12-2019

Analyzing Multigene Stacking and Genome Editing Strategies in Rice

Bhuvan Pathak
University of Arkansas, Fayetteville

Analyzing Multigene Stacking and Genome Editing Strategies in Rice

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology

by

Bhuvan Pathak
Sardar Patel University
Bachelor of Science in Botany, 2002
Sardar Patel University
Master of Science in Biotechnology, 2004
University of Florida
Master of Science in Agronomy, 2010

December 2019
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Vibha Srivastava, Ph.D.
Dissertation Director

_____
Suresh Kumar Thallapuranam, Ph. D.
Committee Member

_____
Ainong Shi, Ph. D.
Committee Member

_____
Andy Pereira, Ph. D.
Committee Member

Abstract

Crop improvement through biotechnology is an integrated effort, incorporating multiple approaches like integration of genes, editing of native genes, and removal of selection marker genes. Before streamlining the protocols, the efficiency and feasibility of the individual approach and their components must be tested. This study evaluated following approaches: 1) stacking an array of genes into a single locus by site-specific integration via Cre-*lox* recombination in rice, 2) determining the efficiency of I-*Sce*I and the *CCR5*-ZFN in the targeted excisions of gene fragments in rice and *Arabidopsis*, and 3) determining the efficiency of CRISPR/Cas9 in generating targeted mutations for genome editing in rice. In gene stacking, >50% site-specific integration lines contained full-length integration of five genes. All genes were properly regulated by their promoters as indicated by the correlation of expression levels of the three constitutively expressed genes with their allelic number, and heat- or cold-induction levels of the two inducible genes. Analysis of I-*Sce*I and *CCR5*-ZFN in rice and *Arabidopsis* found that these overexpressing constructs were refractory to plant transformation. The heat-inducible I-*Sce*I expression in *Arabidopsis* was effective in creating somatic excisions but ineffective in generating heritable excisions. The inducible expression of *CCR5*-ZFN in rice, although transmitted stably to the progeny, appeared ineffective in creating detectable excisions. Finally, the application of CRISPR/Cas9 in rice was found to induce mutations at a high rate, but point-mutations occurred far more frequently than genomic deletions as determined in 114 rice lines including the primary transgenic lines and their progenies for 3 different genes. The heat-shock induced CRISPR/Cas9 was found to create heat-inducible targeted mutations that were inherited by the progeny. Additionally, mutations in the predicted off-target sites were undetectable or found at a lower rate in the heat-shock CRISPR/Cas9 lines as compared to their

frequency in the constitutive-overexpression CRISPR/Cas9 lines. In summary, while Cre-*lox* mediated site-specific integration and CRISPR/Cas9 mediated point-mutagenesis were highly effective in rice genome, application of I-*Sce*I or *CCR5*-ZFN was problematic as tested in *Arabidopsis* and/or rice.

Acknowledgements

I am greatly indebted to Dr. Vibha Srivastava, chairperson of my graduate supervisory committee for her encouragement, inspiration, financial assistance, and expert guidance throughout this project. The daily discussion of science with her has shaped me as a researcher. She is not only a great mentor, but also a great person who has also supported throughout my graduate journey. Her Plant Genetic Engineering class is my one of the most favorite classes, I have ever taken in my life.

I am grateful to my committee members Dr. Suresh Kumar Thallapuranam, Dr. Andy Pereira, and Dr. Ainong Shi for serving on my graduate committee. I have found my teaching idol in Dr. Kumar, whose interaction and personal connection with students, not only kept us motivated in the class, but also had built a strong foundation and understanding of nature and its biological system. The interaction with Dr. Pereira has always made me to smile, which had helped me tremendously in my graduate journey. I have always found Dr. Shi, to be a very calm and composed. His class on plant molecular marker, had immensely helped me in my candidacy exam. I will be forever grateful to him for teaching us this class in such a fun way.

I would like to thank Dr. Fiona Goggin for allowing me to use her laboratory facilities and Dr. Janithri Wickramanayake for teaching me the plate reader for my protein studies. I am heartily thankful to Dr. Jianfeng Xu of Arkansas State University, Jonesboro, AR; Dr. Betty Martin, and Dr. Mourad Benamara of Arkansas Nano and Bio Materials Characterization Facility at UARK, Fayetteville for their guidance and assistance in the confocal microscopy.

My work would not have been completed without a great team of my lab. My heartfelt thanks to Dr. Soumen Nandy, Shan Zhao, Elliot Pruett, Jamie Underwood, Dr. Huazong Guan, and Dr. Flavia Botelho who had not only helped me in my experiments, but we also used to

indulge in many scientific discussions and had fun while working in the lab. I am also heartily thankful to Jessica Kivett for her help and guidance in the greenhouse related activities.

I would like to thank my friends Yheni Dwiningsih, Peter Icalia, Shilpi Agrawal, Sumana Venkat, Anamika Gupta, Dr. Julie Thomas, Dr. Anuj Kumar, Dr. Chirag Gupta, Sara Yingling, Ozgur Azapoglu, Amanda Holder, and Zahra Alizada, Gehendra Bhattarai and Waltram Ravelombola for their constant support and love. With Yheni, I have been constantly inspired to work and stay focused - be it normal days, late nights or weekends and hang out in the weekends and holidays. Special mention goes to Dr. Gulab Rangani and her family for her great friendship and support, who had constantly guided me during my needs.

My journey could not have been completed without the constant inspiration and strength from my family. My parents Paresh Pathak, Jayshree Pathak, Shaunakbhai Bhatt, and Charuben Bhatt; my sisters Dhruva, Shivanididi; brothers Shrirangbhai, and Sumit and my huband Samarth have been a constant pillar of strength, love and affection. My two gems, Rudraansh and Madhav have taught me to sustain all high and low of the life and have always inspired me to work harder. My sisters Harneet (Neeti) and Sharon have been supportive, loving and caring throughout my time in US. They are my big support system who have constantly inspired me to pursue my dreams; helped me to understand myself as an individual and overcome the hurdles with great zeal and patience.

Dedication

To my lifelines Rudraansh and Madhav

Table of Contents

List of Tables

Chapter II

Chapter III

Chapter IV

Chapter V

List of Figures

Chapter II

List of Published Papers

Pathak BP, Pruett E, Guan H, Srivastava V (2019). Utility of I-Sce I and CCR5-ZFN nucleases in excising selectable marker genes from transgenic plants. BMC Research Notes. 12(1):272. https://doi.org/10.1186/s13104-019-4304-2 **(Chapter III)**

Pathak B, Zhao S, Manoharan M, Srivastava V (2019). Dual-targeting by CRISPR/Cas9 leads to efficient point mutagenesis but only rare targeted deletions in the rice genome. 3 Biotech. 9(4):158. https://doi.org/10.1007/s13205-019-1690-z **(Chapter IV)**

Nandy S*, Pathak B*, Zhao S, Srivastava V (2019). Heat-shock-inducible CRISPR/Cas9 system generates heritable mutations in rice. Plant Direct. 3 (5):e00145. https://doi.org/10.1002/pld3.145 **(Chapter V).**

* Equal contributing authors

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Introduction

With a tremendous rise in the world population, estimated to be nine billion by 2050; global agriculture production needs to increase by 60%-110%. Ray et al. (2013) studied four key global crops including maize, rice, wheat, and soybean and observed that these crops increased only at the rate of 1.6%, 1.0%, 0.9%, and 1.3% per year, respectively, which was less than the rate (2.4% per year) needed to double the agriculture production by 2050. If these rates would continue, then it will increase to only ~67%, ~42%, ~38%, and ~55%, respectively, which is much below the threshold level to meet the increasing food demand and food security.

Rice and wheat, each provide 19% of the dietary requirement to the world population. The top three world producers of rice, China, India and Indonesia have so far seen only 1.7%, 1.1%, and 0.8% per year increase in the rice yield, which may affect the global food security (Ray et al. 2013). Therefore, for sustainable agriculture, a number of studies have suggested that it is more important to increase the crop yield in a given area of land, rather than creating more agriculture suitable places (Foley et al. 2011; Godfray et al. 2010; Green et al. 2005; Matson et al. 2006).

The stupendous growth of the human population and the shrinking agriculture land calls for the crop improvement. Crop improvement requires a continuous effort of crop selection with the beneficial traits ensuring optimal productivity even during adverse climatic conditions (Srivastava and Thomson, 2016). Crop improvement is mainly performed by traditional breeding methods sometimes assisted by marker selection. Often a trait introduction like pest and disease resistance (Dong and Ronald, 2019), yield (Breseghello and Coelho, 2013) and nutrition enhancement (Hefferson, 2015), requires a deployment and expression of multiple genes. With an increase in the number of genes, it requires large amount of F2 plants for the selection of

complex traits making the selection process extremely difficult. In trait introduction, traditional breeding also sees the introduction of undesirable alleles owing to chromosomal recombination in trait transfer from the donor parent to the cultivars (Breseghello and Coelho, 2013; Petolino and Kumar, 2016; Srivastava and Thomson, 2016). Given the dearth of the agriculture land, and a time required for traditional breeding, it summons for an alternative approaches of the crop improvement to meet the food and feed demand.

Biotechnology assisted methods like gene stacking by genetic engineering or targeted mutations by either excisions, insertions/deletions or substitutions, are the most sought techniques in the crop improvement to expedite breeding. The trait stacking is defined by the introduction of multiple genes. If these genes are transferred to the same locus or chromosomal segment, they will be co-inherited. This will greatly simplify breeding multigenic traits or multiple traits. The genomic mutations or deletions on the other hand, deals with the generation of mutations (either point or larger) and/or bigger genomic deletions. If these mutations occur naturally, it would take a considerable time to discover and transfer into cultivars (Blanco et al. 2009). Both of these approaches are based on the same principle of inducing a double-stranded break (DSB) in the genome followed by a cellular repair by either homologous recombination (HR) or non-homologous mediated end joining (NHEJ). Crop improvement through biotechnology is an integrated effort, which requires combinations of tools like site-specific recombinases, engineered or rare nucleases, and CRISPR/Cas9. However, before the development of a streamlined protocol for crop improvement through combined use of these tools; the efficiency, feasibility, and functionality of each component must be tested. Therefore, the objectives of this study are:

Objective 1: Evaluation of the structural and expression stability of the multigene stacks in rice developed through Cre-*lox* site-specific integration.

Objective 2: Characterization of I-*Sce*I and *CCR5-ZFN* nuclease activities for targeted excisions in rice and *Arabidopsis* genomes.

Objective 3: Evaluation of CRISPR/Cas9 in generating targeted mutations by (a) constitutive expression of Cas9, (b) heat-inducible expression of Cas9 in the rice genome.

Literature Review

Targeted Gene integration by site specific recombinases

Site-specific gene integration (SSI) or site-specific recombination (SSR) is done by site-specific recombinases. First discovered in bacteria and lower eukaryotes like yeast, they are responsible for phase variation of bacterial virulence and bacteriophage integration in the host genome. Based on the amino acid present at the active site of catalytic domain, they are differentiated into the Serine (S) and Tyrosine (Y) groups (Srivastava and Thomson, 2016).

The Y family contains most well characterized and studied systems of Cre-*lox* (Sauer and Henderson, 1990), FLP-*FRT* (Golic and Lindquist, 1989) and R-*RS* systems (Onouchi et al. 1991) Here, Cre, FLP and R are the recombinase enzymes and *lox*, *FRT* and *RS* are their recognition sites. The *lox*, *FRT* and *RS* recognition sites contain identical left and right arms which consist of inverted repeats flanking a short spacer sequences. These inverted repeats are the binding sites, while the spacer is a DNA nicking site. These identical sequences, make the reaction fully reversible i.e. bidirectional, though excision is favored over integration in the reaction kinetics. On the contrary, the unidirectional Y recombinases contain non-identical recognition sites *att*B (attachment site bacteria) and *att*P (attachment site phage) that participate

in an irreversible recombination when the helper protein excisionase is absent (Srivastava and Thomson, 2016)

The S recombinases have two distinct members; 1) a serine small subfamily which contains β-*six* (Diaz et al 2001*)*, γδ-*res* (Schwikardi and Droge, 2000), CinH-*RS2* (Thomson and Ow, 2006) and ParA-*MRS* (Thomson et al. 2009); 2) large serine subfamily containing phiC31 (Rubtsova et al. 2008), TP901-1 (Stoll et al. 2002), R4 (Olivares et al. 2001) and Bxb1 (Thomson and Ow, 2006).

In the small subfamily *six, res, RS2* and *MRS* are the recognition sites of the β, γδ, CinH and ParA recombinases, respectively. Like *FRT* and *lox* sequences, these recognition sites are also identical. In this subfamily, only excision events have been observed. Also, the excision event is considered as an irreversible reaction because, during the synaptonemal complex formation, these recombinases impart a conformational strain due to which integration is not possible (Mouw et al. 2008). The large serine recombinases have recognition sites of *attP* and *attB*, which yield a hybrid product of *attL* and *attR* upon recombination. These systems work very efficiently for excision, integration and inversion since the conversion of *attP* and *attB* to *attL* and *attR* makes the reverse reaction impossible without the addition of second protein, excisionase (Ghosh et al. 2006; Thorpe et al. 2000).

Cre-*lox* and FLP-*FRT* are the most studied and widely used tool to carry out site-specific recombination reactions. Most of the early studies focused on the efficiency of these systems to carry out excision of marker genes flanked by the recognition sites from the transgene locus (Dale and Ow, 1991; Russell et al. 1992). However, as the system is freely reversible, they can also carry out the site-specific integration (SSI). In order to prevent the reversibility of the reaction, it is necessary to optimize the strategy, which can also provide the stability of the

integration structure. For optimization, two strategies have been mainly used: 1) use of the mutant recognition sites, which can recombine and generate the double mutants' recognition sites to prevent the reaction reversibility, and 2) transient expression of recombinase activity through co-bombardment of the recombinase gene.

