JOURNAL OF TROPICAL LIFE SCIENCE
2023, Vol. 13, No. 2, 257 – 276
http://dx.doi.org/10.11594/jtls.13.02.04

Research Article

Development of CRISPR/Cas9 Construct in Rice (*Oryza sativa* subsp. *indica*) Using Golden Gate Cloning Method Towards Drought Tolerance

Anis Afuza Md Yusof [1], Amin-Asyraf Tamizi [1], Nurul Asyikin Mohd-Zim [1], Siti Syafiqa Abdul Sattar [1], Mohd Syahmi Salleh [1], Nur Sabrina Ahmad Azmi [2], Zamri Zainal [3], Zarina Zainuddin [2], Nurul Hidayah Samsulrizal [1]*

[1] Department of Plant Science, Kulliyyah of Science, International Islamic University of Malaysia, 25200 Kuantan, Pahang
[2] Plant Productivity and Sustainable Resource Unit, Kulliyyah of Science, International Islamic University Malaysia, 25200, Kuantan, Pahang
[3] Institute of Systems Biology, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia

Article history:
Submission September 2022
Revised November 2022
Accepted November 2022

**Corresponding author:*
E-mail:
hidayahsamsulrizal@iium.edu.my

ABSTRACT

Rice (*Oryza sativa*) is a staple food consumed by the majority of the world's population. Climate change, however, has created a significant threat to our food security as it posed severe effects on rice production. The emergence of genome editing technology has opened a new era in crop improvement. Hence, this study aims to develop the CRISPR/Cas9 construct of drought tolerance for *O. sativa* subsp. *indica* cv. IR64 using Golden Gate cloning method. For this purpose, the generation of CRISPR/Cas9 constructs involved several stages, i.e., characterization of *SUMO E2-Conjugating Enzyme* (*OsSCE1*) gene, single-guide RNA (sgRNA) design and vector construction. FGENESH, GeneMarkS, InterProScan, and Blast2GO programmes – were used for the *OsSCE1* gene characterisation. The putative *OsSCE1* gene isolated from IR64 was then verified by sequencing, and the gene was 585 bp long and showed 99% identity with *O. sativa* on chromosome 10. *In silico* analysis concluded the gene is involved in abiotic stress regulation. The 20 bp sgRNA was designed manually with the aid of gRNA prediction programmes including CCTop, and Benchling. The virtual vector was validated using the Golden Gate Cloning approach and later confirmed through sequencing. The assembly involved separate vectors containing the *OsSCE1* sgRNA sequence, plant selectable marker, and Cas9 cassette to construct standardised elements for hierarchical modular cloning (MoClo). This study demonstrated that our format, as the gene insertion are achievable, resulting in a speedier and more efficient assembly process which may contribute to improve drought tolerance in *indica* rice. Further study on the *Agrobacterium*-mediated transformation using the developed construct may be conducted to determine the efficacy of knocking out candidate genes in improving drought tolerance ability *O. sativa*.

Keywords: CRISPR/Cas9, Drought, Golden gate cloning, Oryza sativa, sgRNA

Introduction

Rice is one of the world's important staple foods, consumed by half of the global population. In taxonomy, rice belongs to the order Poales and the family Poaceae [1]. Cultivated rice can be either *Oryza sativa* (Asian rice) or *O. glaberrima* (African rice) and the genus contains both diploid (2n=24) and tetraploid (2n=48) species. Archaeological and historical sources pointed out that *O. sativa* originated from Southeast Asia and derived from perennial *Oryza rufipogon* and annual *Oryza nivira*. It comes from diphyletic origin and consists of two subspecies, *O. sativa* subsp. *indica* (adapted to the tropics), and *O. sativa* subsp. j*aponica* (adapted to tropical uplands and temperate regions) [2, 3]. Genetically, rice genome studies started when the International

How to cite:
Md Yusof AA, Tamizi AA, Zim NAM, et al. (2023) Development of CRISPR/Cas9 Construct in Rice (*Oryza sativa* subsp. *indica*) Using Golden Gate Cloning Method Towards Drought Tolerance. Journal of Tropical Life Science 13 (2): 257 – 276. doi: 10.11594/jtls.13.02.04.

Rice Genome Sequencing Project (IRGSP) was established, and they successfully released a high-quality finished genome sequenced of *japonica* rice and initially became the first monocot plant that had a complete genome sequence in 2005. In addition, efforts were also made to sequence the *indica* rice using the same whole-genome shotgun sequencing method. Consequently, the highly accurate and public IRGSP sequence of *japonica* rice has opened doors for functional characterisation of the rice genome and allowed the identification of genes underlying complex agronomic traits such as drought-related genes including Arginine decarboxylase (*ADC)*, polyphenol oxidase (*PPO)*, transcription factor AP37 (*AP37)* and rice salt and drought-induced RING finger1 (*OsSDIR1)* genes, and recently the rice SUMO-conjugating enzyme (*OsSCE1)* [4, 5]. This is highly useful in the research of unveiling the biological function of rice genes along with the genetic improvement of rice yield and quality. Due to the rising need, extensive research in genetic, biochemistry and physiology have been done to enable further improvement in rice production.

Despite its huge demand, rice faces a major abiotic threat in the form of climate change - the drought [6]. Drought stress impact on plants can be described by water content reduction, diminished leaf water potential, turgor pressure, stomatal activity and decreasing in cell enlargement and growth. In severe water stress, photosynthetic arrest, metabolism disturbance and eventually, plant death could occur [7]. According to [8], drought significantly decreases the agronomic traits of rice, with the largest decreases in biomass and yield. Previous study by [6], indicated that yield reduction of drought susceptible mega rice variety IR64 (*O. sativa* subsp. *indica*) was recorded to be more than 85% as compared to under normal well-watered conditions. Therefore, there is a growing need for drought tolerance in rice, a trait that allows the crop to withstand water deficit conditions [9]. This can be made possible through diverse methods and technologies, including genome editing technology [10].

The advancement of genome editing technology i.e., zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) provides precision in targeting any gene of interest for crop improvement programmes. Recently, the CRISPR/Cas system has been discovered to induce DNA double-strand breaks (DSB) that stimulate error-prone nonhomologous end joining (NHEJ) or homology-directed repair (HDR) at specific genomic locations [11]. For instance, the CRISPR/Cas system has been previously applied to improve many crops such as tomatoes and also rice [12–14]. Hence, utilizing genome editing technology such as CRISPR/Cas system, the manipulation of rice genes to create drought tolerance can be achieved. In addition, CRISPR system is considered highly advantageous compared to other genome editing technologies due to its ability to target any gene locus by just replacing 20 to 25 nucleotides (nts) using single guide RNA (sgRNA) sequence and has been used in not only plants but also other eukaryotic organisms including humans and mice [15–17]. Other than that, the system is efficient since Cas9 enables the targeting of multiple genomic loci simultaneously by co-delivering a combination of sgRNAs to the cells of interest [18].

The *OsSCE1* gene is categorised as a *SUMO Conjugating Enzyme* (*SCE*) and acts as one of the major stress proteins. SUMO refers to Small Ubiquitin like Modifiers and the enzyme is involved in SUMOylation which is a multistep process mediated by E1 (SUMO activating enzyme), E2 (SUMO-Conjugating Enzyme or SCE) and E3 (SUMO ligase) enzymes [19]. SUMOylation is essential in plants for development, hormone signalling, light regulation, flowering time, biotic and abiotic stress responses [20]. The functional characterisation of SUMOylation had been previously performed in *Arabidopsis thaliana* and was recently studied in monocot plant (rice). [5] conducted gene functional analyses through gene knockdown and over-expression studies and it was evident that the knockdown of the *OsSCE1* gene could serve as a solution to drought stress in *japonica* rice. Thus, in this current study, we developed a CRISPR/Cas9 expression vector to target the *OsSCE1* gene, particularly for *O. sativa* subsp. *indica* cv. IR64 in hopes to confer drought tolerance in the cultivar.

Material and Methods

OsSCE1 gene prediction and annotation

Oryza sativa subsp. *indica* gene prediction was conducted using several web-based interface gene prediction software namely FGENESH (http://www.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind) and GeneMarkS (http://exon.gatech.edu/GeneMark/genemarks.cgi) [21, 22]. Then, homology search was conducted with BLAST DIAMOND and proceeded with the BLAST2GO tool by utilising OmicsBox. The OmicsBox software allows for functional gene annotation utilities such as high throughput BLAST, Mapping, InterProScan, and Gene Ontology (GO) annotation [23, 24].

