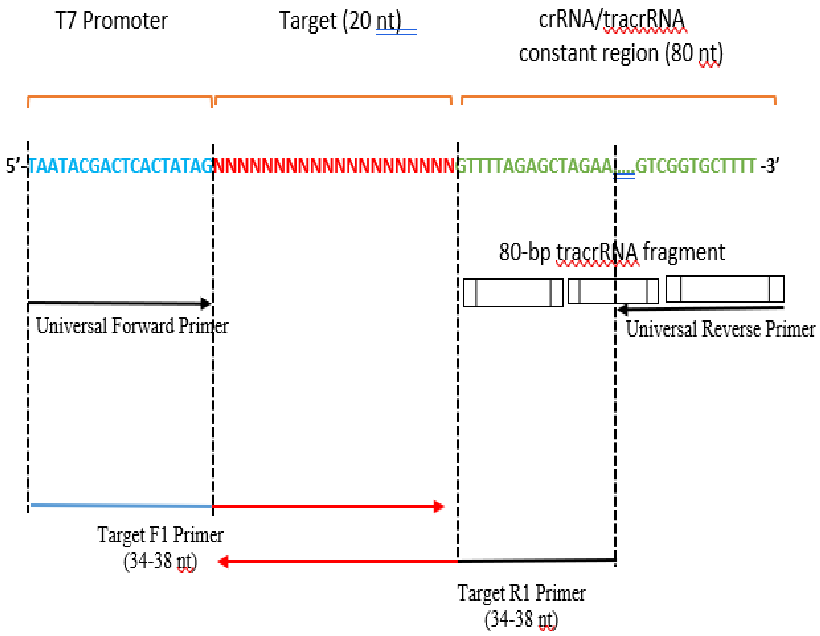

T7 Promoter
Target (20 nt)
crRNA/tracrRNA constant region (80 nt)
5'-TAATACGACTCACTATAGNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAA.....GTCGGTGCTTTT -3'
80-bp tracrRNA fragment
Universal Forward Primer
Universal Reverse Primer
Target F1 Primer (34-38 nt)
Target R1 Primer (34-38 nt)

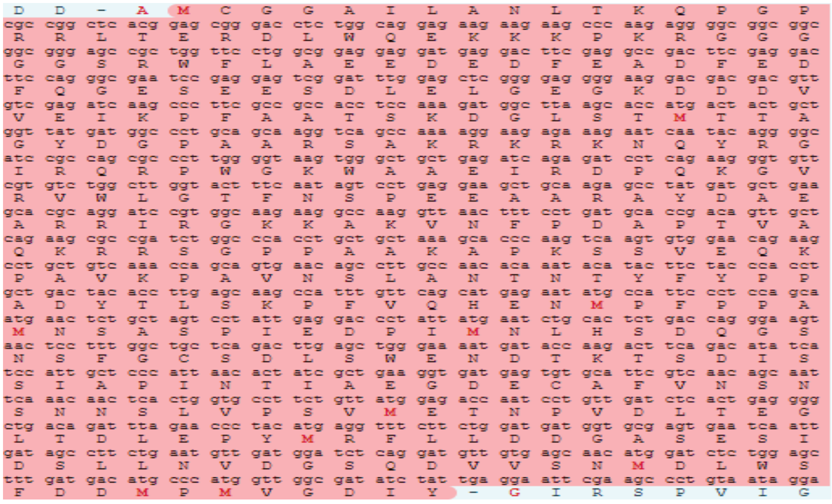
Figure 3. PCR assembly of sgRNA DNA template. A) Sequences of the Target F1 forward and Target R1 reverse oligonucleotides required for synthetic sgRNA template assembly. B) Schematic diagram demonstrating the amplified region using the Target F1 forward and Target R1 reverse oligonucleotides to produce the DNA template for *in vitro* transcription to produce the sgRNA.

2.