In the generation of mutant sites, one of the left or right arm or element (LE or RE) of the recognition was mutated by introducing the 4 - 7 bp mutation (Albert et al. 1995; Srivastava and Ow, 2001; Srivastava and Thomson, 2016). These LE and RE mutants could recombine efficiently due to the cooperativity in binding of the recombinase monomers to a recognition site. This recombination would then result in the doubly mutated RE: LE site and a wild type recognition site. The RE: LE mutants do not bind to the recombinase properly, thus rendering it inactive. This method was used to generate the site-specific integration (SSI) of the transgene in rice and tobacco (Akbudak and Srivastava, 2011; Albert et al. 1995; Chawla et al. 2006; Srivastava et al. 2004). In this approach, the *lox76*, which contains 7 bp mutation in the left arm placed in the genome had recombined with *lox75* located in the donor DNA. *Lox75* also contains 7 bp mutation, but in the right arm as a result *lox75 x lox76* recombination generates a double mutant *lox78* and a wild-type *loxP*. This method enabled recovery of 80-90% transformed clones containing the SSI structure (Srivastava et al. 2004). Additionally, this double mutant had also provided the locus stability despite the presence of Cre with only a few cases of excisions, suggesting the refractory nature of *lox78*.

Unlike *lox*, the *FRT* mutants contain single point mutations in their left or right arms (*FRT46A* and *FRT46T*). These mutants were shown to be effective in the controlling the reversible reaction in *E. coli*; however, in rice, it was found to recombine reversibly (Nandy and Srivastava, 2011). The transient *FLP* expression, by co-transformation of *FLP* gene with the

donor DNA was subsequently used to produce 20-30% SSI clones in rice through FRT_L x FRT_R recombination. The resulting SSI clones lacked co-integration of *FLP* gene, indicating that the transient expression of *FLP* was sufficient to carryout successful site-specific integration in the rice genome (Nandy and Srivastava, 2011). In a recent study in maize (Anand et al. 2019), the use of different heterologous combinations of *FRT* sites (*FRT1 x FRT87*, *FRT1x FRT86* and *FRT1 x FRT12*) showed that *FRT1x FRT86* and *FRT1 x FRT12* generated 3.5 times higher SSI lines than *FRT1 x FRT87*. The *FRT1 x FRT87* combination had shown higher cross-reactivity when *FLP* was transiently expressed leading to more excisions than integration events (Anand et al. 2019). The transient Cre expression and the use of mutant *lox* sites in tobacco had also helped to create stable transgene locus (Albert et al. 1995).

The serine family recombinases that contain the non-identical recognition sites *attP* and *attB* are suitable for gene pyramiding. Sequential transformation of the sites into SSI could theoretically generate a good launching pad for gene integration into a single locus. Because of non-identical nature, these sites cannot carry out a reversible reaction and hence, are ideal for the gene integration. This approach was used to pyramid 3 genes in tobacco with the efficiency of 10-13% by Hou et al (2014) by two rounds of *attPxattB* recombination by Bxb-1 recombinase. In the iterative round of transformation in *Arabidopsis*, De Paepe et al (2013) had obtained 9% (3 of 35) SSI lines containing *eGFP/GUS* and *NPT II* using ΦC31 integrase with *attP x attB* and Cre-*lox* systems. All these lines had also shown stable inheritance of the genes by the next generation. Hence, both types of recombinases (Serine and Tyrosine) can be used for the gene integration. However, tyrosine recombinases have higher efficiency and are more favored for gene integration than serine recombinases.

The precision and efficiency of the DNA integration at a predetermined site are essential components of the recombinase mediated gene integration. Precision of the integration is determined when a single-copy of DNA fragment bordered by recombination sites integrates at the target locus, without any unpredictable gain or loss of the DNA sequences. This accuracy has been frequently observed in both plants and animal systems, when the mutant *lox* and/or mutant/heterologous *FRT* sites are used for the integration (Anand et al. 2019; Chawla et al. 2006; De Paepe et al. 2013; Schetelig et al. 2019; Srivastava et al. 2004; Srivastava and Ow, 2001). Using *lox75 x lox 76*, Chawla et al. (2006) and Srivastava et al. (2004) had recovered ~80% of the precise SSI single-copy lines of rice. Using heterologous *FRT* sites, Anand et al (2019) could obtain 7% of the precise SSI single-copy lines in maize; however, in rice, the efficiency of recovered precise single copy lines by FLP-*FRT* have been reported to be 30% (Nandy and Srivastava, 2011).

During site-specific integration, extra copies of the donor DNA often integrate generating a multicopy insertion patterns on the Southern blots. Around 50% of SSI lines have been reported to contain extra copies in tobacco and rice (Albert et al. 1995; Srivastava et al. 2004). These random integrations do not disrupt the structure of SSI as they are often integrated at far distance from the SSI locus (Chawla et al. 2006; Lowerse et al. 2007), but can influence the gene expression through RNAi mechanism (Akbudak and Srivastava, 2010).

The precision of recombination is also reflected in the expression-stability of the SSI locus through successive generations. It has been well studied that single-copy integration is a crucial aspect for the transgene stability (Srivastava and Thomson, 2016). Integration of the full-length DNA fragment is an important part to determine the stability of gene expression in the SSI locus by keeping each transcription unit intact and avoiding the aberrant transcription. Moreover,

the presence of random integrations also causes the gene silencing in SSI lines, which could be reversed by the segregation of unwanted gene fragments in the next generation (Akbudak et al. 2010; Nandy and Srivastava, 2011; Chawla et al. 2006). Many studies have shown the stable expression of transgenes in the SSI locus through T3 generation in rice (Chawla et al. 2006), *Arabidopsis* (Day et al. 2000; Paepe et al. 2013; Vergunst et al. 1998), soybean (Li et al. 2009; 2010), mustard (Bala et al. 2013), and barley (Kapusi et al. 2012). These studies focused on the expression and stability of only 2 genes in the SSI locus, while the stability, functionality and inheritance of the multigene stack in the SSI locus is yet to be addressed. Therefore, the first objective of this study will evaluate the stability of multigene (5 genes) stack in a single locus in rice developed through Cre-*lox* mediated site specific integration.

In summary, use of the site-specific recombinases can, not only provide the precise integration, but also provide the stability and uniform expression of the genes present in the integration locus.

Zinc Finger Nuclease

Zinc finger (ZF) proteins are the most common type of DNA binding proteins with 8-10 array of fingers. It has been extensively studied in the human genome. The ZF contains Cys2-His2 DNA binding motif, and can bind to any DNA sequences. Each zinc finger consists of ~30 amino acids in a conserved ββα. The amino acids of the alpha helix bind to the 3 bp in the major groove of DNA, with different selectivity (Gaj et al. 2013). The variable selectivity of the ZF protein was the base of the development of zinc finger nuclease (ZFN). The ZFN contained a separate DNA-binding and DNA-cleavage domains. The DNA binding domains consisted of ZF, which can recognize the DNA sequences of 9-18 bp in length. The DNA cleavage domain contained a cleavage domain from *FokI*, a type II restriction enzyme. This cleavage domain does

not have a sequence specificity and thus, cutting could be redirected by the substitution of alternative recognition domain (Kim and Chandrasegaran, 1994; Kim and Pabo, 1998).

The zinc finger nuclease contains target sites known as left and right arms. A 5-7 bp spacer recognized by FokI cleavage site (Caroll, 2014; Gaj et al. 2013) separates each site. When both arms bind to their recognition sequence, the cleavage domain of *FokI* induces the double-stranded break in the target sequence, which is later repaired by the cellular repair machinery through homologous end-joining (HR) or non-homologous end joining (NHEJ). Thus, it can easily be used for genome targeting. Chandrasegaran and their coworkers, who had shown that the artificial assembly of ZFN could induce DSB in the cells and generate chimeras by using multiple targets including homeo-box domain in *Drosophila melanogaster* and yeast Gal4 DNA-binding domain (Kim et al. 1994, 1996 and 1999), demonstrated the first utility of ZFN in the 1990s. The first gene targeting was demonstrated in *yellow* (*y*) gene of *D. melanogaster* (Bibikova et al. 2002) using a pair of three-finger ZFN. The expression of this transgene was induced by heat stress in the fly larvae. This resulted in targeted mutagenesis and gene replacement by homologous recombination in the presence of a donor DNA. Since then, many studies in animals and humans have been reported, especially for their potential utilities in gene therapies. For example, potential of gene corrections by ZFN were demonstrated through numerous studies including the mutant *GFP* correction by a functional *GFP* in human kidney cell lines (Porteus and Baltimore, 2003), healthy gene replacement of the defective *IL2RG* gene for the treatment of severe combined immunodeficiency disease (Urnov et al. 2005), and the successful knockout of C-C chemokine receptor (*CCR5*), a gateway for the entry of HIV in the human cells (Didigu et al.2014; Perez et al. 2008).

In plants, the efficacy and feasibility of ZFN mediated gene targeting was first reported by Lloyd et al (2005) in *Arabidopsis* genome. Authors had chosen 5′-NNCNNCNNC (N6) GNNGNNGNN-3′ as their target site, which could be repeated in the genome at every 4^{18} bp. The expression of ZFN was induced by the heat-shock and the NHEJ mediated mutation frequency was reported to be 19.6%. These mutations were also inherited in 10% of the progeny studied. Like animal and human systems, in plants, ZFN were also primarily used for endogenous sequence modifications or gene corrections. In 2005, Wright et al. reported the restoration of the defective β-Glucuronidase (*GUS*) by ZFN in 10% of tobacco protoplasts by homologous recombination. In another study on tobacco, *SuRA* and *SuRB* were endogenously modified to confer the herbicide resistance by ZFN (Townsend et al. 2009). In maize, the targeted mutagenesis of *IPK1* resulted in the herbicide resistance phenotype (Shukla et al. 2009). In addition, many endogenous genes like *ABI4, ADH1* and *TT4* in *Arabidopsis* were also targeted and the somatic mutation frequencies had ranged from 3-16% with the stable inheritance of the mutations in the subsequent generations (Osakabe et al. 2010; Zheng et al. 2010). A few studies also demonstrated the application of ZFN for the cleavage of larger DNA sequences. In tobacco, Cai et al. (2009) stably transformed a construct which contained a tandem repeat of 540 bp in the two partial *GFP* gene fragment, which was separated by 2.8 kb of the heterologous fragment consisting of ZFN cleavage sites. The expression of ZFN had resulted in the induction of DSB which had deleted 2.8 kb fragment and had restored the functional *GFP*. Petolino et al. (2010) had reported the excision of 4.3 kb of integrated *GUS* gene flanked by ZFN target sites (*CCR5*), in 35% of F1 progenies when the ZFN expressing lines were crossed with target lines. In *Arabidopsis*, deletion of a gene cluster of 55 kb resistant gene locus comprising of eight

tandemly arrayed genes and pseudogenes has also been reported by Voytas (2013), although the results of the study has not been published.

Since there is a limited literature focused on the targeted excisions, the second objective of the current study, was to evaluate the efficiency of targeted excisions in rice by zinc finger nuclease.

Transcription Activator like Effector Nucleases (TALEN)

TALEN are Transcription Activator Like Effector Nucleases. They are derived from transcription activator like proteins in plant pathogen, *Xanthomonas,* which is delivered into the plant host cells. These proteins bind to various plant promoters for the activation of the infection mechanism (Boch et al. 2009; Romer et al. 2007). They consist of 33-35 multiple amino acid repeat binding domains in their left and right arms that recognize single nucleotides (unlike ZFNs, which recognize codon triplets) and a *FokI* cleavage domain. Despite its identical long sequences, it offers a great flexibility in designing than ZFN, and can be used to target any sequences (Gaj et al. 2013; Voyates, 2013). In plants, TALEN have been mainly utilized for inducing targeted mutations, but rarely for sequence excisions. The TALEN mediated mutation in the promoter of *OsSWEET14* gene in rice led to enhanced disease resistance (Li et al. 2012). In the polypoid wheat, the mutation in six *TaMLO* homeologs had enhanced disease resistance against powdery mildew (Wang et al. 2014), while targeted mutagenesis in the *FAD2-1A* and *FAD2-1B* in soybean led to the decreased levels of trans-fatty acids (Haun et al. 2014). In highly polyploid sugarcane, the mutagenesis in caffeic acid O-methyltransferase (*COMT*) had led to 29-32% reduction in the lignin content and improved saccharification efficiency for the biofuel production (Kannan et al. 2018; Jung and Altpeter, 2016). In rice and rapeseed mustard,

knocking-out of cytoplasmic male sterility associated genes, *orf92* and *orf125*, located in mitochondria, had restored the plant fertility (Kazama et al. 2019).

Clustered, Regularly Interspaced, Short Palindromic Repeats (CRISPR)

A third type of site-specific nuclease, distinct from ZFN and TALEN is known as Clustered, Regularly Interspaced, Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas) (proteins) has recently emerged as an alternative tool of genome editing for inducing targeted mutations. In nature, CRISPRs provide an adaptive/acquired immunity against foreign DNA via RNA-guided DNA cleavage (Wiedenheft et al. 2012). In this type of immunity, short segments of foreign DNA, known as "spacers" are integrated in the CRISPR genomic loci. When bacteria are attacked by other bacteria or bacteriophages; in defense, the integrated "spacers" are transcribed and processed into short CRISPR RNA (crRNA). These crRNA bind to the trans-activating crRNA (tracrRNAs) and guide the sequence-specific DSB of the invading pathogenic DNA by Cas enzyme. These Cas enzymes require a 20 bp seed sequence within the crRNA, which is similar to the target sequence and a conserved dinucleotide containing protospacer adjacent motif (PAM) sequence upstream of the seed sequence to bind crRNA region (Gaj et al. 2013; Jinek et al. 2012; Mali et al. 2013; Voytas et al. 2013). Upon binding of crRNA to the target sequence, Cas enzyme recognizes the PAM, and induces a DSB in the seed region. Cas enzymes derived from different bacterial species have different PAM requirement and act on different sites for DSB induction (Swarts and Jinek, 2018). The most widely used Cas is derived from *Streptococcus pyogenes*, known as Cas9. This enzyme recognizes the NGG (N= A/T/C/G) PAM, and induces a DSB between 3rd and 4th nucleotides upstream of the PAM in the seed region. While, other Cas known as Cpf1, derived from *Acidaminococcus sp.* and/or

Lachnospiraceae sp. recognizes TTTV (V=A/C/G) or TTTN and induces a DSB at the distal end of the target sequence (Lee et al. 2019; Zetsche et al. 2017; 2015).