InterPro-Scan

InterPro annotations in OmicsBox software allows for the retrieval of domain/motif information in a sequence-wise manner. InterProScan was executed and run in OmicsBox via the public web service at EBI (https://www.ebi.ac.uk/). The public EMBL-EBI InterPro web-service search the sequences against InterPro's signatures. Member databases available were all selected and the InterProScan was saved in XML file format.

Gene Ontology (GO) Mapping and Annotation using BLAST2GO

Mapping refers to the process of retrieving GO terms associated to the Hits obtained by the BLAST search. Mapping was run and assigned to the BLAST results in the OmicsBox interface. Graph drawing configuration was done for three different GO categories; biological process, molecular function and cellular function and the mapping statistics graphs were saved. The GO annotation refers to the process of selecting GO terms from the GO pool obtained by the Mapping step and assigned them to the protein query sequences, using the most specific annotations with a certain level of reliability. The annotation was performed using default annotation configuration.

Validation of OsSCE1 gene by Polymerisation Chain Reaction (PCR) and Sequencing

The putative sequence of the *OsSCE1* gene was validated using young leaves of *O. sativa* subsp. *indica* var. IR64 as plant material. The CTAB (cetyltrimethylammonium bromide) method was used to extract all plant materials [25–28]. PCR amplifications were performed in a total volume of 25µl containing 1µl of DNA template, 1.5µl of 10µM forward and reverse primers, 12µl of Q5 High-Fidelity 2X Master Mix and 9 µl of nuclease-free water. PCR was initiated by initial denaturation step at 95^{o}C for 3 minutes, followed by 35 cycles of 94^{o}C FOR 30 seconds, 59^{o}C for 45 seconds, 72^{o}C for 1 minute and a final extension at 72^{o}C for 10 minutes. To ensure the amplification's specificity, the gene specific primer (GSP) pair was created from the predicted sequence of the *OsSCE1* gene and it was used to amplify the extracted DNA. The primer's forward and reverse sequences are 5'-CCTGCTACCACTACAACGCT and 5'-GGAGTCACGGGCACAGTTAT, respectively. Amplification products were electrophoretically separated on 1.2% agarose gels with 1X TAE buffer, stained with GelRed® loading dye, and photographed under UV light. A 1 kb DNA ladder was used to estimate the molecular size of the fragments. Then, the PCR products were purified and sent for sequencing.

sgRNA Design

A specific sequence within the gene was chosen to represent the guide RNA (gRNA) to represent the gene of interest (GOI), *OsSCE1*. Selection of target region in the gene sequences is done by selecting a 20 bp within *OsSCE1* gene, on either sense or antisense DNA strand. This selected 20 bp target regions must be located in the first exon of the gene, have a minimum melting temperature (T_m) of 58°C and contain 5' GNNNN NNNNN NNNNN NNNNN NGG 3'. Besides, since Cas9 will cut towards the 3' end of the target, it must overlap with a restriction site in the target sequence. The sgRNA forward and reverse primers were generated manually as well as using the sgRNA prediction tools i.e., CC-Top and Benchling.

CCTop - CRISPR/Cas9 Target Online Predictor

CCTop (https://cctop.cos.uni-heidelberg.de:8043/), enables the users to select specific gRNA predictions based on protospacer adjacent motifs (PAM) site, target organisms, target length, promoters, and other factors [29, 30]. In order to predict the sgRNAs within the *OsSCE1* gene, the putative *OsSCE1* genomic

sequence was uploaded, the *Arabidopsis* U6 promoter (U6 or AtU6) was chosen, and most of the default settings were applied. The generated sgRNAs with scoring were downloaded in a spreadsheet format.

Benchling

Benchling (https://benchling.com/) provides a variety of utilities in their informatics platform, including vector construction and gRNA design [31]. The GOI sequence was uploaded to the programme in FASTA format by using the parameters, that were set to 'single guide' with 20 guide length and PAM site 'NGG (SpCas9,3'side)'. The target sequences were then selected.

Vector Construction

The vector construction was executed and visualised using SnapGene and Benchling. Besides, gRNA design, both programmes allow for *in silico* molecular cloning, such as Golden Gate Cloning. Additionally, the vectors used and constructed can be easily visualised in the form of both linear and circular plasmid. Using the selected gRNA, Level 1 (pICLS01009::AtU6p) was assembled into the destination vector pICH47751 with the use of *BsaI* restriction enzyme. Next, Level 2 was assembled (pICH47732::NOSp::NPTII-OCST, pICH47742:: 35Sp::Cas9-NOST, pICH47751::AtU6::sgRNA, pICH41766) into the destination vector (pAGM4723). The cut-ligation reaction was performed using *BpiI (BbsI)*.

Validation of Constructed Vector

As in the Golden Gate Cloning method, there are three modular cloning levels of plasmid assembly that need to be done accordingly to insert the sgRNA sequence of the GOI along with Cas9 enzymes. Each of the levels was validated using PCR approach and sequencing to ensure the sequence obtained matched with the designed gRNA as well as the sequence of the Cas9 enzymes.

Results and Discussion

OsSCE1 Gene Prediction and Annotation

Gene prediction is a method of gene discovery and is intended for genome-wide annotation and discovery research [32, 33]. To our knowledge, there has been limited literature involving the *OsSCE1* gene, particularly in *O. sativa* subsp. *indica*. To fill this gap, gene prediction and annotation were carried out to identify the *OsSCE1* gene sequence in *indica* subspecies and its underlying function. In this study, *ab initio* approach was used to predict genes directly from nucleotide sequences i.e., FGENESH and GeneMarkS [21, 22]. The program uses statistical models to differentiate the promoter, codon or noncoding regions, as well as intron-exon junctions in genomic sequences. For instance, most common model used in these programmes is the Hidden Marcov Model (HMM) [34]. Based on the outcomes, predicted genes from FGENESH were selected as trustworthy and proven, with a total hit of 3,853 sequences.

Following that, further analysis of the predicted genes proceeds with BLAST. Regarded as one of the most common and efficient bioinformatics tools ever developed, BLAST allows for homology search, enabling users to gain insights into the function and regulatory signals of gene sequences. The 3,853 query sequences were blasted using DIAMOND, a program deemed best for high-throughput settings to find high-scoring segment pairs (HSP). In a comparison study involving another widely known sequence comparison tool, BLASTX, it was revealed that an analysis which took one month with BLASTX took only a few minutes with DIAMOND, indicating the high level of efficiency [35].

Out of a total of 3,853 query sequences, 2674 (69.4%) blast hits were obtained. Prior to Gene Ontology (GO) annotation, the query sequence was subjected to InterProScan. Using the InterPro database, it provides an integrative classification of protein sequences into families while identifying functionally important domains and conserved sites. InterPro integrates 13 protein signature databases into one central resource, such as PANTHER, CATH-Gene3D, and SUPERFAMILY [36]. Most of the query sequences were categorised within the protein kinase domain (IPR000719). The successfully blasted sequences were loaded into Blast2GO for annotation including GO data.

GO provides a structured controlled vocabulary composed of terms which describe gene and protein biological roles. As the function of the activity of a protein can be defined at different levels, GO can be further categorised into three different aspects: molecular function,