Results and discussion

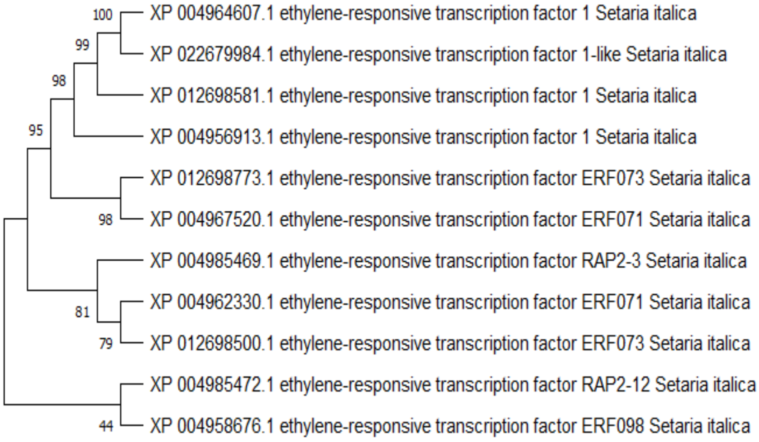
2.1 Identification of a target gene in foxtail millet and sgRNA design

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).

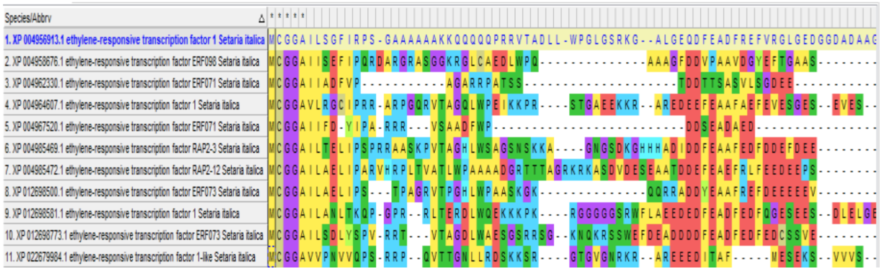


```
 D   D   -   A   M   C   G   G   A   I   L   A   N   L   T   K   Q   P   G   P
cgc cgg ctc acg gag cgg gac ctc tgg cag gag aag aag aag ccc aag agg ggc ggc ggc
 R   R   L   T   E   R   D   L   W   Q   E   K   K   K   P   K   R   G   G   G
ggc ggg agc cgc tgg ttc ctg gcg gag gag gat gag gac ttc gag gcc gac ttc gag gac
 G   G   S   R   W   F   L   A   E   E   D   E   D   F   E   A   D   F   E   D
ttc cag ggc gaa tcc gag gag tcg gat ttg gag ctc ggg gag ggg aag gac gac gac gtt
 F   Q   G   E   S   E   E   S   D   L   E   L   G   E   G   K   D   D   D   V
gtc gag atc aag ccc ttc gcc gcc acc tcc aaa gat ggc tta agc acc atg act act gct
 V   E   I   K   P   F   A   A   T   S   K   D   G   L   S   T   M   T   T   A
ggt tat gat ggc cct gca gca agg tca gcc aaa agg aag aga aag aat caa tac agg ggc
 G   Y   D   G   P   A   A   R   S   A   K   R   K   R   K   N   Q   Y   R   G
atc cgc cag cgc cct tgg ggt aag tgg gct gct gag atc aga gat cct cag aag ggt gtt
 I   R   Q   R   P   W   G   K   W   A   A   E   I   R   D   P   Q   K   G   V
cgt gtc tgg ctt ggt act ttc aat agt cct gag gaa gct gca aga gcc tat gat gct gaa
 R   V   W   L   G   T   F   N   S   P   E   E   A   A   R   A   Y   D   A   E
gca cgc agg atc cgt ggc aag aag gcc aag gtt aac ttt cct gat gca ccg aca gtt gct
 A   R   R   I   R   G   K   K   A   K   V   N   F   P   D   A   P   T   V   A
cag aag cgc cga tct ggc cca cct gct gct aaa gca ccc aag tca agt gtg gaa cag aag
 Q   K   R   R   S   G   P   P   A   A   K   A   P   K   S   S   V   E   Q   K
cct gct gtc aaa cca gca gtg aac agc ctt gcc aac aca aat aca tac ttc tac cca cct
 P   A   V   K   P   A   V   N   S   L   A   N   T   N   T   Y   F   Y   P   P
gct gac tac acc ttg agc aag cca ttt gtt cag cat gag aat atg cca ttc cct cca gca
 A   D   Y   T   L   S   K   P   F   V   Q   H   E   N   M   P   F   P   P   A
atg aac tct gct agt cct att gag gac cct att atg aat ctg cac tct gac cag gga agt
 M   N   S   A   S   P   I   E   D   P   I   M   N   L   H   S   D   Q   G   S
aac tcc ttt ggc tgc tca gac ttg agc tgg gaa aat gat acc aag act tca gac ata tca
 N   S   F   G   C   S   D   L   S   W   E   N   D   T   K   T   S   D   I   S
tcc att gct ccc att aac act atc gct gaa ggt gat gag tgt gca ttc gtc aac agc aat
 S   I   A   P   I   N   T   I   A   E   G   D   E   C   A   F   V   N   S   N
tca aac aac tca ctg gtg cct tct gtt atg gag acc aat cct gtt gat ctc act gag ggg
 S   N   N   S   L   V   P   S   V   M   E   T   N   P   V   D   L   T   E   G
ctg aca gat tta gaa ccc tac atg agg ttt ctt ctg gat gat ggt gcg agt gaa tca att
 L   T   D   L   E   P   Y   M   R   F   L   L   D   D   G   A   S   E   S   I
gat agc ctt ctg aat gtt gat gga tct cag gat gtt gtg agc aac atg gat ctc tgg agc
 D   S   L   L   N   V   D   G   S   Q   D   V   V   S   N   M   D   L   W   S
ttt gat gac atg ccc atg gtt ggc gat atc tat tga gga att cga agc cct gta ata gga
 F   D   D   M   P   M   V   G   D   I   Y   -   G   I   R   S   P   V   I   G
```