Once the DSB is induced, the cell undergoes a repair using two different mechanisms known as non-homologous end joining (NHEJ) or homologous repair (HR). The NHEJ mediated repair is the most common type of cellular repair mechanism (Gaj et al. 2013; Voytas et al. 2013). The NHEJ mostly results in the error-prone repair, generating single nucleotide insertions and/or deletions (Indels), and occasionally larger indels extending from few bp to kilo base pairs. Therefore, like ZFN and TALEN, the CRISPR/Cas system could be targeted to cleave any DNA sequence by reprograming crRNA, and can be used to study the functions of different genes either through knock-in or knockout approaches. Using this concept, Jinek et al. (2012) demonstrated that it is possible to fuse the two RNA molecules of crRNA and tracrRNA *in vitro* known as single guided RNA (sgRNA or gRNA). Authors had delivered the gRNA and Cas9 enzyme into the human cells to target *CLTA* locus and had obtained the targeted mutation frequency of 6-8% (Jinek et al. 2013). In parallel, Mali et al. (2013) and Cong et al. (2013) obtained targeted mutation frequencies of 2-25% when multiple gRNA targeting multiple loci in the human cell lines were multiplexed.

As opposed to the ZFN and TALENs, the specificity of the RNA-guided nuclease, Cas9, is determined by the 20-nucleotide sgRNA. The CRISPR/Cas offers many advantages over ZFN and TALENs. Its low cost and simplicity has enabled its use in many labs across the globe. For the sequence specificity, it only requires the insertion of desired DNA sequences into vector construct for target site selection. The Cas enzyme does not require any alteration, as opposed to the ZFN and TALENs which requires the fusion of *FokI* nuclease domain with its target recognition domain. The simultaneous expression of multiple gRNAs allows studying the

functions of many genes at a time, which is not only economically viable, but also timing effective. Since its first report in 2012, numerous studies have focused on the application of CRISPR/Cas system for genome editing via NHEJ or HR in human, zebrafish and mice and plants (Cong et al. 2013; Hwang et al. 2013; Shan et al. 2013; Wang et al. 2013).

In plants, the two main multiplexed vector systems are currently in use for the design of *Cas9* and sgRNA. In the first type, each individual gRNA contains its species-specific promoter and terminator, and then multiple gRNA cassettes are combined together through golden gate or Gibson assembly (Ma et al. 2015). The *Cas9* contains its own expression cassette with species-specific/constitutive promoters and are co-transformed in the plant cells. The second system developed by Xie et al. (2015) is known as polycistronic tRNA-gRNA (PTG). It uses the mechanism of endogenous tRNA processing system. This vector contains a single promoter to drive multiple gRNAs and a transcription terminator. The *Cas9* expression cassette is fused with the gRNA expression cassette. Thus, it requires delivery of only a single vector in the plants. In the CRISPR studies, both types of vectors have proven efficient in the targeted mutagenesis. For convenience, the first type will be referred as traditional and second one will be referred as PTG systems.

Some of the early reports in *Arabidopsis*, rice, and tobacco using traditional gRNA and Cas9 vector cassette had successfully obtained targeted mutations with the frequency of 10-84% for multiple genes namely *CHL1*, *CHLI1*, *CHLI2*, *BR1*, *JAZ1*, *GA1*, *ROC5*, *SPP*, and *YSA*. These mutations were also successfully inherited in T2/T3 generations (Feng et al. 2013; Gao et al. 2015; Li et al. 2013; Mao et al. 2013). Ito et al. (2015) reported the mutations in *RIN* gene of tomato at three targeted sites, which encode a *MADS*-box transcription factor regulating fruit ripening. The resulting mutants had less ripening and red coloring than controls, suggesting the

pivotal role of *RIN* in fruit ripening. In studying the effect of tissue culture on CRISPR/Cas induced mutagenesis in rice, Mikami et al. (2015) reported that extended tissue culture period increased the mutation frequency mediated by CRISPR/Cas9. Authors observed that mutation frequency in rice was highly dependent on the type of promoter and expression cassette used. The targeting efficiency of CRISPR has now been widely studied in soybean (Du et al. 2016), potato (Wang et al. 2015), populous (Fan et al. 2015), maize (Svitashev et al. 2016) with the targeting efficiencies ranging from 50-100%.

Most of these studies mentioned above, studied the point mutagenesis in the form of indels generated at the DSB site. However, CRISPR/Cas system has also been utilized for bigger genomic deletions. Kapusi et al (2017) studied the putative *EGNase* gene in barley for the genomic excision by dual simultaneous targeting using five different gRNA combinations. Authors had co-transformed the single gRNA expression cassettes to achieve the genomic excisions. Out of 31 T0 plants, six showed monoallelic or biallelic excision of 90-139 bp. Authors described the overall excision efficiency to be 6.7% in T0 plants. The T1 of four T0 plants showed the inheritance of the excision locus. The single targeting mainly generated short indels and was heritable in T1. The overall mutation efficiencies for all five gRNAs ranged from 2.2% to 6.7%. Nekrasov et al. (2017) targeted *SLMlo1* gene in tomato to confer the resistance against powdery mildew disease. The traditional gRNA construct was multiplexed for the dual simultaneous targeting. Out the 10 plants studied, three plants showed the deletion of 48 bp, and the deletions were homozygous or biallelic. Five T1 from one of the T0 plant were also studied for the inheritance of the excision locus. It was observed that, even-though, the plants had homozygous deletions of 48 bp, the pattern of deletion was different from the parent. The *AcPDS* gene in the kiwifruit was targeted to determine the feasibility of CRISPR /Cas9 in genomic

excisions using traditional and PTG vector systems (Wang et al. 2018) using four gRNAs in combinations of two each. Authors obtained the genomic excision in calli using PTG combinations while no genomic deletions were observed using the traditional CRISPR expression cassette. Two expression cassettes of PTG containing two gRNAs each showed the deletions of 755 bp and 271 bp with the overall excision efficiencies of 16.67% and 3.84% in the calli lines, respectively. The individual mutation efficiencies of four gRNAs ranged from 0-8.33%, when the traditional crispr expression cassette was used, but the mutagenic frequency had increased to 65-92% in all the four gRNAs, when the PTG construct was used. The rice *MPK* genes were targeted for the excision of the genomic fragment by simultaneous targeting using PTG system (Minkenberg et al. 2017). Authors selected eight different target sites on four different *MPK* (*MPK1, MPK2, MPK5, and MPK6*) genes and constructed polycistronic tRNA-sgRNA cassette in different combinations. Authors observed excision of 727 bp deletion in three T0 plants (of 14 tested, efficiency-21%) obtained from PTG containing 8 gRNA combinations. The eight T1 from three T0 lines tested for excision inheritance showed the inheritance in either monoallelic and/or homozygous patterns. One of the T1 line, which showed homozygous excision of 727 bp, had also stably inherited the excision locus in its five T2 lines tested. The individual targeting efficiencies (indels) for each gRNA ranged from 67-100%. The four T1 from four T0 plants that harbored mutations (indels) at eight different target sites were used to study the inheritance pattern of the mutations. Authors concluded that the mutations were heritable, but had a different degree of heterozygosity of mutations at different mutation sites. The natural variant of *DEP1* (*dep1* in *Japonica*) which harbors >500 bp deletion has dense, erect panicle, and increased grain yield, has been extensively used in the rice breeding program. As this phenotype is difficult to transfer in the *Indica* variety, Wang et al. (2017) targeted *DEP1* by

CRISPR/Cas9 for the genomic excision to achieve the similar phenotype. For simultaneous targeting, the four different target sites were selected in the region of *DEP1, which* has been deleted in the natural variant. The traditional construct was multiplexed in combinations of either four or two gRNAs and were transformed in the rice. Authors obtained deletions ranging from 200 bp – 767 bp. The average frequency of deletions was highest (24%) for 200 bp deletion and only 9% for the full-length deletions up to 767 bp. The overall excision efficiency was observed to be higher when only two gRNAs were used than combinations of four. The individual mutation frequencies of each gRNA tested was >90% in all the combinations studied. Authors also tested the genomic deletions of up to 10 kb by the simultaneous targeting. They selected three sites of *DEP1* and near/ on the gene Os09g0442100, that is ~8 kb downstream of *DEP1*. Using the same approach as described above, they obtained genomic deletions of 10 kb in only 16 events from 187 T0 events tested (efficiency=9%) when two gRNAs and were used. In case of the use of four gRNAs the simultaneous targeting efficiency of deletions was reduced to only 0.3% (2 of 578 events). Hence, authors concluded that increase in the number of target sites inside the gene could increase the large fragment deletions frequency, but not the full-length deletions. Tian et al. (2017) targeted the *PDS* gene (*CIPDS*) gene in watermelon to study the mutagenic efficiency of CRISPR/Cas9. Two target sites were selected and individual gRNA constructs were made. These constructs were transfected in protoplast. The gRNA1 and gRNA2 had the mutagenic efficiency of 51.6% and 42.1% respectively. A multiplexed vector was also constructed containing two gRNA for the stable transformation and the excision of fragment by dual targeting. All the 16T0 plantlets regenerated, showed editing events on both target sites; however, none of them showed the excision. Authors concluded that, as the distance of two

gRNA was 3.2kb, they were not able to achieve the genomic excision, despite the high individual mutagenic frequency of both gRNAs.

Overall, the genomic excisions occurred at a lower frequency than the point mutations. The frequency of genomic excisions had decreased with the increase in the distance between two target sites. Since, a few studies had focused on the CRISPR mediated genomic deletions; the third objective was to evaluate the efficiency of CRISPR/Cas9 in targeted mutations (longer excisions and/or point mutations) in rice.

Off-target effects of engineered nucleases

A major challenge in the use of engineered nucleases like ZFN and CRISPR/Cas9 is the binding of the nuclease to the unintended genomic sites (off-sites) in the genome that share the similar homology to the on-target site. Targeting of these off - sites and the indels generated because of NHEJ can lead to the gene inactivation or mutation. Multiple off targeting in the genome can lead to the chromosomal rearrangements (Yee, 2016) including chromosomal deletions, translocations and inversions, which can alter the phenotype and bias the data interpretations. There are main three factors that affect the off target activity. First, more homology of target sequence in the genome, increases the likelihood of the off target activity. Second, higher amount of nuclease expressions and third, the long exposure period increases chances of off target activity. In case of ZFN and TALEN, studies by Sanders et al (2013), Pattanayak et al. (2013) and Gullienger al (2014) observed that 21-29% of the off target sites could be cleaved *in vivo*/*in vitro* by the sequence specific ZFN/TALEN in the human cell lines. Unlike ZFN and TALEN, which requires longer target sequence, CRISPR is a simple tool and requires only 20 bp target sequence. Thus, potentially, it is more prone to off target activity than ZFN and TALEN. In the human cell lines, various studies have reported higher off target

mutagenesis ranging from 5-63% (Fu et al. 2013; Veres et al. 2015). However, in plants, only a few studies have reported a low frequency off target effects in cotton, *Arabidopsis* ,rice and soybean (Jacobs et al. 2015; Li et al. 2019; Tang et al. 2018; Zheng et al, 2018), while most of the studies did not find any off target activities in these species (Gao et al. 2015; Lee et al. 2019; Ma et al. 2015; Pan et al. 2016; Sun et al. 2016; Ueta et al. 2017; Young et al. 2019 ; Zhou et al. 2014; Zhu et al. 2015). This could be due to the fact, that many plant species are highly polyploid in nature and they contain many duplications/ repeats especially in the intergenic regions making the sites potentially difficult to analyze (Lee et al. 2019). Therefore, a more controlled approach of Cas9 and sgRNA selections and expressions are needed in order to minimize the off target effects (Yee, 2016).

Therefore, also as a part of third objective, a stress induced approach of Cas9 expression will be tested in two different genes, and their off targets will be studied and compared with the lines containing constitutive expression of Cas9.

All the nucleases namely ZFN, TALEN, and CRISPR/Cas9 induce only DSB in the genome. The DSB is later repaired by the NHEJ and HR mediated cellular repair. The NHEJ is a most common type of repair in the somatic cells, and since this dissertation study had mainly identified the NHEJ mediated mutations, the below section will discuss only on the NHEJ mediated repair.

Cellular repair by non-homologous end joining (NHEJ)

Based on the pattern of repairs and types of factors involved, the non-homologous end joining (NHEJ) are canonical (cNHEJ) or alternate (aNHEJ) (Puchta and Fauser, 2014). In cNHEJ, after the induction of DSB, the Ku heterodimer attaches to the DSB, thus preventing the degradation, followed by ligase 4 mediated repair. In aNHEJ, the DSB induction is followed by

the resection of *3'* at the broken ends. It forms a junction of two single strands at the site of few complementary nucleotides. The ends are trimmed and re-ligation occurs. Often in aNHEJ, micro-homologies are found and there are high chances of loss of the genetic information, in contrast to cNHEJ, which retains the original sequence because of the ligation of the ends. However, micro-homologies are rarely found in the cNHEJ (Puchta and Fauser, 2014). Shen et al. (2017) studied the types of NHEJ for the DSB repair induced by CRISPR/Cas9 in *Arabidopsis*. Authors had generated knockout lines of *CRU3* and *PPO in* the mutant background of *ku80*, required for cNHEJ repair; in *parp1 parp2*, required for aNHEJ, and in triple *ku80 parp1 parp2* mutant. Authors observed that larger deletions were observed in the *ku80* and *ku80 parp1 parp2* mutants, suggesting that when these pathways fail, the third type of uncharacterized repair pathway comes into the play, as it was also observed in the ZFN-mediated DSB repair in a *ku80*, *ku70* and *lig4* mutants of *Arabidopsis* (Osakabe et al. 2010).