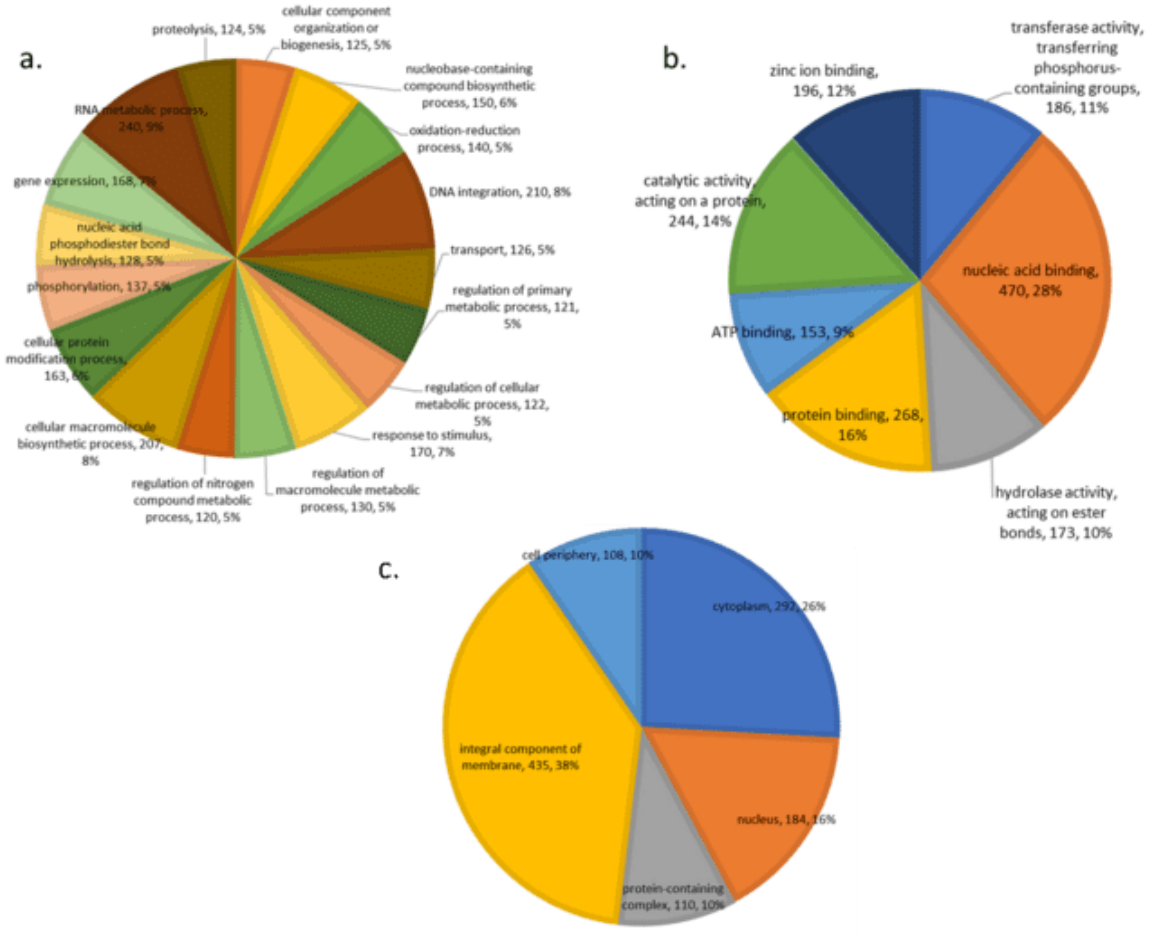


Figure 1. Gene classifactions generated using BLAST2GO software; a. Biological process, b. molecular function and c. cellular components.

biological process and cellular component. This three-way division is based on the following notions: each protein has elementary molecular functions that are usually independent of the environment, like binding or catalytic activities; sets of proteins interact and are involved in cellular processes such as signal transduction, metabolism or RNA processing; and protein could act in different cellular localizations, such as the membrane or nucleus [37]. The query sequences loaded into OmicsBox were successfully assigned based on the Gene Ontology Consortium in biological process (Figure 1a), molecular function (Figure 1b) and cellular components (Figure 1c).

In Figure 1a, the biological process term graph of GO classification displays RNA metabolic process (240 genes) as the most dominant process within the sequences, followed by DNA integration (210 genes), cellular macromolecule biosynthesis process (207 genes), response to stimulus (170), and gene expression (168). For molecular function illustrated in Figure 1b, the query sequences were classed into 7 categories: transferase activity (transferring phosphorus-containing groups), nucleic acid binding, hydrolase activity (acting on ester bonds), protein binding, ATP binding, catalytic activity (acting on a protein) and zinc ion binding. Among them, nucleic acid binding was shown to be the dominant molecular function with 28% (470 genes). The rest of the functions were relatively occurring in similar percentages like protein binding (16%) and catalytic activity (14%).

While in Figure 1c, the sequences were distributed according to its cellular component, which consisted of 5 classes: cell periphery, cytoplasm, nucleus, protein-containing complex, and integral component of membrane. The majority of the sequences fell into an integral component of membrane with the highest percentage of 38% (435), followed by cytoplasm (292), nucleus (184), protein-containing complex (110), and cell periphery (108).

Validation of OsCE1 Gene

Among the 2,674 blast hits, *OsSCE1*, encoding the E2 type SUMO conjugating enzyme gene sequence was predicted (Supplementary 1). The candidate gene was identified to be involved in ubiquitin-conjugating enzyme activity (E2) with the GO ID identified as GO:0061631. This finding corresponds to the characteristics of *OsSCE1* as mentioned by [5]. The E2 is one of the key enzymes in the larger ubiquitin-proteasome system which is associated to abiotic stresses reported in several plants including mung bean, rice, *Arabidopsis* and tobacco [38]. Other than that, the candidate sequence was categorised with the enzyme commission number, E.C:2.3.2.23, also known as the E2 ubiquitin-conjugating enzyme. This finding further confirms the identity of the candidate gene.

Additionally, the candidate gene was also identified to be involved with other GO names such as protein polyubiquitination, ubiquitin-dependent protein catabolic process, ATP binding, and nucleus with the GO IDs GO:0000209, GO:0006511, GO:0005524, and GO:0005634 respectively (Table 1). The verification process was further extended by cross referencing the selected sequence within the NCBI (refseq_protein) database, which yielded 93% identity match to ubiquitin-conjugating enzyme E2 4 (*Oryza sativa japonica* Group, XP_015614603.1). Together, the present findings confirm the candidate gene as the putative *OsSCE1* gene with 585 bp long and encodes for 195 amino acids.

The predicted *OsSCE1* gene has been validated and isolated from IR64 by experimental laboratory through DNA extraction, PCR amplification and sequencing procedures. As mentioned in the methodology, the CTAB method has been used for DNA extraction of *O. sativa* subsp. *indica* var. IR64. This method has been proven by various research where DNA concentration yielded by this method is higher compared to other methods even by using DNA extraction kits [25–28]. In addition, according to [39] this procedure is inexpensive, straightforward, high throughput and PCR compatible.

DNA concentration and its purity are crucial parts for many applications in molecular biology. Impurities in nucleic acid can cause erroneous measurements of DNA concentration and could restrict subsequent labelling responses. In this study, concentration and purity of nucleic acids were measured using NanoDrop™ spectrophotometer, which was evaluated by the ratio of the absorbance at 260nm and 280nm (A_{260}/A_{280}) and 260nm and 230nm (A_{260}/A_{230}). The concentrations of nucleic acid for new leaves 1 (NL1), new leaves 2 (NL2), and new leaves 3 (NL3) were 1916.0 ng/l, 2813.9 ng/l, and 1464.6 ng/l, respectively. These plant samples were harvested with three replicates and were labelled as NL1, and NL2 and NL3. Ratios of A_{260}/A_{280} and A_{260}/A_{230} are used to determine the purity of the extracted DNA and it can be affected by the presence of contaminants in the biological samples or the chemicals used in the extraction process like CTAB [40–42]. The reading of ratio A_{260}/A_{280} for samples NL1, NL2, and NL3 are 2.12, 2.13, and 2.14, respectively, and the reading of ratio A_{260}/A_{230} for samples NL1, NL2, and NL3 are 2.32, 2.26 and 2.23, respectively. [43, 44] reported that ~2.0 reading and above are considered as common measurement and acceptable range for DNA sample purity and all the readings recorded were above 2.0 reading, which reading for ratio A_{260}/A_{230} are slightly higher compared to ratio A_{260}/A_{280} for all three samples.

Table 1. Biological description of putative *OsSCE1* gene

Sequence ID	Description	Length (amino acids)	GO IDs	GO Names	Enzyme Codes
FGENESH_25 886exon(s)	ubiquitin-conjugating enzyme E2 4	195	P:GO:0000209 P: GO:0006511 F:GO:0005524 F:GO:0061631 C: GO:0005634	P: protein polyubiquitination; P: ubiquitin dependant protein catabolic process; F: ATP binding. C: nucleus	E.C:2.3.2.23

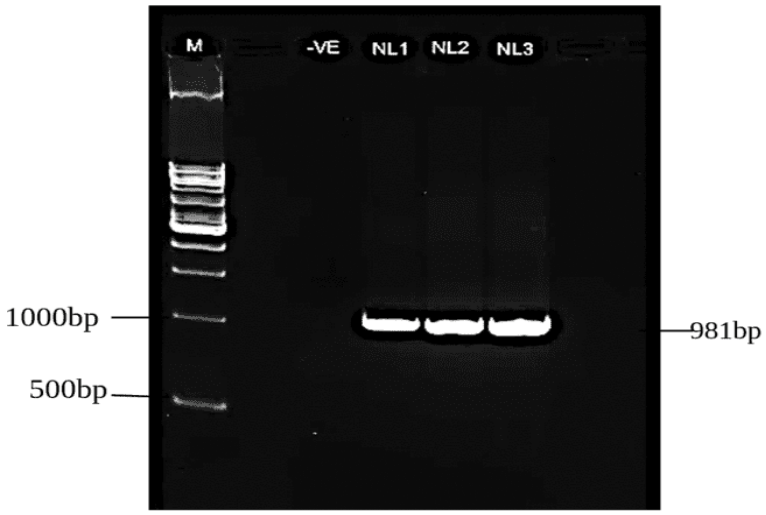


Figure 2. Polymerase chain reaction (PCR) analysis of genomic DNA extracted from *Oryza sativa* subsp. *indica* var. IR64 using gene specific primer. Lane 1, 1kb DNA marker (M); lane 3, non-template control (-VE); lanes 4, 5 and 6, plant samples (NL1, NL2, and NL3).