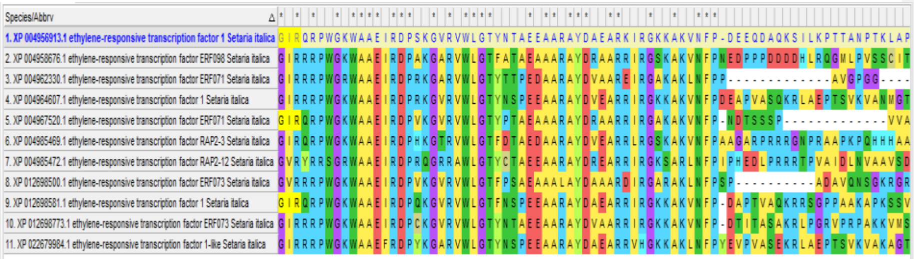
A


100
99
98
95
98
81
79
44
XP 004964607.1 ethylene-responsive transcription factor 1 Setaria italica
XP 022679984.1 ethylene-responsive transcription factor 1-like Setaria italica
XP 012698581.1 ethylene-responsive transcription factor 1 Setaria italica
XP 004956913.1 ethylene-responsive transcription factor 1 Setaria italica
XP 012698773.1 ethylene-responsive transcription factor ERF073 Setaria italica
XP 004967520.1 ethylene-responsive transcription factor ERF071 Setaria italica
XP 004985469.1 ethylene-responsive transcription factor RAP2-3 Setaria italica
XP 004962330.1 ethylene-responsive transcription factor ERF071 Setaria italica
XP 012698500.1 ethylene-responsive transcription factor ERF073 Setaria italica
XP 004985472.1 ethylene-responsive transcription factor RAP2-12 Setaria italica
XP 004958676.1 ethylene-responsive transcription factor ERF098 Setaria italica