References

Akbudak MA and Srivastava V (2011). Improved FLP recombinase, FLPe, efficiently removes marker gene from transgene locus developed by Cre–*lox* mediated site-specific gene integration in rice. Mol Biotechnol. 49(1):82-9.

Albert H, Dale EC, Lee E and Ow DW (1995). Site-specific integration of DNA into wild-type and mutant *lox* sites placed in the plant genome. The Plant J. 4:649-59.

Alonso-Blanco C, Aarts MG, Bentsink L, Keurentjes JJ, Reymond M, Vreugdenhil D and Koornneef M (2009). What has natural variation taught us about plant development, physiology, and adaptation? The Plant Cell. 21(7):1877-96.

Anand A, Wu E, Li Z, TeRonde S, Arling M, Lenderts B, Mutti JS, Gordon-Kamm W, Jones TJ and Chilcoat ND (2019). High efficiency Agrobacterium-mediated site-specific gene integration in maize utilizing the FLP-FRT recombination system. Plant Biotech J. Feb 1.

Bala A, Roy A, Das A, Chakraborti D and Das S (2013). Development of selectable marker free, insect resistant, transgenic mustard (Brassica juncea) plants using Cre/*lox* mediated recombination. BMC Biotech.1:88.

Bertier LD, Ron M, Huo H, Bradford KJ, Britt AB and Michelmore RW (2018). High-resolution analysis of the efficiency, heritability, and editing outcomes of CRISPR/Cas9-induced

modifications of NCED4 in lettuce (Lactuca sativa). G3: Genes, Genom, Genet. 8:1513-21.

Bibikova M, Golic M, Golic KG and Carroll D (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics. 3:1169-75.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A and Bonas U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 326 (5959):1509-12.

Breseghello F and Coelho AS (2013). Traditional and modern plant breeding methods with examples in rice (Oryza sativa L.). J Agr Food Chem. 35:8277-86.

Cai CQ, Doyon Y, Ainley WM, Miller JC, DeKelver RC, Moehle EA, Rock JM, Lee YL, Garrison R, Schulenberg L and Blue R (2009). Targeted transgene integration in plant cells using designed zinc finger nucleases. Plant Mol Biol. 69(6):699-709.

Carroll D and Beumer KJ (2014). Genome engineering with TALENs and ZFNs: repair pathways and donor design. Methods. 69(2):137-41.

Chawla R, Ariza-Nieto M, Wilson AJ, Moore SK and Srivastava V (2006). Transgene expression produced by biolistic-mediated, site-specific gene integration is consistently inherited by the subsequent generations. Plant Biotech J. 4:209-18.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA and Zhang F (2013). Multiplex genome engineering using CRISPR/Cas systems. Science. 339:819-23.

Dale EC and Ow DW (1991). Gene transfer with subsequent removal of the selection gene from the host genome. Proc. Natl. Acad. Sci. USA. 88(23):10558-62.

Day CD, Lee E, Kobayashi J, Holappa LD, Albert H and Ow DW (2000). Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. Genes & Dev. 14(22):2869-80.

De Paepe A, De Buck S, Nolf J, Van Lerberge E and Depicker A (2013). Site-specific T–DNA integration in A rabidopsis thaliana mediated by the combined action of CRE recombinase and ϕ C 31 integrase. The Plant J. 75(1):172-84.

Diaz V, Servert P, Prieto I, Gonzalez MA, Martinez AC, Alonso JC and Bernad A (2001).New insights into host factor requirements for prokaryotic beta-recombinase-mediated reactions in mammalian cells. J Biol Chem 276, 16257– 16264.

Didigu CA, Wilen CB, Wang J, Duong J, Secreto AJ, Danet-Desnoyers GA, Riley JL, Gregory PD, June CH, Holmes MC and Doms RW (2014). Simultaneous zinc-finger nuclease editing of the HIV coreceptors ccr5 and cxcr4 protects CD4+ T cells from HIV-1 infection. Blood. 123(1):61-9.

Dong OX and Ronald PC (2019). Genetic engineering for disease resistance in plants: recent progress and future perspectives. Plant Physiol. 180(1):26.

Du H, Zeng X, Zhao M, Cui X, Wang Q, Yang H, Cheng H and Yu D (2016). Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. J Biotech. 217:90-7.

Fan D, Liu T, Li C, Jiao B, Li S, Hou Y and Luo K (2015). Efficient CRISPR/Cas9-mediated targeted mutagenesis in Populus in the first generation. Scientific. Rep. 5:12217.

Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P, Cao F, Zhu S, Zhang F, Mao Y and Zhu JK (2013). Efficient genome editing in plants using a CRISPR/Cas system. Cell Res. 23(10):1229.

Foley JA, Ramankutty N, Brauman KA, Cassidy ES, Gerber JS, Johnston M, Mueller ND, O'Connell C, Ray DK, West P and Balzer C (2011). Solutions for a cultivated planet. Nature. 478(7369):337.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK and Sander JD (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotech. 31(9):822.

Gaj T, Gersbach CA and Barbas III CF (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol.31:397-405.

Gao J, Wang G, Ma S, Xie X, Wu X, Zhang X, Wu Y, Zhao P and Xia Q (2015). CRISPR/Cas9-mediated targeted mutagenesis in Nicotiana tabacum. Plant Mol Biol. 87:99-110.

Ghosh P, Wasil LR and Hatfull GF (2006). Control of phage Bxb1 excision by a novel recombination directionality factor. PLoS Biol. 4(6):e186.

Godfray HC, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM and Toulmin C (2010). Food security: the challenge of feeding 9 billion people. Science. 327(5967):812-8.

Golic KG and Lindquist S (1989). The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. Cell .59(3):499-509.

Guilinger JP, Pattanayak V, Reyon D, Tsai SQ, Sander JD, Joung JK and Liu DR (2014). Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. Nature Methods. 11(4):429.

Haun W, Coffman A, Clasen BM, Demorest ZL, Lowy A, Ray E, Retterath A, Stoddard T, Juillerat A, Cedrone F and Mathis L (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotech J. 7:934-40.

Hefferon K (2015). Nutritionally enhanced food crops; progress and perspectives. International J Mol. Sci. 16(2):3895-914.

Hou L, Yau YY, Wei J, Han Z, Dong Z and Ow DW (2014). An open-source system for in planta gene stacking by Bxb1 and Cre recombinases. Mol. Plant. 17(12):1756-65.

Ito Y, Nishizawa-Yokoi A, Endo M, Mikami M andToki S (2015). CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening. Biochem Bioph Res Co. 467(1):76-82.

Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA (2015). Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotech. 15(1):16.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA and Charpentier E (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science. 337(6096):816-21.

Jinek M, East A, Cheng A, Lin S, Ma E and Doudna J (2013). RNA-programmed genome editing in human cells. elife. 2: e00471.

Jung JH and Altpeter F (2016). TALEN mediated targeted mutagenesis of the caffeic acid O-methyltransferase in highly polyploid sugarcane improves cell wall composition for production of bioethanol. Plant Mol Biol. 92(1-2):131-42.

Kannan B, Jung JH, Moxley GW, Lee SM and Altpeter F (2018). TALEN-mediated targeted mutagenesis of more than 100 COMT copies/alleles in highly polyploid sugarcane improves saccharification efficiency without compromising biomass yield. Plant Biotech J. 16(4):856-66.

Kapusi E, Corcuera-Gómez M, Melnik S and Stoger E (2017). Heritable genomic fragment deletions and small indels in the putative ENGase gene induced by CRISPR/Cas9 in barley. Front Plant Sci. 8: 540.

Kapusi E, Kempe K, Rubtsova M, Kumlehn J and Gils M (2012). phiC31 integrase-mediated site-specific recombination in barley. PLoS One. Sep 14; 7(9):e45353.

Kazama T, Okuno M, Watari Y, Yanase S, Koizuka C, Tsuruta Y, Sugaya H, Toyoda A, Itoh T, Tsutsumi N and Toriyama K (2019). Curing cytoplasmic male sterility via TALEN-mediated mitochondrial genome editing. Nature Plants. 5 (7):722.

Kim JS and Pabo CO (1998). Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. Proc. Natl. Acad. Sci. USA. 17; 95 (6):2812-7.

Kim YG and Chandrasegaran S (1994). Chimeric restriction endonuclease. Proc. Natl. Acad. Sci. USA. 91(3):883-7.

Klimek-Chodacka M, Oleszkiewicz T, Lowder LG, Qi Y and Baranski R (2018). Efficient CRISPR/Cas9-based genome editing in carrot cells. Plant Cell Rep. 37(4):575-86.

Lee K, Zhang Y, Kleinstiver BP, Guo JA, Aryee MJ, Miller J, Malzahn A, Zarecor S, Lawrence-Dill CJ, Joung JK and Qi Y (2019). Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. Plant Biotech J. 17(2):362-72.

Li J, Manghwar H, Sun L, Wang P, Wang G, Sheng H, Zhang J, Liu H, Qin L, Rui H and Li B (2019). Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. Plant Biotech J. 17(5):858-68.

Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM and Sheen J (2013). Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and Nicotiana benthamiana using guide RNA and Cas9. Nature Biotech.31 (8):688.

Li T, Liu B, Spalding MH, Weeks DP and Yang B (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. Nature Biotech. 30(5):390.

Li Z, Moon BP, Xing A, Liu ZB, McCardell RP, Damude HG and Falco SC (2010). Stacking multiple transgenes at a selected genomic site via repeated recombinase-mediated DNA cassette exchanges. Plant Physiol. 154(2):622-31.

Li Z, Xing A, Moon BP, McCardell RP, Mills K and Falco SC (2009). Site-specific integration of transgenes in soybean via recombinase-mediated DNA cassette exchange. Plant Physiol. 151(3):1087-95.

Lloyd A, Plaisier CL, Carroll D and Drews GN (2005). Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. . Proc. Natl. Acad. Sci. USA. 102(6):2232-7.

Louwerse JD, van Lier MC, van der Steen DM, de Vlaam CM, Hooykaas PJ and Vergunst AC (2007). Stable recombinase-mediated cassette exchange in *Arabidopsis* using Agrobacterium tumefaciens. Plant Physiol.145 (4):1282-93.

Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, Wang B, Yang Z, Li H, Lin Y andXie Y (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol Plant 8(8):1274-84.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE and Church GM (2013). RNA-guided human genome engineering via Cas9. Science. 339(6121):823-6.

Mao Y, Zhang H, Xu N, Zhang B, Gou F and Zhu JK (2013). Application of the CRISPR–Cas system for efficient genome engineering in plants. Mol Plant 6(6):2008-11.

Matson PA andVitousek PM (2006). Agricultural intensification: will land spared from farming be land spared for nature? Cons Biol. 20(3):709-10.

Mikami M, Toki S and Endo M. (2015). Parameters affecting frequency of CRISPR/Cas9 mediated targeted mutagenesis in rice. Plant Cell Rep. 34(10):1807-15.

Minkenberg B, Xie K and Yang Y (2017). Discovery of rice essential genes by characterizing a CRISPR-edited mutation of closely related rice MAP kinase genes. The Plant J. 89(3):636-48.

Mouw KW, Rowland SJ, Gajjar MM, Boocock MR, Stark WM and Rice PA (2008). Architecture of a serine recombinase-DNA regulatory complex. Mol Cell. 30(2):145-55.

Nandy S and Srivastava V (2011). Site-specific gene integration in rice genome mediated by the FLP–FRT recombination system. Plant Biotech J. 9(6):713-21.

Nekrasov V, Wang C, Win J, Lanz C, Weigel D and Kamoun S (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. Sci. Rep. 7(1):482.

Olivares EC, Hollis RP and Calos MP (2001). Phage R4 integrase mediates site-specific integration in human cells. Gene, 278, 167– 176.

Onouchi H, Yokoi K, Machida C, Matsuzaki H, Oshima Y, Matsuoka K, Nakamura K and Machida Y (1991). Operation of an efficient site-specific recombination system of Zygosaccharomyces rouxii in tobacco cells. Nucleic Acids Res. 19(23):6373-8.

Osakabe K, Osakabe Y and Toki S (2010). Site-directed mutagenesis in *Arabidopsis* using custom-designed zinc finger nucleases. Proc. Natl. Acad. Sci. USA. 107(26):12034-9.

Osakabe K, Osakabe Y and Toki S (2010). Site-directed mutagenesis in *Arabidopsis* using custom-designed zinc finger nucleases. Proc. Natl. Acad. Sci. USA. 107(26):12034-9.

Pan C, Ye L, Qin L, Liu X, He Y, Wang J, Chen L and Lu G (2016). CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. Sci. Rep.6: 24765.

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA and Liu DR (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nature Biotech. 31(9):839.

Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, Wang N, Lee G, Bartsevich VV, Lee YL and Guschin DY (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nature Biotech. 26(7):808.

Petolino JF and Kumar S (2016). Transgenic trait deployment using designed nucleases. Plant Biotech J. 14(2):503-9.

Petolino JF, Worden A, Curlee K, Connell J, Moynahan TL, Larsen C and Russell S (2010). Zinc finger nuclease-mediated transgene deletion. Plant Mol Biol. 73(6):617-28.

Porteus MH and Baltimore D (2003). Chimeric nucleases stimulate gene targeting in human cells. Science. 300(5620):763-.

Puchta H and Fauser F (2014). Synthetic nucleases for genome engineering in plants: prospects for a bright future. The Plant J. 78(5):727-41.