Table 2. Description of the selected sgRNA sequence for *OsSCE1* gene

gRNA tools	*OsSCE1* gene sequence	Labelled	Sequence	GC Content (%)
Benchling (manually)	From NCBI validated	SG1	GTCTCAGACCTCCACTCTGTC**GGG**	57.1
CCTop	From NCBI validated	SG2	CACGGTGCGCCTGCGGCAAG**CGG**	65.0
Benchling (manually)	From putative *OsSCE1* gene	SG3	GCTCATGATGTCCGACTACA**AGG**	50.0

These three samples were proceeded with PCR analysis by using a gene specific primer (GSP) pair and the PCR products were separated electrophoretically along with 1kb DNA ladder (Figure 2). The fragments size obtained for three samples were similar which were 981 bp and the sequences obtained from sequencing were shown in Supplementary 2.

The amplicon intensity on the gel and spectrophotometer results led to the selection of NL2 for sequencing since it had the highest DNA concentration for further experiments. The primer pair that was created from the predicted sequence of the *OsSCE1* gene was used to amplify the homologous gene from IR64 using forward and reverse primer; 5'-CCTGCTACCACTACAACGCT and 5'-GGAGTCACGGGCACAGTTAT. The sequence obtained was confirmed with NCBI blast. The sequence blast result showed the closest hit with *O. sativa* subsp. *indica* Group cultivar Teqing SUMO E2 conjugating enzyme SCE1- like protein (Os10g0536000) with an identity value of 99%.

sgRNA Sequence Design

As plants require a long generation time, it is critical to select an efficient sgRNA that could effectively target the desired site. This study aims to generate gene knockout by disrupting exons that are shared by all transcript variants of the *OsSCE1* gene. In order to do so, targeting functional protein domains is necessary to result in the loss-of-function mutations. Once the selected sgRNA targets the protein-coding region, it results in frame-shift mutation and subsequent premature stop codons, leading to mRNA elimination by nonsense-mediated mRNA decay [46]. Furthermore, it is necessary to target the early exon in the putative *OsSCE1* gene in order to ensure frame-shift mutation and no protein is going to be translated.

The *CRISPR/Cas9* system could be programmed to virtually cleave any sequence preceding a 5'-NGG-3' PAM sequence. Nevertheless, the success rate is not always guaranteed with regard to all sites predicted to be targeted [47]. Therefore, to overcome this

obstacle, the use of gRNA prediction tools was incorporated to design an efficient sgRNA. In this study, the sgRNA sequence was predicted using gRNA prediction tools, namely, CCTop and Benchling. The top hit sgRNA sequence demonstrating the high-efficiency score was retrieved and analysed. The candidate sgRNA sequence was selected based on its respective efficiency score to reduce potential off-targets.

The sgRNA must contain the following sequence pattern: "GNNNNNNNNNNNNNNNNNNN". The finding suggests that since the sequence was close to the predicted sgRNA sequence site from CCTop and Benchling, the manual site could be an excellent choice of sgRNA in this study. The selected sgRNA sequence had a PAM sequence bolded nucleotide base in Table 2. Besides the sgRNA sequence adhered to a characteristic essential to sgRNA effectiveness in plants by having a G/C composition of greater than 50%. According to [47], 97% of sgRNA have a G/C content between 30% and 80%, indicating it as an important efficiency criterion.

Vector Construction

The vector construction was first executed virtually using Benchling, which applied the concepts of the Golden Gate Cloning method. The cloning method relies on the use of type II restriction enzyme and allows for the assembly of several DNA fragments in a defined linear order in a vector in a single step. In addition, this cloning technique is characterized by great cloning efficiency and the ability to ligate up to 24 fragments in a single cloning procedure [48]. In the DNA assembly prior to Level 1, the amplification of sgRNA is required, where the sgRNA needs to be amplified with pICH86966::AtU6p::sgRNA_PDS construct as a template in Figure 3, as this construct carry guide RNA scaffold for CRISPR/Cas9 systems. This procedure aimed in producing single guide RNA (sgRNA) as a guide for Cas9 endonuclease activity, as it contains 5'-NGG-3' PAM sequence that can reduce the off-target site.

Guide RNA designed in forward direction (20 to 25 bp) was annealed to the scaffold and amplified producing PCR product of size almost to 200 bp (Figure 3). The obtained results were in the range of the expected band size, which is in between 150 to 200 bp, thus, it indicates that the

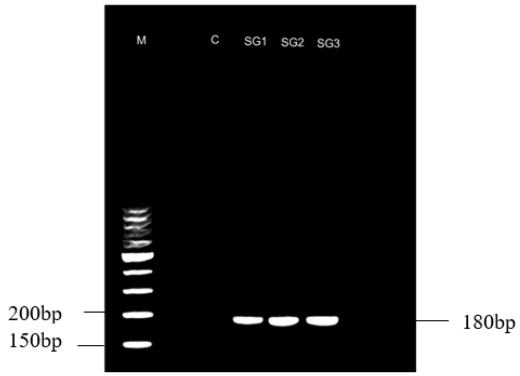


Figure 3. Polymerase chain reaction (PCR) analysis of Level 0 CRISPR. Lane 1, 100bp DNA marker (M); lane 3, non-template control (C); lane 4, 5, and 6, gRNAs labelled SG1, SG2, and SG3, respectively, amplified with pICH86966::AtU6p:: sgRNA_PDS construct.

three gRNAs were successfully annealed with the scaffold in the pICH86966::AtU6p::sgRNA_PDS construct.

Then, the resulting PCR product and vector pICSL01009::AtU6p (SpecR) were delivered into Level 1 destination vector pICH47751, producing Level 1 AtU6p::sgRNA, with the sgRNA placed under the *Arabidopsis* U6 promoter. The cut-ligation reaction utilised the *BsaI* restriction enzyme and T7 DNA ligase. Once the cut-ligation process was done, the products were incubated overnight in 4^{o}C before being transformed into the competent cell *E. coli* strain DH5α. The mutation of *recA1* and *endA1* in competent cells DH5α functions to support blue-white screening since the destination vector pICH47751 contain LacZ alpha gene fragment that is used for the screening of transformant colony. The mutation of those genes can increase the stability and promotes better transformation efficiencies [49]. Figure 4 showed single colonies that grew after the transformation product, with different volumes, on Luria Bertani (LB) media supplemented with carbenicillin antibiotics and X-gal for blue-white screening. As a result, the colour of the colonies will serve as a marker for the existence of the GOI [50]. All of the plates showed many white colonies and only a few blue colonies appeared after being incubated overnight at 37^{o}C. As white colony indicated that the colony contained construct with insert (recombinant) while blue colony contained construct without insert [50].

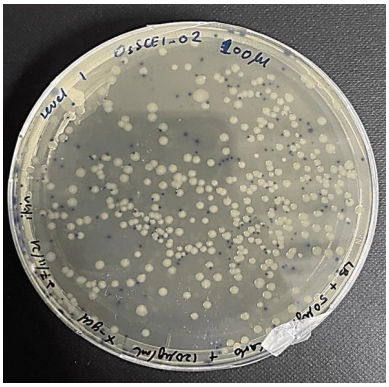


Figure 4. Single colonies on LB media supplemented with carbenicillin antibiotics and X-gal. LB agar plate was spread with 100µl of DH5α cells transformed with ligation products of CRISPR Level 1 constructs.

Figure 5. Bacterial colonies formed after transformation of CRISPR Level 2 vector and incubated at 37°C overnight in the

The single white colonies were picked and proceeded with plasmid purification and PCR-analysed using the same forward and reverse primers used in PCR of CRISPR plasmid Level 0. The purified PCR products were separated on 1.2% 1X TAE gel electrophoresis for 65 minutes at 100V. Then, all of the PCR products were sent for sequencing to confirm the insert (Supplementary 3-6). After sequencing result analysis, three plasmids were picked and proceeded with cut-ligation for CRISPR Level 2.