B



```
Species/Abbrv                                                              △ *****
1. XP 004956913.1 ethylene-responsive transcription factor 1 Setaria italica   MCGGAILSGFIRPS-GAAAAAAKKQQQQQPRRVTADLL-WPGLGSRKG--ALGEQDFEADFREFVRGLGEDGGDADAAG
2. XP 004958676.1 ethylene-responsive transcription factor ERF098 Setaria italica   MCGGAIISEFIPQRDARGRASGGKRGLCAEDLWPQ--------------AAAGFDDVPAAVDGYEFTGAAS--------
3. XP 004962330.1 ethylene-responsive transcription factor ERF071 Setaria italica   MCGGAIIADFVP-----------AGARRPATSS--------------------TDDTTSASVLSGDEE-----------
4. XP 004964607.1 ethylene-responsive transcription factor 1 Setaria italica   MCGGAVLRGCIPRR-ARPGQRVTAGQLWPEIKKPR----STGAEEKKR--AREDEEFEAAFAEFEVESGES--EVES--
5. XP 004967520.1 ethylene-responsive transcription factor ERF071 Setaria italica   MCGGAIIFD-YIPA-RRR---VSAADFWP-------------------------DDSEADAED----------------
6. XP 004985469.1 ethylene-responsive transcription factor RAP2-3 Setaria italica   MCGGAILTELIPSPRRAASKPVTAGHLWSAGSNSKKA----GNGSDKGHHHADIDDFEAAFEDFDDEFDEE--------
7. XP 004985472.1 ethylene-responsive transcription factor RAP2-12 Setaria italica   MCGGAILAELIPARVHRPLTVATLWPAAAADGRTTTAGRKRKASDVDESEAATDDEFEAEFRLFEEDEEPS--------
8. XP 012698500.1 ethylene-responsive transcription factor ERF073 Setaria italica   MCGGAILAELIPS---TPAGRVTPGHLWPAASKGK--------------QQRRADDYEAAFREFDEEEEEV--------
9. XP 012698581.1 ethylene-responsive transcription factor 1 Setaria italica   MCGGAILANLTKQP-GPR--RLTERDLWQEKKKPK----RGGGGGSRWFLAEEDEDFEADFEDFQGESEES--DLELGE
10. XP 012698773.1 ethylene-responsive transcription factor ERF073 Setaria italica   MCGGAILSDLYSPV-RRT---VTAGDLWAESGSRRSG--KNQKRSSWEFDEADDDDFEADFEDFEDCSSVE--------
11. XP 022679984.1 ethylene-responsive transcription factor 1-like Setaria italica   MCGGAVVPNVVQPS-RRP--QVTTGNLLRDSKKSR----GTGVGNRKR--AREEEDITAF-----MESEKS--VVVS--
```



```
Species/Abbrv
1. XP 004956913.1 ethylene-responsive transcription factor 1 Setaria italica       GIRQRPWGKWAAEIRDPSKGVRVWLGTYNTAEEAARAYDAEARKIRGKKAKVNFP-DEEQDAQKSILKPTTANPTKLAP
2. XP 004958676.1 ethylene-responsive transcription factor ERF098 Setaria italica  GIRRRPWGKWAAEIRDPAKGARVWLGTFATAEAAARAYDRAARRIRGSKAKVNFPNEDPPPDDDDHLRQGMLPVSSCIT
3. XP 004962330.1 ethylene-responsive transcription factor ERF071 Setaria italica  GIRRRPWGRWAAEIRDPRKGARVWLGTYTTPEDAARAYDVAAREIRGAKAKLNFPP-------------AVGPGG----
4. XP 004964607.1 ethylene-responsive transcription factor 1 Setaria italica       GIRRRPWGKWAAEIRDPRKGVRVWLGTYNSPEEAARAYDVEARRIRGKKAKVNFPDEAPVASQKRLAEPTSVKVANMGT
5. XP 004967520.1 ethylene-responsive transcription factor ERF071 Setaria italica  GIRQRPWGKWAAEIRDPVKGVRVWLGTYPTAEAAARAYDRAARRIRGAKAKVNFP-NDTSSSP-------------VVA
6. XP 004985469.1 ethylene-responsive transcription factor RAP2-3 Setaria italica  GIRQRPWGKWAAEIRDPHKGTRVWLGTFDTAEDAARAYDVEARRLRGSKAKVNFPAAGARPRRRGNPRAAPKPQHHHAA
7. XP 004985472.1 ethylene-responsive transcription factor RAP2-12 Setaria italica GVRYRRSGRWAAEIRDPRQGRRAWLGTYCTAEEAARAYDREARRIRGKSARLNFPIPHEDLPRRRTPVAIDLNVAAVSD
8. XP 012698500.1 ethylene-responsive transcription factor ERF073 Setaria italica  GVRRRPWGKWAAEIRDPVKGVRVWLGTFPSAEAAALAYDAAARDIRGARAKLNFPSP----------ADAVQNSGKRGR
9. XP 012698581.1 ethylene-responsive transcription factor 1 Setaria italica       GIRQRPWGKWAAEIRDPQKGVRVWLGTFNSPEEAARAYDAEARRIRGKKAKVNFP-DAPTVAQKRRSGPPAAKAPKSSV
10. XP 012698773.1 ethylene-responsive transcription factor ERF073 Setaria italica GIRRRPWGKWAAEIRDPCKGVRVWLGTYNTAEEAARAYDVAARRIRGKKAKVNFP-DTITASAKRLPGRVPRPAKKVMS
11. XP 022679984.1 ethylene-responsive transcription factor 1-like Setaria italica GIRRRPWGKWAAEFRDPYKGARVWLGTYNSPEEAARAYDAEARRVHGKKAKLNFPYEVPVASEKRLAEPTSVKVAKAGT
```