Ray DK, Mueller ND, West PC and Foley JA (2013). Yield trends are insufficient to double global crop production by 2050. PloS One 8(6):e66428.

Römer P, Hahn S, Jordan T, Strauß T, Bonas U and Lahaye T (2007). Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. Science. 318(5850):645-8.

Rubtsova M, Kempe K, Gils A, Ismagul A and Weyen- J-Gils M (2008). Expression of active Streptomyces phage phiC31 integrase in transgenic wheat plants. Plant Cell Rep. 27:1821– 1831.

Russell SH, Hoopes JL and Odell JT (1992). Directed excision of a transgene from the plant genome. Mol Genet Genomics. 234(1):49-59.

Sander JD, Ramirez CL, Linder SJ, Pattanayak V, Shoresh N, Ku M, Foden JA, Reyon D, Bernstein BE, Liu DR and Joung JK (2013). *In silico* abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. Nucleic Acids Res. 41(19):e181.

Sauer B and Henderson N. (1990). Targeted insertion of exogenous DNA into the eukaryotic chromosome by the cre recombinase. New Biol. 2: 441– 449.

Schwikardi M and Droge P. (2000). Site- specific recombination in mammalian cells catalyzed by gammadelta resolvase mutants: implications for the topology of episomal DNA. FEBS Letters. 471: 147– 150.

Shen H, Strunks GD, Klemann BJ, Hooykaas PJ and de Pater S (2017). CRISPR/Cas9-induced double-strand break repair in *Arabidopsis* nonhomologous end-joining mutants. G3: Genes, Genom, Genet. 7(1):193-202.

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X and Choi VM (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature. 459(7245):437.

Srivastava V and Thomson J (2016). Gene stacking by recombinases. Plant Biotech J. 14(2):471-82.

Stoll S, Ginsburg DS and Calos MP (2002). Phage TP901-1 site-specific integrase functions in human cells. J. Bacter. 184:3657– 3663.

Sun Y, Zhang X, Wu C, He Y, Ma Y, Hou H, Guo X, Du W, Zhao Y and Xia L (2016). Engineering herbicide-resistant rice plants through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. Mol. Plant. 9(4):628-31.

Svitashev S, Schwartz C, Lenderts B, Young JK and Cigan AM (2016). Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes. Nature Comm. 7:13274.

Swarts DC and Jinek M (2018). Cas9 versus Cas12a/Cpf1. Structure–function comparisons and implications for genome editing. Wiley Interdisciplinary Reviews: RNA 9(5):e1481.

Tang X, Liu G, Zhou J, Ren Q, You Q, Tian L, Xin X, Zhong Z, Liu B, Zheng X and Zhang D (2018). A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. Genome Biol. 19(1):84.

Thomson JG and Ow DW (2006). Site- specific recombination systems for the genetic manipulation of eukaryotic genomes. Genesis. 44:465– 476.

Thomson JG, Chan R, Smith J, Thilmony R, Yau YY, Wang Y and Ow DW (2012). The Bxb1 recombination system demonstrates heritable transmission of site-specific excision in *Arabidopsis*. BMC Biotech. 12: 9.

Thomson JG, Yau YY, Blanvillain R, Chiniquy D, Thilmony R and Ow DW (2009). ParA resolvase catalyzes site-specific excision of DNA from the *Arabidopsis* genome Transgenic Res. 18: 237– 248.

Thorpe HM, Wilson SE and Smith MC (2000). Control of directionality in the site-specific recombination system of the Streptomyces phage φC31. Mol Microbiol. 38(2):232-41.

Tian S, Jiang L, Gao Q, Zhang J, Zong M, Zhang H, Ren Y, Guo S, Gong G, Liu F and Xu Y. (2017). Efficient CRISPR/Cas9-based gene knockout in watermelon. Plant Cell Rep. 36(3):399-406.

Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK and Voytas DF (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature. 459(7245):442.

Ueta R, Abe C, Watanabe T, Sugano SS, Ishihara R, Ezura H, Osakabe Y and Osakabe K (2017). Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. Sci. Rep. 7(1):507.

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD and Holmes MC (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature. 435(7042):646.

Veres A, Gosis BS, Ding Q, Collins R, Ragavendran A, Brand H, Erdin S, Cowan CA, Talkowski ME and Musunuru K (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell. 15(1):27-30.

Vergunst AC and Hooykaas PJ (1998). Cre/lox-mediated site-specific integration of Agrobacterium T-DNA in *Arabidopsis* thaliana by transient expression of cre. Plant. Mol. Biol. 38(3):393-406.

Voytas DF (2013). Plant genome engineering with sequence-specific nucleases. Ann. Rev. Plant Biol. 64: 327-350.

Wang S, Zhang S, Wang W, Xiong X, Meng F and Cui X (2015). Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. Plant. Cell. Rep. 34(9):1473-1476.

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C and Qiu JL (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nature Biotech. 32(9):947.

Wang Y, Geng L, Yuan M, Wei J, Jin C, Li M, Yu K, Zhang Y, Jin H, Wang E and Chai Z (2017). Deletion of a target gene in Indica rice via CRISPR/Cas9. Plant Cell Rep. 36(8):1333-1343.

Wang Z, Wang S, Li D, Zhang Q, Li L, Zhong C, Liu Y and Huang H (2018). Optimized paired-sgRNA/Cas9 cloning and expression cassette triggers high-efficiency multiplex genome editing in kiwifruit. Plant Biotech J .16(8):1424-33.

Wiedenheft B, Sternberg SH and Doudna JA (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature. 482(7385):331.

Wright DA, Townsend JA, Winfrey Jr RJ, Irwin PA, Rajagopal J, Lonosky PM, Hall BD, Jondle MD and Voytas DF (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. The Plant J. 44(4):693-705.

Xie K, Minkenberg B and Yang Y (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. USA. 112(11):3570-5.

Yee JK (2016). Off-target effects of engineered nucleases. FEBS J. 283(17):3239-48.

Young J, Zastrow-Hayes G, Deschamps S, Svitashev S, Zaremba M, Acharya A, Paulraj S, Peterson-Burch B, Schwartz C, Djukanovic V and Lenderts B (2019). CRISPR-Cas9 editing in maize: systematic evaluation of off-target activity and its relevance in crop improvement. Sci. Rep. 9(1):6729.

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, Van Der Oost J, Regev A and Koonin EV (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 163(3):759-71.

Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, DeGennaro EM, Winblad N, Choudhury SR, Abudayyeh OO, Gootenberg JS and Wu WY (2017). Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. Nature Biotech. 35(1):31.

Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T and Joung JK (2010). High frequency targeted mutagenesis in *Arabidopsis* thaliana using zinc finger nucleases. Proc. Natl. Acad. Sci. USA 107(26):12028-33.

Zhang Q, Xing HL, Wang ZP, Zhang HY, Yang F, Wang XC and Chen QJ (2018). Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in *Arabidopsis* and its prevention. Plant Mol Biol. Mar 1; 96(4-5):445-56.

Zhou H, Liu B, Weeks DP, Spalding MH and Yang B (2014). Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Res. 42(17):10903-14.

Zhu J, Song N, Sun S, Yang W, Zhao H, Song W and Lai J (2016). Efficiency and inheritance of targeted mutagenesis in maize using CRISPR-Cas9. J Genet Genome. 43(1):25-36.

CHAPTER II

EVALUATION OF THE STRUCTURAL AND EXPRESSION STABILITY OF THE MULTIGENE STACK IN RICE DEVELOPED THROUGH CRE-*LOX* SITE-SPECIFIC INTEGRATION

Abstract

A multigene locus coding for multiple traits is an important new breeding technique. Random DNA integration such as integration of large T-DNA and targeted integrations based on double-stranded break repair and site-specific recombination mechanisms have been used for stacking multiple genes into a single locus. However, investigations on the stability of the multigene stacked locus are limited. Here, a multigene locus developed by Cre-*lox* site-specific recombination system in rice was studied in 28 independent lines and their progeny. This site-specific integration locus consisted of 5 genes consisting of 3 expressed by strong constitutive promoters (*Ubi: NPT, 35S: GUS, 35: GFP*) and 2 expressed by inducible promoters (*AtRD29a:AtDREB1A* and *HSP: pporRFP*). Twenty-one of these recovered site-specific integration lines contained a full-length integration of the 5-gene stack, and expressed the constitutive and inducible genes according to their promoter specificity. Gene expression of *NPT*, *GUS* and *GFP* as determined by enzyme activity or protein levels in the progeny plants was found to be similar among site-specific integration lines, and showed correlation with allelic state of the locus. Expression of inducible genes (*AtDREB1A* and *pporRFP*) by heat- or cold-inducible promoters was also found to be duly regulated by heat or cold treatments. These data indicate that Cre-*lox* site-specific recombination in rice generates a high rate of precise full-length integrations of multigene DNA fragments. The resulting multigene stacked locus stably expresses each gene in the primary transgenic plants and their progeny, and the expression of inducible genes in the stacked locus was not disturbed by the surrounding strong promoters. In conclusion, Cre-*lox* site-specific integration is an effective approach for developing multigene stacked locus expressing constitutive or inducible genes.

Introduction

Continuous development of the practical approaches for gene stacking is important for crop improvement, as often-multigene introduction for expression of complex traits such as disease resistance, agronomic characters are required. In conventional breeding, an increase in the number of genes exponentially increases the number of F2 plants needed to screen for multigene stacked lines. Combining transgenes by breeding is also challenging, as it requires multiple rounds of crossings to generate a pure line and limit linkage drag (Srivastava and Thomson, 2016). Through biotechnology, however, concern for linkage drag is removed as insertion of the gene could occur directly into the cultivated variety. Introduction of complex traits would require integration of multiple genes. By stacking these genes into one chromosomal block or genetic locus, breeding into multiple adapted cultivars would be greatly simplified. Thus, strategies for inserting multiple genes into a single locus are needed. Further, strategies are needed for directing genes into specified genomic sites to avoid disruption of host genes and creating unfavorable mutations (Petolino and Kumar, 2016). Therefore, targeted gene integration can be deployed for creating multigene stacks in the plant genomes. Two different approaches of targeted integrations are available: 1) double-stranded break (DSB) repair, and 2) site-specific recombination (SSR) (Petolino and Kumar, 2016; Srivastava and Thomson, 2016). In the first type, engineered nucleases like ZFN, TALEN and CRISPR-Cas9 generates a DSB in the genome, which stimulates the cellular repair machinery. The transgene integrates into the genome as a by-product of the DNA repair between the targeted cleavage site (Kumar et al. 2016; Moehle et al. 2007; Petolino and Kumar, 2016), albeit at a lower frequency (Cai et al. 2009; D'Haullin et al. 2008; Wright et al. 2005). For example, in maize, using zinc finger nuclease, promoter-less herbicide resistance gene was introduced using an endogenous promoter-

trap strategy (Shukla et al. 2009). However, the precise events recovered were five-fold less than the random integration events produced by conventional transformation approach. Through targeted integration approaches, transgene integration could be directed to the 'safe harbor' locations in the genome, which allow high and stable expression of the transgenes without interfering the neighboring gene function (Petolino and Kumar, 2016). In humans, these regions were identified to be 50 kb from the *5'* end of any gene, at least 300 kb from any cancer related gene, ~ 300 kb from any microRNA, location outside a transcription unit, and location outside the ultra-conserved region. Introduction of beta-globin transgene in these safe harbor regions of thalassemia-induced pluripotent stem cells, led to higher expression of the gene, without affecting the neighboring gene's expression. (Papapetrou et al. 2011). In plants, these sites are proposed to be in the non-coding regions. Cantos et al. (2014) studied 'safe harbors' in rice by introducing the *GUS* and directing integrations by ZFN that could target multiple coding or non-coding genomic sites. Authors analyzed >100 transgenic events that mapped to 28 genomic regions but found only 1 that was in non-coding region and showed high expression. Thus, identification of 'safe harbors' in the plant genomes will require experimental validation through transgene integration and gene expression analysis. However, identification of safe harbors in polyploid plants could become more complicated.

Site-specific recombination (SSR) driven by well-characterized SSR systems such, as Cre-*lox* is a simple reaction leading to predictable outcomes (Gaj et al. 2014; Grindley et al. 2014; Ow 2002; Sauer, 1994; Srivastava and Thomson, 2016). The Cre-*lox* recombination works efficiently in many plant cells that have been used for different applications including transgene integration (Day et al. 2000; Srivastava et al. 2004).

Other SSR systems that have been successfully used in obtaining the site-specific integrations (SSI) in plants include FLP-*FRT*, R-*RS* and Bxb1 (Hou et al, 2014; Nandy and Srivastava, 2011; Nanto and Ebinuma, 2008). Site-specific integrations by SSR are generally recovered at a high rate and contain precise integrations. In two separate studies done on tobacco and rice, site-specific integration events by Cre-*lox* were recovered at equal or higher rates in comparison to the conventional transformation approaches such as *Agrobacterium* mediated random T-DNA integration, particle bombardment or protoplast transformation (Day et al. 2000; Srivastava et al. 2004). About 80% of the recovered events contained precise site-specific integrations, 50% of which were single-copy (SC) site-specific integrations (SSI) devoid of additional random integrations (Srivastava et al. 2004).

The SC-SSI lines of tobacco and rice were found to express the transgene at more or less same levels between transgenic lines (Chawla et al. 2006; Day et al. 2000; Srivastava et al. 2004) however, tobacco SSI lines developed by protoplast transformation method also showed gene silencing that was correlated with promoter hyper-methylation (Day et al. 2000). No silencing was observed in rice SC-SSI lines developed by gene gun method indicating the role of foreign DNA dosage and transient overexpression in DNA methylation.

The FLP-*FRT* recombination system is also effective in directing transgene integrations in rice and maize (Li et al. 2009; Nandy et al. 2011). Transgene expression produced by SSI locus generated by FLP-*FRT* recombination, when DNA was delivered by gene gun in rice, was also found to be within two – three-fold variation between independent transgenic lines (Nandy and Srivastava, 2012). This indicates SSI locus developed by gene gun, owing to its precise integration structure is expressed predictably, and not subject to epigenetic modifications triggered by transient overexpression. Finally, SSI lines of tobacco developed by R/RS site-

specific recombination system were found to express transgene at similar levels (Nanto et al. 2009).