The level 2 assembly was executed by introducing Level 1 construct (pICH47732::NOSp ::NPTII-OCST, pICH47742:: 35Sp::Cas9-NOST, pICH47751::AtU6::sgRNA, pICH41766) into Level 2 destination vector pAGM4723. The cut-ligation reaction was done using *BpiI* (*BbsI*), producing Level 2 (NPTII-Cas9-sgRNA). Like CRISPR Level 1, cut-ligated product was then transformed in the competent cells, DH5α, yet the difference for CRISPR Level 2, the marker used for this level exhibit different colour of the colonies as well as the selective antibiotics used is kanamycin [51]. The destination vector pAGM4723 contain selectable marker known as CRed, which is an artificial bacterial operon that responsible for canthaxanthin biosynthesis [51]. Therefore, red-white colonies (instead of blue-white colonies) were observed on the LB media supplemented with kanamycin (Figure 5). Similar to Level 1, the white colony was selected as it should carry the intended inserts, to proceed with the confirmation by sequencing. However, the number of white colonies recovered for Level 2 assembly was much lower compared to that of in Level 1 *E. coli* transformation plates. This was mostly due to more plasmids self-ligated in Level 2 step, therefore the rate of successful ligation of the intended insert reduced.

Since the amount of white colony is limited, hence only one white colony was picked for each plate and proceeded with the plasmid purification and amplified the purified plasmids were PCR-analysed with forward (5'-GGGATGAC GCACAATCCCAC-3') and reverse (5'-TATGCGCCAGCGCGAGATAG-3') primers designed for Level 2. The primer pair designed in this level is meant to amplify the region between 35S promoter and Cas9 regions. In Figure 6, the PCR product were separated electrophoretically on 1.2% 1X TAE gel electrophoresis for 65 minutes at 100V. The PCR products for sample SG1.201 (one single-white colony from SG1 plates with 20µl spread of bacterial broth) was

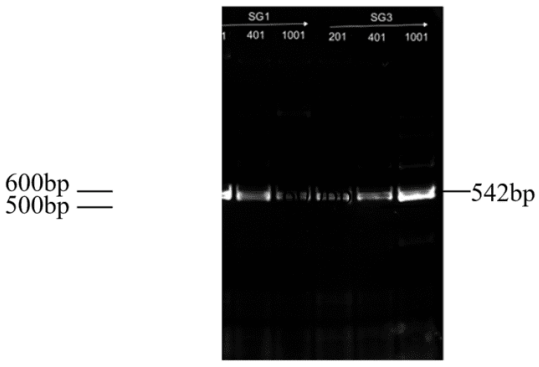
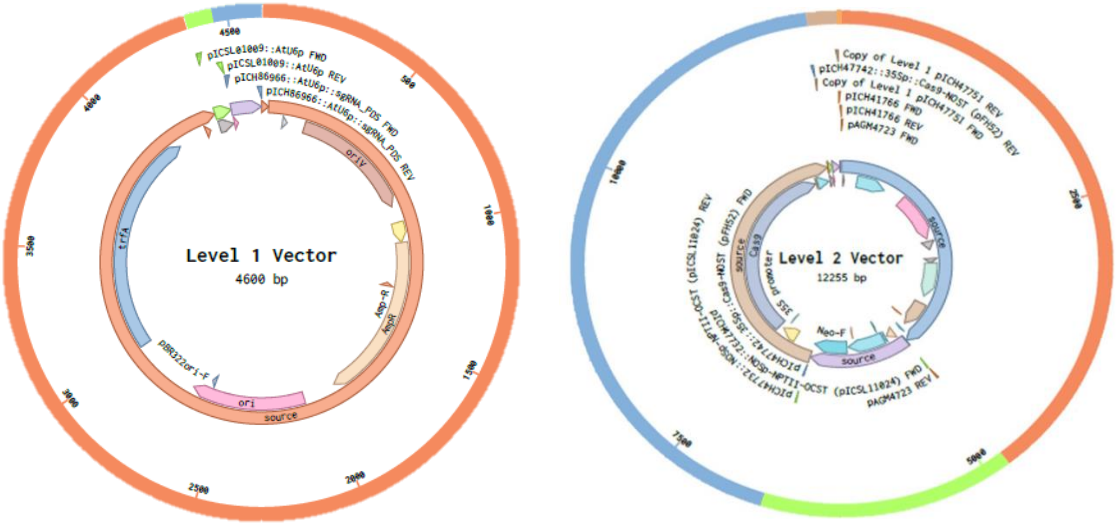


Figure 6. Polymerase chain reaction (PCR) analysis of Level 2 CRISPR. Lane 1, 100bp DNA marker (M); lane 2, non-template control (-VE); lane SG1 (201, 401, and 1001) and SG3 (201, 401, 1001), plasmid of white colonies amplified with 35S-Cas9 primer.



a **b**

Figure 7. Circular maps of: (a) Level 1 vector pICH47751::AtU6::sgRNA, (b) Level 2 Vector (NPTII-Cas9-sgRNA).

then sent for sequencing Cas9 sequence validations (Supplementary 7-8).

The resulting vectors of Level 1 (4600 bp) and Level 2 (12255 bp) intended for *OsSCE1* gene knockout are visualised in the form of a circular plasmid map in Figure 7a and Figure 7b, respectively. Together, with the successful inclusion of *OsSCE1* sgRNA sequence within the Level 2 vector, it marks the completion of CRISPR construct for *Oryza sativa* subsp. *indica* cv. IR64 targeting *OsSCE1* gene through the Golden Gate Cloning method.

Conclusion

The CRISPR/Cas9 construct is a novel method for systematically characterising and comprehending the molecular function of genes. The detailed generation CRISPR/Cas9 construct for *OsSCE1* gene in *O. sativa* subsp. *indica* was successfully completed and represented in this study. This study showed that our approach makes it possible to incorporate such inclusions, leading to a quicker and more effective assembly process, which help to increase rice's tolerance toward stresses. The usefulness of proposed gene knockouts in *O. sativa* for drought tolerance will be further investigated during the *in vivo* stage to ensure a successful outcome.

Acknowledgment

The abstract of this work was presented during the International Conference of Bioscience and Medical Engineering (ICBME 2022). This work is supported by the Ministry of Higher Education, Malaysia FRGS/1/2019/STG05/UIAM/03/8. We also like to express our gratitude to the International Islamic University Malaysia for its technical support.

References

1. Nagata T, Hosaka-Sasaki A, Kikuchi S (2016) The evolutionary diversification of genes that encode transcription factor proteins in plants. In: Gonzalez DH, eds. Plant Transcription Factors. Academic Press. 73-97.
2. Sasaki T (2001) Rice genome analysis to understand the rice plant as an assembly of genetic codes. Photosynthesis research 70 (1): 119-127. doi:10.1023/A:1013844411049.
3. Nair KP (2019) Utilizing crop wild relatives to combat global warming. Advances in Agronomy 153: 175-258. doi: 10.1016/bs.agron.2018.09.001.
4. Oladosu Y, Rafii MY, Samuel C et al. (2019) Drought Resistance in Rice from Conventional to Molecular Breeding: A Review. International Journal of Molecular Sciences 20 (14). doi: 10.3390/ijms20143519.
5. Nurdiani D, Widyajayantie D, Nugroho S (2018) *OsSCE1* Encoding SUMO E2-Conjugating Enzyme Involves in Drought Stress Response of *Oryza sativa*. Rice Science 25 (2): 73–81. doi: 10.1016/j.rsci.2017.11.002.
6. Salleh MS, Nordin MS, Shahari R et al. (2021) Response of primed rice (*Oryza sativa*) seeds towards reproductive stage drought stress. Sains Malaysiana 50 (10): 2913–2921. doi: 10.17576/jsm-2021-5010-06.
7. Singh CM, Kumar B, Mehandi S, Chandra K (2012) Effect of Drought Stress in Rice: A Review on Morphological and Physiological Characteristics. Trends in Biosciences 5 (4): 261–265.
8. Zhang J, Zhang S, Cheng M et al. (2018) Effect of drought on agronomic traits of rice and wheat: A meta-analysis. International Journal of Environmental Research and Public Health 15 (5). doi: 10.3390/ijerph15050839.