C

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 0122679984.1 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.

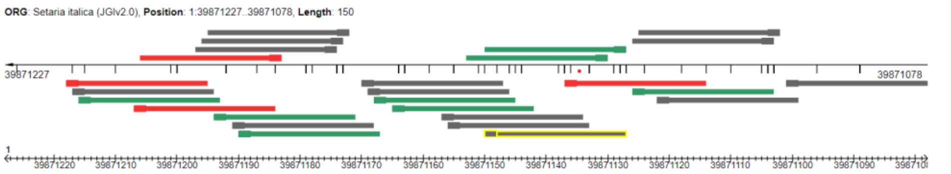
In the CRISPRa mechanism for transcriptional activation, the transcriptional activation domain (TAD) recruited by the dCas9 needs to be positioned in the promoter region within 200 bp upstream of TSS (4). In order to identify the region upstream of the promoter, it was essential to determine the position of the TSS. Figure 5 shows the nucleotide 500 bp upstream of the ATG of the *SiERFI.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).

```
>Seita.1G342700 | scaffold_1:39868092..39871080 reverse
TTGAAGATACAATTTGAAAAGCAAAATAATTGTTAAAGTGAGGATGGCGCGCTG
AGTGAGTTTTCAGCGTGTGATGATCTTTGTAGGAGAATGAAACATGCCCTCTAAACC
TCGGAATCTTGCATTCTTGCCTTACTGTTTGAGGGATATGTTCATCAAATTATATATT
CACAAAGGGGCGACATGTGCAAATTTGGCCTCTACAGCCTACTAGCACCATCTACTA
CACCTTCTCCGTTTCTCTTACTACCCCGTCGCTTCCGCTTCACGTTTCCGCCCGTTCGG
CTGACGTGGGGACCCCACCGCGCGTCCCAGCACCTGCACACTGCTCCGAGCGGCGGC
CAGCCAATATATATATATATATAGGGCCCACCCAGTCCGTGCTCTCCAACATTTCGCCC
AGGTAAAACTCGTTTACCATTTCGCTCTCGCGACTCGCGATTTTTTTTTCACTGTGCT
CGTTCCGGCGCGCGGCGGCGGAGCTCCCGACGACGACTGAGCCATG
```

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/CRISPR2/) is suitable for designing highly efficient sgRNA with minimal off-target effects. CRISPR-P 2.0 uses a scoring system for rating the off-targeting potential and on-targeting efficiency of sgRNAs for *Streptococcus pyogenes* Cas9, the most commonly used CRISPR-Cas9 system (27). The scoring system is based on the latest knowledge about *Streptococcus pyogenes* Cas9 genome editing. Detailed information of the guide sequence is generated, consisting of: GC content, restriction endonuclease site, microhomology sequence flanking the targeting site (microhomology score), and the secondary structure of sgRNA. Figure 6 provides a graphic genome model of mapping *SiERF1.1* target sequence (150 sequence upstream of the *ERF1.1* promoter inclusive of the TSS) to the *Setaria italica* v2.2 genome through CRISPR-P 2.0 design tool.