Molecular stacks with multiple traits is a challenging task, as it may require a sequential transformations (Petolino and Kumar, 2016) and have more chances of random integrations, which affect both, stability and expression of the genes within the locus. In the trait stacking, it is also necessary to introduce the more number of regulatory elements (Que et al. 2010) for a broad-spectrum trait development like having simultaneous traits herbicide resistance and higher yield genes stacked in one locus. In the trait stacking studies (irrespective of the approaches used for the transgenes integration), only a few genes expressed under the constitutive promoters have been analyzed for their stability and expression over the successive generations (Ainley et al. 2013; Akbudak et al. 2010; Chawla et al. 2006; Srivastava et al. 2004).

The current study attempted to stack five genes, three genes under constitutive promoters and two under inducible at a single locus using the Cre-*lox* mediated recombination. We recovered >75% precise SSI events, all of which showed stable, heritable expression of all five stacked genes at transcript and/or protein levels. Expression of the two inducible genes controlled by heat or cold-inducible promoters was found to be properly regulated as indicated by low/undetectable expression at room temperature and abundant expression upon heat or cold treatments. Similar to previous reports (Akbudak et al. 2010; Chawla et al. 2006), this study also found higher gene expression in biallelic homozygous lines as compared to the monoallelic hemizygous lines.

This study validated the feasibility of gene stacking by Cre-*lox* recombination, and determined the stability of the genes within the stack in rice for the development of gene stacking methods for crop improvement.

Materials and Methods

Vector construction

The multigene vector, pNS64 (Fig. 1b), was developed for the current study. This vector was developed through standard restriction cuts and ligation methods. The *pporRFP* and gene cassettes from pUC vectors were ligated one by one into pAA12 backbone that contains a promoterless neomycin phosphotransferase II (*NPTII*) gene followed by 35S:GFP:nos*3'* and 35S:GUS:nos*3'* cassettes between *loxP* and *lox75* (Akbudak and Srivastava, 2017). Hence , pNS64 contained between *loxP* and *lox75* following fragments: 1) promoterless selection marker gene, neomycin phosphotransferase (*NPT II*), 2) green fluorescent protein (*GFP*) under CaMV 35 S promoter, (3) β-glucuronidase (*GUS*) under CaMV 35S promoters, 4) *Arabidopsis thaliana* dehydration responsive element B1A (*AtDREB1A)* under *Arabidopsis* cold inducible *rd29a* promoter, and 5) red fluorescent protein (*pporRFP*) from coral *Porites porites* (Alieva et al. 2008) under soybean heat inducible *Gmhsp17.5E* promoter. The *pporRFP* was obtained from pANIC 6A vector, while *rd29a* and *AtDREB1A* were amplified from *Arabidopsis* genomic DNA. For the vector assembly, first individual vectors of *Gmhsp17.5e: pporRFP* (pNS54) and *Atrd29a:AtDREB1A* (pNS55) were generated. The pNS54 was cut with *Xba*I and ligated with pAA12 (Akbudak and Srivastava, 2017) to generate pNS59. Later, pNS59 was *Bgl*II digested, dephosphorylated with CIP, and ligated with *Bgl*II digested pNS55 to generate a five multigene construct pNS64.

Rice transformation

The rice line T5 (Taipei 309) which contains a Cre-*lox* target site as determined by pVS52 construct (Fig. 1a) and described by Srivastava and Ow (2002) was used in the present study. The scutellar callus of T5 was developed on 2N6D media. The five µg of plasmid pNS64

coated on 1µm gold particles was bombarded by PDS1000/He gene gun on 3-4 weeks old scutellar callus. The bombarded callus was selected on 100 mg/l geneticin^{TM} to isolate the site-specific integration (SSI) lines, which were transferred to the regeneration media supplemented with 100 mg/L geneticin to develop transgenic SSI plants. All tissue culture protocols were followed as mentioned by Nishimura et al (2006).

PCR and Southern analysis

The primary transgenic SSI plant lines and T1 progeny were subjected to PCR and Southern analysis. The genomic DNA was isolated using CTAB method and checked on the 0.8% agarose gel. The polymerase chain reaction (PCR) was performed using Emerald Amp MAX PCR Master Mix (Takara Bio, CA, USA) using the primers given in Table 3. The PCR cycling conditions consisted of initial template denaturation at 95°C for 4 min followed by 40 cycles of 95°C for 1min, annealing at 58/60° for 1 min, 72°C extension at 1 or 2 min depending on amplicon length and the final extension at 72°C for 15 minutes. Southern blot analysis was performed on these plants using ^{32}P-labeled DNA probes of GFP, RFP, GUS, and AtDREB1A. Genomic DNA were digested with *Eco*RI overnight, fractioned on 0.8% agarose gel, blotted on nylon membrane, and hybridized with the probes using the standard southern hybridization method.

T1 seedlings germination

T1 seedlings of the SSI lines were used for expression analysis and protein assays. The T1 seeds of the SSI lines were germinated on ½ MS media without selection for 7-10 days. All the seedlings were tested for GFP and/or GUS activities before using them for the gene expression (RT-qPCR) and protein assays. T5 was used as a negative control.

Primer efficiency evaluation for expression analysis

The qPCR primers were designed using IDT primer quest tool and the melt curve was predicted using U-melt (Dwight et al. 2011). For each gene, two-three primer pairs were tested for its efficiency. The efficiency for *AtDREB1A* and *pporRFP* primers was tested on 2-fold diluted genomic DNA, while for *NPT, GUS,* and *GFP,* 10-fold diluted cDNA was used. The list of the primers is given in Table 3.

Expression analysis by RT-qPCR

For *GUS, GFP* and *NPT* qPCR, the 7-10 days old seedlings maintained at room temperature were used for the expression analysis**.** While, for *AtDREB1A*1a and *pporRFP*, the seedlings were cold-shocked on ice for 20 hours or heat-shocked for 3 hours at 42°C, respectively. The respective controls of the *AtDREB1A* and *pporRFP* were maintained at the room temperature. Total RNA was isolated using Trizol (Invitrogen) and quantified using Nano-drop 2000 (Thermo-Fisher Inc). Two microgram of total RNA was treated with RQ1-RNAse free DNase (Promega Inc) for the removal of genomic DNA, and one microgram of the DNase-treated RNA was used for the cDNA synthesis using PrimeScript RT reagent kit (Takara Bio, CA, USA). The expression analysis was performed using TB green Premix Ex Taq II (Takara Bio, CA, USA) on Bio-Rad CFX 96 C1000 with following conditions: 95°C for 30 sec and 40 cycles of 95°C for 5 sec + 60°C for 30 sec. The product specificity was verified by the melt curve analysis. The C_t values of all the genes were normalized against 7Ubiquitin or Ubiquitin fused protein reference genes (Pabuayon et al. 2016; Xie et al. 2015). The relative expression was calculated against T5 negative control and the untreated controls (room temperature) using delta-delta C_t method (Livak and Schmittgen, 2001). Each line contained two to three biological replicates with two technical replications.

GUS fluorometric analysis

GUS activity in the leaves was detected by Jefferson (1987) method. The young leaf tissue was submerged in the GUS staining solution consisting of the 1 mM X-Gluc (Gold Biotechnologies, St. Louis, MO, USA) and incubated at 37 °C overnight. The quantitative measurement of the GUS activity was done as described in the Versa-flurometer (Bio-Rad, CA) guide. Briefly, the protein from ~50-100mg 10-day-old seedling was extracted in the GUS extraction buffer (50 mM $NaPO_4$, 10 mM β-mercaptoethanol (β-ME), 10 mM Na_2EDTA, 0.1% SDS and 0.1% triton X-100). The total protein was estimated using Bradford reagent (VWR). All samples were normalized to 10 µg of the total protein for the quantitative measurement of the GUS activity. All normalized samples were added to 500 µl of assay buffer (100 mM 4-methylumbelliferyl b-D-glucuronide, MUG; β-ME and GUS extraction buffer) and were incubated for 30 minutes, followed by the addition of 1 ml of the 1x stop buffer (Na_2CO_3). The activity was detected in the Versa-flurometer equipped with 360± 5nm excitation filter and 390± 5nm emission filter. A standard curve was prepared with the dilution series of 4-methylumbelliferone (4-MU) in the stop solution for the calculation of the GUS activity. A unit of GUS activity was defined as nmol 4-MU produced per minute from each milligram of the soluble protein (nmol/min/mg)

GFP fluorometric assay

The GFP expression was checked in the 5-10 day old seedlings under the Leica 56D stereoscope fitted with the 440-460nm excitation and 500-560nm (band pass) emission filters (Night Sea, Lexington, MA). For the quantitative estimation, the GFP positive seedlings were ground in extraction buffer consisting of 10 mM Tris–EDTA, pH 8.0 at 4 °C and centrifuged at 13,000 rpm for 20 min to collect the supernatant. In this experiment, a high GFP expressing line,

C30-1, generated in the Zinc Finger Nuclease study was used as a reference. The total protein was estimated using Bradford reagent (VWR). All samples were normalized to 10 µg of protein for the quantitative estimation. The expression (fluorescence) was estimated using Versa-flurometer (Bio-Rad Inc) equipped with 490 ± 5 nm excitation filter and a 510 ± 5 nm emission filter. The 19000 range and low gain was set using C30-1 extract, and all T1 lines were measured against it. A unit of GFP was defined as relative fluorescence units/ten microgram of total protein (RFU/10 µg of total protein).

NPTII ELISA

NPTII enzyme linked immunosorbent assays (ELISA) were conducted according to the manufacturer's instructions (Agdia, Elkhart, IN). Briefly, ~50 mg of fresh leaf from 1-month old greenhouse grown plants samples were ground in the protein extraction buffer (PEB1) provided in the kit. For ELISA, the protein extracts and the enzyme conjugates were sequentially added, followed by wash steps as per the manufacturer's instructions. The NPTII provided in the kit and the T5 protein extract were used as a positive and negative controls respectively. ELISA plates were read at A650 in the Synergy Biotek Cytation 3. The ratio of the absorbance of samples to T5 negative control was used as the measure of NPTII expression. For each line, two to three biological replicates were tested with the two technical replicates.

Confocal Imaging

The confocal microscopy for the detection of RFP and GFP in the 7-10 days T1 seedlings was performed at Arkansas Nano and Bio Materials Characterization Facility, University of Arkansas, Fayetteville. The seedlings were heat-shocked as described in the expression analysis section. The imaging was done at 24, 48 and 72 hours post heat shock in the roots. The images were captured using a Leica TCS SP5 (Buffalo Grove, IL. USA) microscope by the bandwidth

adjustment for the fluorescence detection. For roots imaging, the samples were excited using 514 Argon and 594 HeNe laser channels and emission was collected at 542-582 mm for GFP and 610-710 nm for RFP. For leaf imaging, samples were excited at 514 Argon laser channel and emission was collected at 590-610 nm for blocking the chlorophyll auto-fluorescence. The leaf images were captured through sequential scan to prevent the bleed-through between chlorophyll auto-fluorescence and fluorescent protein(s). Since, the T1 seedlings contained constitutively expressed GFP, imaging of green fluorescence was used to locate the tissue in which red fluorescence was subsequently determined. The GFP positive C30-1 and the parental T5 seedlings were used as controls. Using C30-1 seedlings, it was ensured that the RFP signals originated from the RFP emission spectra, and not from the bleed-through from GFP Argon laser channel. For all samples, first the gain, zoom and offset was adjusted for T5 negative control, and then all images were captured using the same parameters at 20x magnification.

Results

Molecular Strategy

This study utilized the Cre-*lox* mediated site specific integration at a *T5* locus in rice cv. Taipei 309. This locus has a single copy of T-DNA (Fig. 1a.) containing a *lox76* site that serves as the target of gene integration through *lox75* x *lox76* recombination catalyzed by Cre recombinase (Akbudak and Srivastava, 2017; Akbudak and Srivastava, 2011; Srivastava et al. 2004; Srivastava and Ow, 2002). The *lox76* site is placed between maize ubiquitin promoter (*ZmUbi1*) and the *cre* coding sequence. The donor DNA pNS64 (Fig. 1b) contains genes-of-interests between *lox75* and *loxP* sites along with a promoter-less marker gene for selecting site-specific integrations through promoter-trap strategy. Upon delivery of the pNS64 into Cre-expressing T5 cells, *lox75* and *loxP* would undergo rapid recombination separating gene

construct from the vector backbone. Next, recombination between of the gene construct circle containing *lox75* site with *lox76* site at the *T5* locus will result in site-specific integration of the genes. This integration structure is selected on geneticin™ as *NPT* gene is turned on through promoter trapping at the *T5* locus. The recombination between *lox75* and *lox76*, the two single mutant *lox* sites, generates a double-mutant *lox* at the integration site, preventing reversibility of the recombination (Fig. 1c). The SSI plant lines developed by the biolistic delivery of pNS64 into T5 line were analyzed by PCR and Southern blot hybridization to determine integration structure and copy number, followed by gene expression analysis of the stacked genes: *NPT, GFP*, *GUS*, *AtDREB1A,* and *pporRFP* genes.