9. Sahebi M, Hanafi MM, Rafii MY et al. (2018) Improvement of Drought Tolerance in Rice (*Oryza sativa* L.): Genetics, Genomic Tools, and the WRKY Gene Family. Biomed Research International 2018: 1-20. doi: 10.1155/2018/3158474.
10. Bi H, Yang B (2017) Gene editing with TALEN and CRISPR/Cas in rice. Progress in molecular biology and translational science. 149: 81-98. doi: 10.1016/bs.pmbts.2017.04.006.
11. Gaj T, Gersbach CA, Barbas III CF (2013) ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology (31) 7. doi: 10.1016/j.tibtech.2013.04.004.ZFN.
12. Zhou H, Liu B, Weeks DP et al. (2014) Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Research 42 (17): 10903–10914. doi: 10.1093/nar/gku806.
13. Uluisik S, Chapman NH, Smith R et al. (2016) Genetic improvement of tomato by targeted control of fruit softening. Nature Biotechnology 34 (9): 950–952. DOI: 10.1038/nbt.3602.
14. Zainuddin Z, Mohd-Zim NA, Azmi NSA et al. (2021) Genome editing for the development of rice resistance against stresses: A review. Pertanika Journal of Tropical Agricultural Science 44 (3): 599–616. doi: 10.47836/pjtas.44.3.06.
15. Fujihara Y, Ikawa M (2014) CRISPR/Cas9-based genome editing in mice by single plasmid injection. In: Doudna JA, Sontheimer EJ, eds. Methods in Enzymology. Academic Press. 546: 319–336. doi: 10.1016/B978-0-12-801185-0.00015-5.
16. Shen B, Zhang J, Wu H et al. (2013) Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell Research 23 (5): 720–723. doi: 10.1038/cr.2013.46.
17. Horvath P, Barrangou R (2013) RNA-guided genome editing à la carte. Cell Research 23 (6): 733–734. doi: 10.1038/cr.2013.39.
18. Ran FA, Hsu PD, Wright J et al. (2013) Genome engineering using the CRISPR-Cas9 system. Nature Protocols 8 (11): 2281–2308. doi: 10.1038/nprot.2013.143.
19. Nigam N, Singh A, Sahi C et al. (2008) SUMO-conjugating enzyme (*Sce*) and FK506-binding protein (FKBP) encoding rice (*Oryza sativa* L.) genes: Genome-wide analysis, expression studies and evidence for their involvement in abiotic stress response. Molecular Genetics and Genomics 279 (4): 371–383. doi: 10.1007/s00438-008-0318-5.
20. Rosa MTG, Almeida DM, Pires IS et al. (2018) Insights into the transcriptional and post-transcriptional regulation of the rice SUMOylation machinery and into the role of two rice SUMO proteases. BMC Plant Biology 18 (1): 1–18. doi: 10.1186/s12870-018-1547-3.
21. Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Research 29 (12): 2607–2618. doi: 10.1093/nar/29.12.2607.
22. Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biology 7 (1): S10. doi: 10.1186/gb-2006-7-s1-s10.
23. Hasan S, Huang L, Liu Q et al. (2022) The Long Read Transcriptome of Rice (*Oryza sativa* ssp. *japonica* var. *Nipponbare*) Reveals Novel Transcripts. Rice 15 (29): 1–17. doi: 10.1186/s12284-022-00577-1.
24. Zuo Y, Verheecke-Vaessen C, Molitor C et al. (2022) *De novo* genome assembly and functional annotation for *Fusarium langsethiae*. BMC Genomics 23 (158): 1-10. doi: 10.1186/s12864-022-08368-0.
25. Yu C, Cohen LH (2004). Tissue sample preparation - Not the same old grind. https://www.chromatographyonline.com/view/class-is-back-in-session-more-questions-on-your-extraction-knowledge. Accessed date: November 2021.
26. Psifidi A, Dovas CI, Bramis G, et al. (2015). Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and long-term DNA banking using blood samples. PLoS one. 10 (1): e0115960. doi: 10.1371/journal.pone.0115960.
27. Zhang G, Weiner JH (2000) CTAB-mediated purification of PCR products. BioTechniques 29 (5): 982–986. doi: 10.2144/00295bm11.
28. Chen S, Borza T, Byun B et al. (2017). DNA markers for selection of late blight resistant potato breeding lines. American Journal of Plant Sciences. 8 (6): 1197–1209. doi: 10.4236/ajps.2017.86079.
29. Stemmer M, Thumberger T, del Sol Keyer M, Wittbrodt J, Mateo JL (2015) CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. PLOS ONE 10 (4): e0176619. doi: 10.1371/journal.pone.0124633.
30. Labuhn M, Adams FF, Ng M et al. (2018) Refined sgRNA efficacy prediction improves largeand small-scale CRISPR-Cas9 applications. Nucleic Acids Research 46 (3): 1375–1385. doi: 10.1093/nar/gkx1268.
31. Doench JG, Fusi N, Sullender M et al. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotechnology 34 (2): 184–191. doi: 10.1038/nbt.3437.
32. Allen JE, Pertea M, Salzberg SL (2004) Computational gene prediction using multiple sources of evidence. Genome Research 14 (1): 142–148. doi: 10.1101/gr.1562804.
33. Wang Z, Chen Y, Li Y (2004) A brief review of computational gene prediction methods. Genomics, Proteomics bioinformatics 2 (4): 216–221. doi: 10.1016/S1672-0229(04)02028-5.
34. Yu Y, Santat LA, Choi S (2006) Bioinformatics packages for sequence analysis. Applied Mycology and Biotechnology 6: 143–160. doi: 10.1016/S1874-5334(06)80009-2.
35. Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. Nature Methods 12 (1): 59–60. doi: 10.1038/nmeth.3176.
36. Blum M, Chang HY, Chuguransky S et al. (2021) The InterPro protein families and domains database: 20 years on. Nucleic Acids Research 49 (D1): D344–D354. doi: 10.1093/nar/gkaa977.
37. Yu B, Hinchcliffe M, eds. (2011) In Silico Tools for Gene Discovery. 1st Edition. New Jersey: Humana Totowa.
38. Liu W, Tang X, Qi X et al. (2020) The Ubiquitin Conjugating Enzyme: An Important Ubiquitin Transfer Platform in Ubiquitin-Proteasome System. International Journal of Molecular Science 21 (8): 2894. doi: 10.3390/ijms21082894.

39. Rani B, Sharma VK (2016) A Modified CTAB Method for Quick Extraction of Genomic DNA from Rice Seed/Grain/Leaves for PCR Analysis. International Journal of Science and Research Methodology 4 (4): 254–260.
40. Moreira PA, Oliveira DA (2011) Leaf age affects the quality of DNA extracted from *Dimorphandra mollis* (Fabaceae), a tropical tree species from the Cerrado region of Brazil. Genetics and Molecular Research 10 (1): 353–358. doi: 10.4238/vol10-1gmr1030.
41. García-Alegría AM, Anduro-Corona I, Pérez-Martínez CJ et al. (2020) Quantification of DNA through the nanodrop spectrophotometer: Methodological validation using standard reference material and sprague dawley rat and human DNA. International Journal of Analytical Chemistry 2020: 1-9. doi: 10.1155/2020/8896738.
42. Matlock B (2015). Assessment of Nucleic Acid Purity. https://tools.thermofisher.com/content/sfs/brochures/TN52646-E-0215M-NucleicAcid.pdf. Accessed date: August 2022.
43. Abdel-Latif A, Osman G (2017) Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. Plant Methods 13 (1): 1–9. doi: 10.1186/s13007-016-0152-4.
44. Yeates C, Gillings MR, Davison AD, Altavilla N, Veal DA. Methods for microbial DNA extraction from soil for PCR amplification. Biological Procedures Online 14 (1): 40-47. doi: 10.1251/bpo6.
45. Van Campenhout C, Cabochette P, Veillard AC et al. (2019) Guidelines for optimized gene knockout using CRISPR/Cas9. Biotechniques 66 (6): 295–302. doi: 10.2144/btn-2018-0187.
46. Lucena-Aguilar G, Sánchez-López AM, Barberán-Aceituno C, Carrillo-Ávila JA, López-Guerrero JA, Aguilar-Quesada R (2016) DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. Biopreservation and Biobanking 14 (4): 264–270. doi: 10.1089/bio.2015.0064.
47. Liang G, Zhang H, Lou D, Yu D (2016) Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Scientific Reports 6: 21451. doi: 10.1038/srep21451.
48. Marillonnet S, Grützner R (2020) Synthetic DNA Assembly Using Golden Gate Cloning and the Hierarchical Modular Cloning Pipeline. Current Protocol Molecular Biology 130: e115. doi: 10.1002/cpmb.115.
49. Renzette N (2011) Generation of transformation competent *E. coli*. Current Protocols in Microbiology 22: A3L.1-A3L.5. doi: 10.1002/9780471729259.mca03ls22.
50. Sadeghi S, Ahmadi N, Esmaeili A, Javadi-Zarnaghi F (2017) Blue-white screening as a new readout for deoxyribozyme activity in bacterial cells. RSC Advances 7: 54835–54843. doi: 10.1039/c7ra09679h.
51. Weber E, Engler C, Gruetzner R et al. (2011) A modular cloning system for standardized assembly of multigene constructs. PLOS One 6 (2): e16765. doi: 10.1371/journal.pone.0016765.