Commented [SM1]: Does it directly use the genome sequence and look for matches throughout?



A

	Sequence	Region	%GC
guide1	TATATTGGCTGGCCGCCGCTCGG	Intergenic	60
guide2	GCTTCACGTTTCCGCCCGTTCGG	Intergenic	60

guide3	GAACGGGCGGAAACGTGAACGG	Intergenic	60
guide4	AACGTGAAGCGGAAGCGACGGGG	Intergenic	60
guide6	GGGTCCCCACGTCAGCCGAACGG	Intergenic	70
guide7	CCCCACGTCAGCCGAACGGGCGG	Intergenic	75

B

Figure 6 Mapping of *SiERF1.1* target sequence to the *Setaria italica* v2.2 genome through CRISPR-P 2.0 design tool using 150 sequence upstream of the *SiERF1.1* promoter inclusive of the TSS as an input. **A)** Distribution of sgRNA targets in Setaria italica genome. The presumably best target sequences in red (score > 0.50) and the intermediates in green (0.20 < score < 0.50). **B)** DNA target sequences for six selected sgRNA, their region in the genome and %GC.

2.2 Advanced selection of sgRNAs

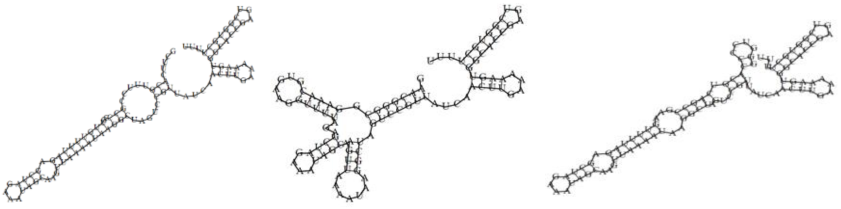
The CRISPR-P 2.0 design tool employs a scoring module to evaluate the sgRNAs based on sequence features of sgRNAs, which leads to improvement of on-target efficiency and the construct a predictive model to design critically active sgRNAs (27, 28). The choice of the targeting site is the most critical 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 *SiERF.1.1* 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 *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 (30). 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 *SiERF1.1* promoter region generated by CRISPR- P 2.0.

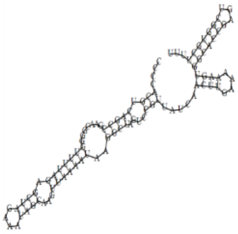
Guide	**On-score**	**Micro-Score**	**Secondary structure**	**Structure features**
guide 1	0.6774	77.64	**UAUAUUGGCUGGCCGCCGCU**GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU **.(((.(((((((((.....(**(((((((.((((....))))...))))))))).)))))).)))).)))..((((....))))((((((((...))))))))... (-30.50)	3 TSL;0 GSL;9 CBP;13 TBP;0 IBP

guide 2	0.6070	67.62	GCUUCACGUUUCCGCCCGUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU .....(((.....(((..((((((((((.((((....))))...))))))))))))).....))).....((((....))))(((((((...)))))))... (-24.70)	3 TSL;0 GSL;3 CBP;8 TBP;0 IBP
guide 3	0.6011	0.00	GAACGGGCGGAAACGUGAAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU .(((((((.(((((......)))))...((((....)))).((((.......)))).)))))))....((((....))))(((((((...)))))))... (-27.80)	5 TSL;0 GSL;7 CBP;12 TBP;0 IBP
guide 4	0.5539	0.00	AACGUGAAGCGGAAGCGACGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU ....((((((((((((....(((((((.((((....))))...)))))))...)))..)))))).)))((((....))))(((((((...)))))))... (-30.40)	3 TSL;0 GSL;12 CBP;12 TBP;0 IBP
guide 6	0.4469	69.43	GGGUCCCCACGUCAGCCGAAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU ((....))(((.(((((...(((((((.((((....))))...)))))))..)))).)..))).....((((....))))(((((((...)))))))... (-26.20)	4 TSL;1 GSL;5 CBP;10 TBP;2 IBP
guide 7	0.4385	73.22	CCCCACGUCAGCCGAACGGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU ....(((.(((((.......(((((((.((((....))))...)))))))..)))).)..))).....((((....))))(((((((...)))))))... (-23.20)	3 TSL;0 GSL;5 CBP;8 TBP;0 IBP

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.