Characterization of transgenic lines

A total of 29 geneticin-resistant, primary transgenic plant lines (T0) were obtained by transformation of T5 line with pNS64. One of which was albino, and therefore, removed from the study. The 28 putative SSI lines were subjected to molecular characterization by PCR and Southern hybridization (Table 1). PCR with primers Ubi1960 and KanR and BamH1pporRFP and Cre2333 indicated the presence of predicted SSI junctions (Fig. 1c) .The primer pair Ubi and RevATG was used to check if the biallelic/monoallelic integration had occurred at *T5* site (Fig. 1a). The PCR analysis revealed that all the 28 lines contained the predicted SSI structure (Fig. 2a-b), except line #1, which appeared to be truncated at junction 2. The PCR with Ubi and RevATG revealed that six lines lacked PCR amplification indicating biallelic integration at *T5* site (Fig. 2c, Table 1). Next, the genomic DNA of the SSI lines was digested with *Eco*RI and probed with *GFP*, *pporRFP*, *AtDREB1A,* and *GUS* on a Southern blot. All 28 lines showed the presence of the predicted 3.2 kb band on *GFP* hybridization, confirming precise junction 1, and 27 lines showed 2.1 kb band on *pporRFP* hybridization, confirming SSI junction 2 (Fig. 3a).

Fifteen of 28 SSI lines also contained random integrations, indicated by additional bands on Southern blots (Fig. 3a). Subsequent hybridizations with *GUS* and *AtDREB1A* probes showed the expected 2.5 kb and 2.1 kb bands, respectively, in 18 SSI lines (Fig. 3b). The remaining either did not show hybridization or showed a lower or higher band, indicating truncation in the SSI structure (Fig. 3b; Table 1) Southern hybridization analysis clearly distinguished between single copy (SC) and multi-copy (MC) lines (Fig. 3a-b). SSI lines that are free of additional random integrations are called SC, while those that contain additional integrations are called MC lines. Southern hybridization also revealed clonal lines among MC lines indicated by the presence of identical hybridization pattern (Fig. 3a-b; Table 1). Three clonal groups (lines 2 and 3, 5 and 6, 16 and 17) were identified by *GFP* or *pporRFP* hybridization (Fig. 3a-b). Accordingly, only one of the two clonal lines was used in subsequent study. A subset of 17 SSI lines was subjected to PCR with GusF982 and cre2333 to determine the presence of a full-length integration from *GUS* through *pporRFP*. Amplification of the 4 kb fragment in this PCR indicated full-length integration and corroborated with Southern data. Lines that lacked *GUS* or *AtDREB1A* integration in Southern blots failed to amplify 4 kb band, while that showed expected bands on Southern blots showed 4 kb band (Fig. 2d and 3a-b). A total of 13 SC and 8 MC SSI lines containing all 5 genes were recovered, while the remaining 7 contained imprecise junction or a truncation within the structure. Of these, 9 SC and 6 MC SSI lines were healthy and fertile, while the remaining 4 SC and 2 MC lines did not set the seeds. Progeny seedlings (T1) of the 15 fertile SSI lines were screened by GFP expression using fluorescence stereoscope. Twelve lines produced both GFP+/GFP- progeny, while the remaining 3 generated all GFP+ progeny. This data agrees with PCR prediction of monoallelic/biallelic integration in these SSI lines (Table 1). All fertile SSI lines were included in gene expression analysis.

Transgene expression analysis

Transgene expression was studied at transcriptional and post-translational levels *i.e.* by quantifying mRNA levels and/or measuring protein activity in T1 progeny of the SSI lines. Based on the T1 seed availability, 12 lines (9 SC and 3 MC) were used for the qRT-PCR analysis, and 17 lines [9 SC, 8 MC including 1 truncated (#15), and 1 imprecise (#1)] for the protein assays (Table 1). Among 9 SC lines, three lines #11, 20 and 29 had biallelic integration, while other 6 were monoallelic. The analysis at transcript levels (RT-qPCR) was performed on all the five genes to verify their expression. The protein assay was performed on the four genes, where GFP and GUS activity was measured by fluorometric assay, NPT by ELISA and pporRFP fluorescence by confocal microscopy. Line #1 and 15 were included only in the protein assay. The results for protein and transcript expression will be shown as the comparisons between 1) SC and MC; and 2) Monoallelic and biallelic integrants of SC lines.

Transcript levels

Primer efficiency evaluation

When the multiple genes are stacked at a single locus, their expression levels vary due to the stability of locus, promoter strengths and pattern of integration. The variability in the expression levels among different sample types, quality of cDNA, copy number of the transcripts contribute to the dissimilarities in the qPCR efficiency and thus leads to erroneous results (Ruijter et al. 2013; Sreedharan et al. 2018). Therefore, the primer efficiency for each of the genes was also evaluated on either cDNA or genomic DNA. For all genes, the correlation coefficient (R^2) ranged from 0.955 to 1 suggesting a reliable reproducibility of the results but the efficiency varied from 58% to 118%. Given the balanced GC content of the *pporRFP* and *AtDREB1A*, the qPCR efficiency of these genes ranged from 83 - 108%, while in the *GFP, NPT*

and *GUS*, a large variation in the PCR efficiencies were observed, which was due to the higher GC content of the genes (Bustin and Hugget, 2017) (Table 2). Only highly efficient primers were selected for qRT-PCR analysis, efficiencies of which were calculated between 93 – 106% (Table 2).

Constitutive NPT, GUS, and GFP transgene transcript abundance in SSI lines

Twelve SSI lines were subjected to qRT-PCR and the transcript levels were quantified relative to the T5 negative control. In *NPT*, expression ranged from 600 – 4000x among 12 lines (Fig. 4a). The highest expression level of ~4000x was observed in the SC line #12, and the lowest in the MC line #19. Although more variation in transcript levels was estimated within the SC lines, SC lines in general, had three-fold higher expression than the MC lines (Fig. 4a).
In *GFP*, all 12 lines showed the transcript levels ranging from 28,000 –120,000x. The highest expression was observed in the SC line #11 and the lowest in MC line #16 (Fig. 4b). Like *NPT*, the transcript levels of *GFP* varied to higher extent within SC lines; however, MC lines displayed somewhat lower levels than the SC lines (Fig. 4b).

In *GUS*, the transcript levels ranged from 64 – 5000x. The highest expression of 5000x was observed in SC line #11, while the lowest expression of 64x was seen in the SC line #12 (Fig.4c). Thus, a greater variability of transcript levels was observed in the *GUS* gene with three lines expressing 64 – 750x (#9, 10, and 12). Among SC and MC, no significant difference in the expression levels was observed.

Among 6 monoallelic and 3 biallelic SC lines (Fig. 2c, Table 1), no significant difference in the *NPT* transcript levels were observed (Fig. 4d), while 2.5x higher transcript levels were estimated for *GFP* and *GUS* in the biallelic lines as compared to monoallelic lines (Fig. 4e, f).

Inducible AtDREB1A and RFP gene expression analysis in SSI lines

The *AtDREB1A* and *pporRFP* were placed under cold-inducible *AtRD29A* and heat-inducible *GmHSP17.5e* promoters, respectively. For *AtDREB1A* expression analysis, ten-day-old seedlings were cold-shocked on ice for 20 hours, and for inducing *pporRFP* expression analysis, seedlings were heat-shocked for three hours at 42ºC. The controls for both genes were maintained at the room temperature. The expression was calculated relative to T5 negative and the room temperature controls.

In *AtDREB1A,* relative to T5, the 12 lines showed cold-induced expression from 200-1200x (Fig. 5a). Line #9 had highest expression of 1200x, and line #21 had the lowest expression of 200x. The room temperature expression in these lines ranged from 8-40x (Fig. 5a). Relative to treatment, all 12 lines showed fold-induction that ranged from 8 – 75x. Line #14 had highest induction levels of 75x, and line #27 had the lowest induction level of 8x (Fig. 5b). When the expression levels between SC and MC lines were compared, no significant changes in the expression levels were observed (Fig 5a, b). In the SC lines, the expression levels between monoallelic and biallelic integrants (Fig. 5c-d) were also found to be similar (Fig. 5c-d) .

In *pporRFP,* heat-induced expression levels ranged from 60 - 870x in the SSI lines relative to T5 (Fig. 6a). The highest expression of 870x was observed in line #10 and lowest of 60x in line #27. No significant difference in the expression levels were seen between SSI lines (Fig. 6a). With respect to treatment (42° C for 3 h), the induced expression levels ranged from 25 - 2100x. The highest fold-induction was observed in line #19 and the lowest in line #27 (Fig. 6b). A greater variability in the induction levels within SC and MC lines were observed. For example, SC line #27 had lowest induction level of 25x, while line #29 had induction level of 600x. A similar trend was observed in MC lines, where line #19 had highest induced levels of 2100x,

while lines #14 and 16 had only ~300x induced levels. However, the induction levels were not significantly different between SC and MC lines (Fig. 6b). In the SC lines, the biallelic lines did not show significant increase in induced expression when compared to that of the monoallelic lines (Fig. 6c-d).

In summary, all SSI lines in this study were found to properly express the 5 stacked genes at the *T5* locus. The genes controlled by strong constitutive promoters (*Ubi: NPT, 35S: GFP, 35S: GUS*) showed strong levels of transcripts in the 9 SC and 3 MC lines. The inducible genes (*AtRD29a:AtDREB1A* and *HSP: pporRFP*) in all SSI lines were found to express at basal levels, and enhance abundantly upon cold or heat-treatment. The correlation of allelic state with transcript abundance was also observed in a subset of lines, especially in *GFP* and *GUS* genes.

Estimation of protein levels:

All protein assays were carried out using 17 SSI lines that consisted of 9 SC and 8 MC lines with T5 as negative control. In NPT II ELISA, all the 17 lines tested positive for NPT II (Fig. 7a) and the absorbance ranged from 12 - 25 relative to T5 negative control. The lowest absorbance ratio of 12 was observed in line #12, and highest in line #30. No significant difference was observed between biallelic (average ratio of 23) and monoallelic (average ratio of 22) SC lines (Fig. 7b) or between SC and MC lines (average ratio of 21 for both) (Fig. 7c). The NPT II ELISA and subsequent GUS and GFP protein assays also included truncated line #15 and the imprecise line #1. Line #15 had a truncation of the *GUS* and *AtDREB1a* gene, while other three genes were present (Fig. 3a-b; Table 1) and line # 1 had only first junction containing *NPT* and *GFP*, while other three genes were absent (Fig. 2d, 3a; Table 1). Both of these lines also tested positive for NPTII and had ratio of 19 and 23 for line #1 and #15, respectively (Fig. 7a).

In *GFP*, all the lines tested positive for GFP fluorescence with the fluorescence level ranging from ~7300 - 2280 RFU/10µg of total protein. Among all lines, SC line #11 had the highest expression of 7299 RFU and line #1 had the lowest of 2286 RFU. The truncated line #15 had GFP expression of 4093 RFU, almost in the similar range as observed in other MC lines that carried full-length integration. Among SC lines, the average GFP levels in the biallelic lines (6668 RFU) were almost two-fold higher than that in the monoallelic lines (4273 RFU; Fig. 8b). However, no significant difference was found among SC (average of 5132 RFU) and MC (4331 RFU) lines (Fig. 8c).

In GUS, all lines, which showed 2.5 kb band in the Southern blot (Fig. 3a), tested positive for the GUS staining. In PCR or Southern blot analysis, 15 SSI lines were found to contain *GUS* gene, while the remaining 2 lines lacked *GUS* integration. Accordingly, 15 SSI lines showed histochemical GUS staining (Fig. 9a). Estimation of GUS enzymatic activity by MUG assay in these lines showed that the activity ranged from 70 - 280 nmol/min/mg protein (Fig. 9b). The highest and lowest GUS activities were detected in line #29 and #10, respectively. As predicted truncated lines, #1 and #15 did not show GUS activity (Fig. 9a-b). Similar to GFP, the biallelic SSI lines (average of 213 nmol/min/mg of protein) had 2x higher GUS activity than the monoallelic (average of 132 nmol/min/mg of protein) SSI lines (Fig. 9c). No significant difference in the GUS activity was observed between SC (average of 173 nmol/min/mg of protein) and MC (average of 216 nmol/min/mg of protein) lines (Fig. 9d).

In summary, expression of the genes expressed by constitutive promoters was measured by ELISA, protein fluorescence or enzyme activity. The expression variation in these assays was found to be much lower (2 – 4x) than that seen in transcript measurements (2-83X). The biallelic SC lines had almost two times higher GUS and GFP activity when compared to the monoallelic

SC lines. However, this trend was not seen in the NPT expression. Significant differences in the protein levels between SC and MC lines were also not observed. This data corroborated with qRT-PCR, where biallelic SSI had two-fold higher expression than monoallelic SSI in GUS and GFP, but not in NPT.

pporRFP expression analysis by confocal microscopy

Induced expression of *ppor*RFP controlled by HSP promoter was studied by confocal microscopy. First, a time course study was done for determining the optimal time for detecting *ppor*RFP fluorescence in heat-induced seedlings by confocal imaging. For this, roots of 7-10 days old T1 seedlings of SC lines #9 were subjected to heat-shock treatment and imaged at 24, 48 and 72 hours post-treatment. Red fluorescence was undetectable at room temperature or after 24 h of treatment. It was weakly detectable after 48 h, and optimally detected after 72 h of treatment (Appendix Fig. 1). Therefore, *ppor*RFP was studied after 72 h of heat treatment in all SSI lines. Later, all the 12 (9 SC and 3 MC) lines were screened for *ppor*RFP detection; however, RFP fluorescence could be captured in only four lines #9 ,10,11, and 12 in roots and/or shoots. Induced RFP fluorescence was observed in the roots of all four lines (Fig. 10 – 13). Line 9 showed a clear induced RFP expression in both the leaf blade and root, while line 10 showed room temperature expression in the leaf margins in addition to the induced expression in leaf blades and main root (Fig. 10 - 11). Line 11 did not show induced expression in the leaf blades but a clear induced RFP expression was captured in the roots (Fig. 12). Line 12 showed highest induced RFP expression of all in the leaf and roots (Fig. 13). In summary, confocal imaging confirmed induced expression of *pporRFP* in four SSI lines. Although some RFP expression at room temperature was found in one of the lines, induced levels of expression was observed in these lines in shoots and/or roots.