Supplementary 1

Putative *OsSCE1* gene (From Rice Genome Chromosome 10)

Nucleotide sequence

ATGTCCTCCCCGTCCAAGCGCGCCGAGATGGACCTCATGAAGCTCATGATGTCCGACTACAAGGTGGAGATGGTG
AACGACGGCATGCAGGAGTTCTTCGTGGAGTTCCGCGGCCCGACCGAGTCCATCTACCAGGGCGGCGTGTGGAAG
GTGCGCGTGGAGCTCCCGGACGCCTACCCGTACAAGTCCCCGTCCATCGGCTTCGTGAACAAGATCTACCACCCG
AACGTGGACGAGATGTCCGGCTCCGTGTGCCTCGACGTGATCAACCAGACCTGGTCCCCGATGTTCGGCGAGATC
ACCCTCGTGCTCGTGATCATCTCCACCGACCTCGTGAACGTGTTCGAGGTGTTCCTCCCGCAGCTCCTCCTCTAC
CCGAACCCGTCCGACCCGCTCAACGGCGAGGCCGCCGCCCTCATGATGCGCGACCGCCCGGCCTACGAGCAGAAG
GTGAAGGAGTACTGCGAGAAGTACGCCAAGCCGGAGGACGCCGGCGTGACCCCGGAGGACAAGTCCTCCGACGAG
GAGCTCTCCGAGGACGAGGACGACTCCGGCGACGACGCCATCCTCGGCAACCCGGACCCG

Amino acid sequence

MSSPSKRAEMDLMKLMMSDYKVEMVNDGMQEFFVEFRGPTESIYQGGVWKVRVELPDAYPYKSPSIGFVNKIYHP
NVDEMSGSVCLDVINQTWSPMFGEITLVLVIISTDLVNVFEVFLPQLLLYPNPSDPLNGEAAALMMRDRPAYEQK
VKEYCEKYAKPEDAGVTPEDKSSDEELSEDEDDSGDDAILGNPDP

Supplementary 2

Sequencing results of *OsSCE1* gene (Isolated from this study)
Nucleotide sequence

PLUS STRAND

CTNCCCCCTCTTCACCCAATAACGCTATGCGCTCGTGGGAGATGAGCCGACGCCGATGCGTCCGTCCCCTCGCCT
CCCTCCCTGCTCGCTCCCTCCAAGGCCAGCCGCCACCATACGTGAGTCTCCCCGCTTCCAGCCACCCACCACCCT
GCCCAGCCGCCTCACCTCCCTCCCTCCGTCCATGTCCACTGCCTCCTCAGCTCCTCTCGCTGGCTCGGCGGCCCG
GCAACAAGGCTCTTTCTGCACAACCCCCCTGCACGGTGCGCCTGCGGCAAGCGGCAGGTGCTCAGAGTCTCAGAC
CTCCACTCTGTCGGGCAAGGAGCGGACGAGGACCTTGACAGGTGAGTTTTTCATAATCTATTTCTGGTATTCCCA
TTTCTCAAATGCCTAAATGGCCTAGATCACCTATAGATCTTCTGATTCCATTCAATTTGTTTTAAACTTAATAGA
GATAGAAGGTGAAGAGGCTTAGCGATGAAATCTATAGATGATTGATGGGGTATAAAGAGAAGAATAAAGATGACC
AGAAACATTTCAAAATGCAATTGAGACGGGTAAAACCATTCTAAATGCTTTCCGTGGAAAAAAAATCCGCCATTT
TTGTTGATGAGTATGATGGAGATGCATCTTTCAAAGAGCAAATAGCTGTCATTTTGAGGTTAGATGCTGTTTCTT
GCGTTGTTGTTTTTTTCCTTTCGAAACTATAATAATTTGTATTGTACTTATTATGTATTAGTGTGGGTCTCATTT
ATGGTATGGGATAGTAGCAATTAGATTTGGCCTAGGGCATGAAGAAATCCTGGGTCCGCCACTGAAGATTACTAT
ATATCCCCCCATATGCAAGATAAAGTGTTTTGGAAGTGATGTTGATTATGAAGACAGTAGTGATGGTGACTATAA
TCCATCTTTCAGCAACTAAATCTTACAACATTCAGATCCTTGTACCCTGTTAATTCAATACTGGCGCCNGGNANC
CAAAAA

MINUS STRAND

TTTTTGGNTNCCNGGCGCCAGTATTGAATTAACAGGGTACAAGGATCTGAATGTTGTAAGATTTAGTTGCTGAAA
GATGGATTATAGTCACCATCACTACTGTCTTCATAATCAACATCACTTCCAAAACACTTTATCTTGCATATGGGG
GGATATATAGTAATCTTCAGTGGCGGACCCAGGATTTCTTCATGCCCTAGGCCAAATCTAATTGCTACTATCCCA
TACCATAAATGAGACCCACACTAATACATAATAAGTACAATACAAATTATTATAGTTTCGAAAGGAAAAAAACAA
CAACGCAAGAAACAGCATCTAACCTCAAAATGACAGCTATTTGCTCTTTGAAAGATGCATCTCCATCATACTCAT
CAACAAAAATGGCGGATTTTTTTTCCACGGAAAGCATTTAGAATGGTTTTACCCGTCTCAATTGCATTTTGAAAT
GTTTCTGGTCATCTTTATTCTTCTCTTTATACCCCATCAATCATCTATAGATTTCATCGCTAAGCCTCTTCACCT
TCTATCTCTATTAAGTTTAAAACAAATTGAATGGAATCAGAAGATCTATAGGTGATCTAGGCCATTTAGGCATTT
GAGAAATGGGAATACCAGAAATAGATTATGAAAAACTCACCTGTCAAGGTCCTCGTCCGCTCCTTGCCC**GACAGA**
GTGGAGGTCTGAGACTCTGAGCACCTGCCGCTTGCCGCAGGCGCACCGTGCAGGGGGGTTGTGCAGAAAGAGCCT
TGTTGCCGGGCCGCCGAGCCAGCGAGAGGAGCTGAGGAGGCAGTGGACATGGACGGAGGGAGGGAGGTGAGGCGG
CTGGGCAGGGTGGTGGGTGGCTGGAAGCGGGGAGACTCACGTATGGTGGCGGCTGGCCTTGGAGGGAGCGAGCAG
GGAGGGAGGCGAGGGGACGGACGCATCGGCGTCGGCTCATCTCCCACGAGCGCATAGCGTTATTGGGTGAAGAGG
GGGNAG

Supplementary 3

SG1 Forward Level 1

>SG1.FWDPRIMER
GTCTCAGACCTCCACTCTGTCGGG

>SG1.201F
CANCCGGGAACACTAGAATTCGAGCTCGGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATA
ATGATAGAGTCGACATAGCGATTGTCTCAGACCTCCACTCTGTCGGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT
AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGC
TTTACTTGTCTTCTGCACGAG

>SG1.202F
CCCCCCGGGAAAACTAAGAATTCGAGCTCGGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATA
TAATGATAGAGTCGACATAGCGATTGTCTCAGACCTCCACTCTGTCGGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTAC
GCTTTACTTGTCTTCTGCACGAAAA

>SG1.401F
TCGAGCTCGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCG
ATTGTCTCAGACCTCCACTCTGTCGGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA
AAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTTTACTTGTCTTCTGCACGA
AATT

>SG1.402F
CCCACCCGGGAAAACTAGAATTCGAGCTCGGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATA
TAATGATAGAGTCGACATAGCGATTGTCTCAGACCTCCACTCTGTCGGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTAC
GCTTTACTTGTCTTCTGCACGA

>SG1.1001F
ACCCCGGNAAAACTAGAATTCGAGCTCGGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATA
ATGATAGAGTCGACATAGCGATTGTCTCAGACCTCCACTCTGTCGGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT
AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGC
TTTACTTGTCTTCTGCACGAA

>SG1.1002F
ACCCGGGAAAACTAGAATTCGAGCTCGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATAAT
GATAGAGTCGACATAGCGATTGTCTCAGACCTCCACTCTGTCGGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTT
TACTTGTCTTCTGCACGAA

Supplementary 4

SG1 Reverse Level 1

>SG1.RVS
TGTGGTCTCAAGCGTAATGCCAACTTTGTAC

>SG1.201R
TTTGTAAGAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTT
GCTATTTCTAGCTCTAAAACCCCCGACAGAGTGGAGGTCTGAGACAATCGCTATGTCGACTCTATCATTATATAAACTAAGCTG
CTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCACTGCA
TCCACCCCAGTACAA

>SG1.202R
TTTGAAANAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTT
GCTATTTCTAGCTCTAAAACCCCCGACAGAGTGGAGGTCTGAGACAATCGCTATGTCGACTCTATCATTATATAAACTAAGCTG
CTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCACTGCA
TCCACCCCAGTACAAA

>SG1.401R
CTTTGTCAGAAAGCTGGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC
TTGCTATTTCTAGCTCTAAAACCCCCGACAGAGTGGAGGTCTGAGACAATCGCTATGTCGACTCTATCATTATATAAACTAAGC
TGCTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCACTG
CATCCACCCCAGTACAACNT