Guide 2 Guide 3 Guide 6



Guide 7

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'TAATACGACTCACTATAGNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAG
TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT-3'

Guide 2

5'TAATACGACTCACTATAG**GCTTCACGTTTCCGCCCGTT**GTTTTAGAGCTAGAAATAGCA
AGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT-3'

Guide 7

5'TAATACGACTCACTATAG**CCCCACGTCAGCCGAACGGG**GTTTTAGAGCTAGAAATAGC
AAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT-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 *ERVII*, 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

1. Moradpour M, Abdulah SNA. CRISPR/dCas9 platforms in plants: strategies and applications beyond genome editing. Plant Biotechnol J. 2020;18(1):32–44.
2. Harish KS. Current status of potential applications of repurposed Cas9 for structural and functional genomics of plants. Mini review. Biochem and Biophys Res Commun. 2016;480(4):499-50713.
3. Konermann S, Brigham MD, Trevino AE, Abudayyeh OO, Barcena C, Hsu PD, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015;517(7536):61422–7.

4 Gilbert LA, Horlbeck MA, Adamson B, Jacqueline E, Chen Y, Whitehead EH, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell. 2014;159(3):647–61.

5. Shahmuradov IA, Umarov RK, Solovyev V V. TSSPlant: A new tool for prediction of plant Pol II promoters. Nucleic Acids Res. 2017;45(8).
6. Durairaj M, Gurumurthy G, Nachimuthu V, Muniappan K, Balasubramanian S. Dehulled small millets: The promising nutricereals for improving the nutrition of children. Matern Child Nutr. 2019;15(Suppl 3): e12791.
7. Kam J, Puranik S, Yadav R, Manwaring HR, Pierre S, Srivastava RK, Yadav RS. Dietary Interventions for Type 2 Diabetes: How Millet Comes to Help. Front Plant Sci. 2016;7(1454):1-14
8. Lee SH, Chung I-M, Cha Y-S, Park Y. Millet consumption decreased serum concentration of triglyceride and C-reactive protein but not oxidative status in hyperlipidemic rats. Nutr Res. 2010;30(4):290-6.
9. Saxena R, Vanga SK, Wang J, Orsat V, Raghavan V. Millets for food security in the context of climate change: A review. Sustain [Internet]. 2018;10(7). Available from: www.mdpi.com/journal/sustainability
10. Ni X, Xia Q, Zhang H, Cheng S, Li H, Fan G, et al. Updated foxtail millet genome assembly and gene mapping of nine key agronomic traits by resequencing a RIL population [Internet]. Vol. 6, GigaScience. 2017. Available from: https://academic.oup.com/gigascience/article-abstract/6/2/giw005/2929393
11. Loo YY, Billa L, Singh A. Effect of climate change on seasonal monsoon in Asia and its impact on the variability of monsoon rainfall in Southeast Asia. Geosci Front [Internet]. 2015 ;6(6):817–23. Available from: https://www.sciencedirect.com/science/article/pii/S16749871140

12. Matsuura A, An P, Murata K, Inanaga S. Effect of pre-and post-heading waterlogging on grow; th and grain yield of four millets. Plant Prod Sci [Internet]. 2016 [cited 2020 Sep 8];19(3):348–59. Available from: https://www.jstage.jst.go.jp/article/pps/15/4/15_323/_article/-char/ja/

13. Araki H, Hossain MA, Takahashi T. Waterlogging and Hypoxia have Permanent Effects on Wheat Root Growth and Respiration. Journal of Agronomy and Crop Science. 2012;198(4):264-75

14. Yamauchi T, Shimamura S, Nakazono M, Mochizuki T. Aerenchyma formation in crop species: A review. F Crop Res. 2013;152:8-16

15. Irfan M, Hayat S, Hayat Q, Afroz S, Ahmad A. Physiological and biochemical changes in plants under waterlogging. Protoplasma. 2010;241(1):3-17.

16. Mendiondo GM, Gibbs DJ, Szurman-Zubrzycka M, Korn A, Marquez J, Szarejko I, Maluszynski M, King J, Axcell B, Smart K, Corbineau F And Holdsworth MJ. Enhanced waterlogging tolerance in barley by manipulation of expression of the N-end rule pathway E3 ligase PROTEOLYSIS6. Plant Biotechnology J. 2016;40(1):40-5

17. Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, et al. (2006). Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature 2006;442:705-08.

18. Gil-Monreal M, Giuntoli B, Zabalza A, Licausi F, Royuela M. ERF-VII transcription factors induce ethanol fermentation in response to amino acid biosynthesis-inhibiting herbicides. J Exp Bot. 2019;70(20):5839–51.

19. Luan H, Guo B, Shen H, Pan Y, Hong Y, Lv C, et al. Overexpression of barley transcription factor HvERF2.11 in Arabidopsis enhances plant waterlogging tolerance. Int J Mol Sci. 2020;21(6).

20. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9.

21. Arora K, Panda KK, Mittal S, Mallikarjuna MG, Rao AR, Dash PK, et al. RNAseq revealed the important gene pathways controlling adaptive mechanisms under waterlogged stress in maize. Sci Rep [Internet]. 2017 [cited 2020 Sep 8];7(1). Available from: https://www.nature.com/articles/s41598-017-10561-1

22. Ebrahimi M, Abdullah SNA, Aziz MA, Namasivayam P. A novel CBF that regulates abiotic stress response and the ripening process in oil palm (*Elaeis guineensis*) fruits. Tree Genet Genomes. 2015;11(3).

23. Nakano T, Suzuki K, Fujimura T, Shinshi H. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant. Physiol. 2006;140:411-32.

24. Tora L, Timmers HTM. The TATA box regulates TATA-binding protein (TBP) dynamics in vivo. Trends Biochem Sci. 2010 Jun 1;35(6):309-14.

25. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotechnol. 2014;32(7):677–83.

26. Manghwar H, Lindsey K, Zhang X, Jin S. CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing. Trends Plant Sci [Internet]. 2019;24(12):1102–25. Available from: https://doi.org/10.1016/j.tplants.2019.09.006

27. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9- mediated gene inactivation. Nat Biotechnol. 2014;32(12):1262–7.

28. Liang G, Zhang H, Lou D, Yu D. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Sci Rep [Internet]. 2016;6:1–8. Available from: http://dx.doi.org/10.1038/srep21451

29. Fu Y, Sander JD, Reyon D, Cascio VM, Keith JJ. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014;32(3):279–84.

30. Liu H, Ding Y, Zhou Y, Jin W, Xie K, Chen L-L. CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. Molecular Plant. 2017;10:530–32

31. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016 Feb;34(2):184–91.

32. Labuhn M, Adams FF, Ng M, Knoess S, Schambach A, Charpentier EM, et al. Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications Nucleic Acids Res. 2018; 46:1375-85.

33. Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. CCTop: an intuitive, flexible and reliable crispr/cas9 target prediction tool PLoS One. 2015; 10.

34. Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification Nat Meth. \ 2014; 11:122-24.

35. Concordet JP, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens Nucleic Acids Res. 2018; 46: W242-W245.

36. Park J, Bae S, Kim JS. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites Bioinformatics. 2015; 31: 4014-16.

37. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. Cell [Internet].

2013;154(6):1380–9. Available from: http://dx.doi.org/10.1016/j.cell.2013.08.021

38. Li Z, Zhang D, Xiong X, Yan B, Xie W, Sheen J, et al. A potent Cas9-derived gene activator for plant and mammalian cells. Nat Plants [Internet]. 2017;3(12):930–6. Available from: http://dx.doi.org/10.1038/s41477-017-0046-0