>SG1.402R
TCCTTTGTAAGAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAA
CTTGCTATTTCTAGCTCTAAAACCCCCGACAGAGTGGAGGTCTGAGACAATCGCTATGTCGACTCTATCATTATATAAACTAAG
CTGCTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCACT
GCATCCACCCCAGTACAAA

>SG1.1001R
TCCTTTGTNAGGAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA
ACTTGCTATTTCTAGCTCTAAAACCCCCGACAGAGTGGAGGTCTGAGACAATCGCTATGTCGACTCTATCATTATATAAACTAA
GCTGCTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCAC
TGCATCCACCCCAGTACAA

>SG1.1002R
TTTGTNAGGAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACT
TGCTATTTCTAGCTCTAAAACCCCCGACAGAGTGGAGGTCTGAGACAATCGCTATGTCGACTCTATCATTATATAAACTAAGCT
GCTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCACTGC
ATCCACCCCAGTACA

Supplementary 5

SG3 Forward Level 1

>SG3.FWD
GCTCATGATGTCCGACTACAAGG

>SG3.201F
CGNGNAAACTAAGAATTCGAGCTCGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATAATGA
TAGAGTCGACATAGCGATTGCTCATGATGTCCGACTACAAGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCC
GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTTTAC
TTGTCTTCTGCACGA

>SG3.202F
NCGGGANCAACTAGAATTCGAGCTCGGAGTGATCAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATAATGA
TAGAGTCGACATAGCGATTGCTCATGATGTCCGACTACAAGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCC
GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTTTAC
TTGTCTTCTGCACGA

>SG3.401F
ACCGTGNAAACTAAGAATTCGAGCTCGGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATAA
TGATAGAGTCGACATAGCGATTGCTCATGATGTCCGACTACAAGGGGTTTTAAAGCTAAAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGGGGCACCGAGTCGGGGCTTTTTTTCTAAACCCACCTTTCTTGTACAAAGTTGGCATTACCCTT
TACTTGTCTTCTGCAGGAA

>SG3.402F
ACCGGGNAAACTAGAATTCGAGCTCGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATAATG
ATAGAGTCGACATAGCGATTGCTCATGATGTCCACTACCAGGGGGTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGGTAGTCC
GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAAACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTTTAC
TTGGCTTCTGCACGGA

>SG3.1001F
ACCGGGAAAACTAGAATTCGAGCTCGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATAATG
ATAGAGTCGACATAGCGATTGCTCATGATGTCCGACTACAAGGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTTTA
CTTGTCTTCTGCACGAA

>SG3.1002F
ACCCGGGAAAACTAAGAATTCGAGCTCGGAGTGATCAAAAGTCCCACATCGATCAGGTGATATATAGCAGCTTAGTTTATATAA
TGATAGAGTCGACATAGCGATTGCTCATATGTCCGACTACAAGGGGTTTTAGAGGCTAGAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTT
TACTTGTCTTCTGCACGA

Supplementary 6

SG3 Reverse Level 1

>SG3.RVS
TGTGGTCTCAAGCGTAATGCCAACTTTGTAC

>SG3.201R
GNNAGAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCT
ATTTCTAGCTCTAAAACCCCTTGTAGTCGGACATCATGAGCAATCGCTATGTCGACTCTATCATTATATAAACTAAGCTGCTAT
ATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCACTGCATCCA
CCCCAGTACAAA

>SG3.202R
NTTTGTCAGAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACT
TGCTATTTCTAGCTCTAAAACCCCTTGTAGTCGGACATCATGAGCAATCGCTATGTCGACTCTATCATTATATAAACTAAGCTG
CTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCACTGCA
TCCACCCCAGTACAA

>SG3.401R
TCNTTTGTAAGAAAGCTGGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA
ACTTGCTATTTCTAGCTCTAAAACCCCTTGTAGTCGGACATCATGACCAATCGCTATGTCAACTCTATCATTATATAAACTAAC
CTGCTATATATCACCTGATCGATGGGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCCCAAAGTGGGGCCCACT
GCTTCCACCCCAGTAAAAA

>SG3.402R
CTTTGTNAGGAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC
TTGCTATTTCTAGCTCTAAAACCCCTTGTAGTGGACATCATGAACAATCGCTATGTCCACTCTATCATTATATAAACTAAGCTG
CTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAACTCGAATTCTAGTTTGGCTTCCCAGAATGGGGCCCACTGCA
TCCACCCCAGTACAAAA

>SG3.1001R
CATTTGTAAGAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC
TTGCTATTTCTAGCTCTAAAACCCCTTGTAGTCGGACATCATGAGCAATCGCTATGTCGACTCTATCATTATATAAACTAAGCT
GCTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCACTGC
ATCCACCCCAGTACAA

>SG3.1002R
CCTTTGTAAGAAAGCTGGGTCTAGAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC
TTGCTATTTCTAGCCTCTAAAACCCCTTGTAGTCGGACATATGAGCAATCGCTATGTCGACTCTATCATTATATAAACTAAGCT
GCTATATATCACCTGATCGATGTGGGACTTTTGATCACTCCGAGCTCGAATTCTAGTTTGTCTTCACAGAGTGGGGCCCACTGC
ATCCACCCCAGTACAA

Supplementary 7

SG1.201 Forward Level 2 *E. coli*

```
LVL2.SEQ                TCTCCGACTACGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTG
L2E_SG1201_Forward      ------------------------------------------------------------


LVL2.SEQ                ATAATAAAGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAG
L2E_SG1201_Forward      ----------------------CCATGAATGATAGAGGGAACAGTGATAACGTCCCCTCAG
                                              **   **  ********** ********************

LVL2.SEQ                AAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCA
L2E_SG1201_Forward      AAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCA
                        ************************************************************

LVL2.SEQ                CACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATA
L2E_SG1201_Forward      CACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATA
                        ************************************************************

LVL2.SEQ                AAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGGCCC
L2E_SG1201_Forward      AAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGGCCC
                        ************************************************************

LVL2.SEQ                AAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGG
L2E_SG1201_Forward      AAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGG
                        ************************************************************

LVL2.SEQ                TGAAAGTTATTACTCTGAAGTCTAAGCTGGTTTCAGATTTCAGAAAGGACTTTCAGTTTT
L2E_SG1201_Forward      TGAAAGTTATTACTCTGAAGTCTAAGCTGGTTTCAGATTTCAGAAAGGACTTTCAGTTTT
                        ************************************************************

LVL2.SEQ                ATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGG
L2E_SG1201_Forward      ATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGG
                        ************************************************************

LVL2.SEQ                TAGGCACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACT
L2E_SG1201_Forward      TAGGCACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACT
                        ************************************************************

LVL2.SEQ                ATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCA
L2E_SG1201_Forward      ATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAAG--G
                        ********************************************************** *

LVL2.SEQ                CCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGG
L2E_SG1201_Forward      CCACAAAA----------------------------------------------------
                        ** * **
```

Supplementary 8

SG1.201 Reverse Level 2 *E. coli*

```
LVL2.SEQ                TCTCCGACTACGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTG
L2E_SG1201_REVRCOMP     ----------TTACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTG
                                  **********************************************

LVL2.SEQ                ATAATAAAGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAG
L2E_SG1201_REVRCOMP     ATAATAAAGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAG
                        ************************************************************

LVL2.SEQ                AAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCA
L2E_SG1201_REVRCOMP     AAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCA
                        ************************************************************

LVL2.SEQ                CACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATA
L2E_SG1201_REVRCOMP     CACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATA
                        ************************************************************

LVL2.SEQ                AAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGGCCC
L2E_SG1201_REVRCOMP     AAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGGCCC
                        ************************************************************

LVL2.SEQ                AAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGG
L2E_SG1201_REVRCOMP     AAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGG
                        ************************************************************

LVL2.SEQ                TGAAAGTTATTACTCTGAAGTCTAAGCTGGTTTCAGATTTCAGAAAGGACTTTCAGTTTT
L2E_SG1201_REVRCOMP     TGAAAGTTATTACTCTGAAGTCTAAGCTGGTTTCAGATTTCAGAAAGGACTTTCAGTTTT
                        ************************************************************

LVL2.SEQ                ATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGG
L2E_SG1201_REVRCOMP     ATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGG
                        ************************************************************

LVL2.SEQ                TAGGCACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACT
L2E_SG1201_REVRCOMP     TAGGCACTGCACTTATCAAAAAATATCCCAAGCTAGAATGCGAATTTGCCTACGCAGACC
                        ********************************* ****  *******  **** ****
```