Discussion

The current study aimed to evaluate the expression of 5 genes stacked at a single locus. Our lab had previously reported stable expression of the reporter *GUS* gene from the site-specific integration (SSI) locus developed by Cre-*lox* recombination in the rice genome (Akbudak et al. 2011; Chawla et al. 2006). In the present study, expression of multigene stack consisting of three constitutive and two inducible genes was studied. These genes were integrated by Cre-*lox* recombination in the same rice genomic locus used in these previous studies, called *T5* in cv. Taipei-309 (Srivastava and Ow, 2002). Site-specific integration is a desirable approach for developing multigene stacks as it generates precise integration of the foreign DNA, which in turn allows stable expression of the integrated genes. This method is also a reliable approach for developing a higher number of single-copy lines as integration of only one copy of the foreign DNA is supported by Cre-*lox* recombination. While additional copies could get randomly integrated into the genome, they are likely to segregate in subsequent generations, yielding a clean SSI line.

Single-copy locus shows lower expression variability between transgenic lines and consistent expression in subsequent generations. Site-specific recombination mediated gene integration strategy has not been exploited for multigene stacking, so far, and only a limited information is available about the stability of multigene locus developed by integration of long T-DNA by *Agrobacterium* or co-bombardment of multiple vectors by particle bombardment (Anand et al. 2019; Collier et al. 2018). A general observation is that expression of multigene transgenic loci is highly variable due to copy# variation and unpredictable transgene integration site. A lower variation is observed in the single-copy lines; however, conventional methods of plant transformation generate only a few single-copy lines within an experiment. Site-specific

integration approach, on the other hand, delivers a higher percentage of single-copy lines. This study found that 48% of transgenic lines developed by Cre-*lox* mediated site-specific integration contained a single-copy of the DNA harboring 5 genes. Molecular analysis of 28 SSI lines showed that 96% lines (a total of 27) contained precise SSI structure at both integration junctions, 48% of which were SC, 30% MC with 1 - 3 additional copies, and 22% were truncated in the middle of the integrated fragment (Table 1). These observations were similar to the observations of Srivastava et al. (2004) and Chawla et al. (2006), who reported recovery of more than 50% precise SC lines and ~20% imprecise lines in their experiments of Cre-*lox* mediated site-specific integration of a transgene in rice.

The transgene locus stability was determined by the expression of all 5 genes at transcript levels for 9 SC and 3 MC lines; and at functional level (protein activity) for 4 genes (*NPT, GFP, GUS and pporRFP)* for 9 SC and 8 MC lines. We were interested to know if the copy number had an effect on the transcript levels on the five genes. In the constitutively expressing *GFP* and *GUS*, no significant difference in the expression levels were observed between SC and MC lines, while in *NPT*, the SC lines had 2.5x higher expression than MC lines (Fig. 4a-c). Ubiquitin promoter has been shown to be stronger in monocots than *35S* promoter (Christensen and Quail, 1997), thus differences in the promoters could have accounted for the difference in the *NPT* transcript levels in SC and MC lines, but not for the *GFP* and *GUS* transcripts. The induced expression of cold-inducible *AtRD29a:DREB1A* and heat-inducible *GmHSP17.5E: pporRFP* was observed in all of the 12 lines (Fig. 5a-b, 6a-b). Like *GFP* and *GUS*, no difference in the induced expression of either genes was observed in SC and MC lines (Fig. 5b, 6b). MC lines had only 1 - 3 additional copies of *NPT*, *GFP* and/or *pporRFP*, while no additional copies of *GUS* and *AtDREB1A*, were detected. Since the analysis was done in T1 plants, additional transgene

copies could segregate from the SSI locus, which may have a negative or positive effect on total gene expression in a plant (Chawla et al. 2006). This study also observed the variation in transcript abundance between T1 plants of the MC SSI line, which could be explained by SSI segregation from additional copies. For example, 9x, 6.6x, and 4.1x variation in the *NPT*, *GUS* and *GFP* transcript levels was observed in the two plants of the MC line #16 used in this study (Appendix Fig. 2).

We also sought to determine if the site-specific integration displayed the characteristic allelic gene dosage effect. Out of six biallelic SC lines, only three were analyzed due to plant sterility or low seed availability (Table 1). In agreement to Akbudak et al. (2010), Chawla et al. (2006) and Srivastava et al. (2004), ≥2 fold higher expressions of *GUS* and *GFP* were observed in biallelic lines as compared to the monoallelic lines (Fig. 4e-f). However, in *NPT*, the allelic effect was not statistically significant (Fig. 4d). The *pporRFP* and *AtDREB1A* also did not display any allelic effects (Fig. 5c-d, 6c-d), possibly due to their inducible nature of the expression.

At the protein levels, 2-fold expression variation was observed for NPTII among 17 SSI lines consisting of SC and MC lines (Fig. 7a). This consistency is likely due to the selectable nature of the *NPTII* gene. Antibiotic selection of SSI lines possibly ensures recovery of lines that express the gene at consistently high levels. In GFP and GUS activities, variation among the SSI lines was higher i.e. 4-fold. Importantly, allelic dosage effect was observed in both GFP and GUS activities among the SC lines, and generally, the transcript (RT-qPCR) levels corroborated with the estimated protein activities in the SC lines. For example, SC line #11 had the highest *GFP* and *GUS* transcript levels as well as the protein activities (Fig 4b-c, 8a, 9b), while line #21 had intermediate expression of *GFP* and *GUS* transcripts and the proteins (Fig 4b-c, 8a, 9b).

Similarly, MC line #19 had intermediate expression of *GFP* and *GUS* at both protein and transcript levels. This shows abundant transcription and translation in the SSI lines, and rules out the possibility of aberrant RNA formation that leads to post-transcriptional silencing. However, some anomalies in the expression levels of transcript and proteins were also observed, e.g., line #12 had low *GUS* transcript levels, but intermediate GUS activity (Fig. 4a, 9b).

The confocal imaging was performed to determine the functionality of *ppor*RFP in 12 SSI lines; *ppor*RFP is a 25.1 kD dsRed type tetramer protein derived from coral *Porites porites* (Alieva et al. 2008). The kinetic properties and/or crystal structure of *ppor*RFP has not been studied, and only a fraction of studies have used it for imaging, e.g., in switch-grass and tobacco, where ubiquitin promoter or *35S* promoter was used for expressing *ppor*RFP (King et al. 2014; Lin et al. 2017; Liu et al. 2014; Mann et al. 2012a; Mann et al. 2012b). In our study, *pporRFP* was expressed by the heat-inducible promoter. Further, dsRed protein requires an extended maturation time to attain a functional state (Jakobs et al. 2000; Sacceti et al. 2002), hence, we performed a time-course experiment at 24, 48, and 72 h post-heat shock on SC line #9 to determine, if *ppor*RFP also required a similar maturation time for a functional red chromophore. In agreement to these studies, we observed that the highest fluorescence was seen at 72 hours post heat-shock treatment (Appendix Fig. 1), implying that *ppor*RFP also requires extended maturation time. We could not test the fluorescence beyond 72 hours due to the deteriorating tissue quality and loss of *GFP* fluorescence. GFP imaging helped to track the tissue in confocal microscopy. Interestingly, the area with high GFP fluorescence had shown high *ppor*RFP fluorescence in roots, while in shoots only a smaller area was found to have a *ppor*RFP, despite having high GFP fluorescence in the four lines that were imaged for the *ppor*RFP fluorescence (Fig. 10-13). Recently, Jansing and Buyel (2019) showed that despite low mRNA levels in

tobacco plants, dsRed fluorescence gradually increased up to 5 days, indicating maturation time needed for dsRed subsequent to translation of its mRNA. Also they observed the highest protein levels in the younger leaves than older ones. We had observed that line #12 had highest fluorescence intensity among the four lines imaged. This line had a slow germination and was almost 4 days younger to that of T5 negative control and other T1 lines (# 9, 10 and 11), possibly supporting a higher *ppor*RFP fluorescence. Line #12 as mentioned above, had low transcript levels of *GUS* but high induced levels of *pporRFP* transcripts. Out of 12 lines screened, only four showed induced RFP fluorescence, although all 12 showed abundantly induced *pporRFP* transcripts upon heat-shock treatment. Lack of detectable induced RFP fluorescence in the remaining eight lines cannot be explained but could have bearing with the complex nature of *ppor*RFP maturation that was intractable in some lines due to the induced expression system used for *ppor*RFP expression.

AtDREB1A is a dehydration responsive element binding/C-repeat binding factor from *Arabidopsis*. It is a key trans-activation factor that has been shown to provide tolerance to abiotic environmental stresses such as cold, drought and salinity (Stockinger et al. 1997; Gilmour et al. 1998; Liu et al. 1998; Shinozaki and Yamaguchi-Shinozaki, 2000; Thomashow, 2001) in many plant species such as rice (Datta et al. 2012; Kasuga et al. 1999), tobacco (Kasuga et al. 2004), wheat, (Pellegrineschi et al.2004) and potato (Behnam et al., 2007). It contains *AP2/EREBP* DNA-binding domain that controls the expression of stress inducible genes including *rd29a*. In potato, the expression of *AtRD29a:AtDREB1A* had shown an increased resistance to the chilling stress in the T0 and T1 seedlings (Behnam et al. 2007), while in *Indica* rice, it has been shown to provide the drought stress resistance (Latha et al. 2019). When this multigene stacking study was conducted, only a few SSI lines had set T1 seeds, which were not sufficient for phenotyping

for the abiotic stress tolerance; therefore, the phenotyping will be carried out in the T2 generation of the selected lines.

As the complexity, has increased in the modern agriculture, need for the genetically enhanced crops with multiple traits have become more urgent. In future, multigene stacking will likely involve stacking of broad-spectrum traits like insect resistance along with the value added traits, such as nutritional enhancement and high yields. These goals could be realized by stacking multiple genes through biotechnology applications such as recombinase-mediated multigene integration. The resulting locus will be easy to breed into different varieties for cultivation in diverse ecosystems.

References

Ainley WM, Sastry-Dent L, Welter ME, Murray MG, Zeitler B, Amora R, Corbin DR, Miles RR, Arnold NL, Strange TL and Simpson MA (2013). Trait stacking via targeted genome editing. Plant Biotech J. 11(9):1126-34.

Akbudak MA, More AB, Nandy S and Srivastava V (2010). Dosage-dependent gene expression from direct repeat locus in rice developed by site-specific gene integration. Mol Biotech. 45(1):15-23.

Akbudak MA and Srivastava V (2017). Effect of gene order in DNA constructs on gene expression upon integration into plant genome. 3 Biotech. 7(2):94.

Alieva NO, Konzen KA, Field SF, Meleshkevitch EA, Hunt ME, Beltran-Ramirez V, Miller DJ, Wiedenmann J, Salih A and Matz MV (2008). Diversity and evolution of coral fluorescent proteins. PLoS One. 3(7):e2680.

Behnam B, Kikuchi A, Celebi-Toprak F, Kasuga M, Yamaguchi-Shinozaki K and Watanabe KN (2007). *Arabidopsis* rd29A: DREB1A enhances freezing tolerance in transgenic potato. Plant Cell Rep. 26(8):1275-82.

Bustin S and Huggett J (2017). qPCR primer design revisited. Biomol Detect Quant. 14:19-28.

Cai CQ, Doyon Y, Ainley WM, Miller JC, DeKelver RC, Moehle EA, Rock JM, Lee YL, Garrison R, Schulenberg L and Blue R (2009). Targeted transgene integration in plant cells using designed zinc finger nucleases. Plant Mol Biol. 69(6):699-709.

Cantos C, Francisco P, Trijatmiko KR, Slamet-Loedin I and Chadha-Mohanty PK (2014). Identification of "safe harbor" loci in indica rice genome by harnessing the property of zinc-finger nucleases to induce DNA damage and repair. Front Plant Sci. 5:302.

Chawla R, Ariza-Nieto M, Wilson AJ, Moore S and Srivastava V (2006). Transgene expression produced by biolistic-mediated, site-specific gene integration is consistently inherited by the subsequent generations. Plant Biotech J. 4(2):209-18.

D'Halluin K, Vanderstraeten C, Stals E, Cornelissen M and Ruiter R (2008). Homologous recombination: a basis for targeted genome optimization in crop species such as maize. Plant Biotech J. 6(1):93-102.

Datta K, Baisakh N, Ganguly M, Krishnan S, Yamaguchi Shinozaki K, and Datta SK (2012). Overexpression of Arabidopsis and rice stress genes' inducible transcription factor confers drought and salinity tolerance to rice. Plant Biotech J. 10(5):579-86.

Day CD, Lee E, Kobayashi J, Holappa LD, Albert H and Ow DW (2000). Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. Genes Dev. 14(22):2869-80.

Dwight Z, Palais R and Wittwer CT (2011). uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. Bioinformatics. 27(7):1019-20.

Gidoni D, Srivastava V and Carmi N (2008). Site-specific excisional recombination strategies for elimination of undesirable transgenes from crop plants. In Vitro Cell Dev-Pl. 44(6):457-67.

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM and Thomashow MF (1998). Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. Plant J. 16:433–442.

Jakobs S, Subramaniam V, Schönle A, Jovin TM and Hell SW (2000). EGFP and DsRed expressing cultures of Escherichia coli imaged by confocal, two-photon and fluorescence lifetime microscopy. FEBS Letters. 479(3):131-5.

Jefferson RA (1987). Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep .5(4):387-405.

Jansing J and Buyel JF (2019). The correlation between DsRed mRNA levels and transient DsRed protein expression in plants depends on leaf age and the 5′ untranslated region. Biotechnol J. 14(3):1800075.

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K and Shinozaki K (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nature Biotech. 17:287–291

Kasuga M, Miura S, Shinozaki K and Yamaguchi-Shinozaki K (2004). A combination of the *Arabidopsis* DREB1A gene and stress-inducible rd29A promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer. Plant Cell Physiol. 45(3):346–350.

King ZR, Bray AL, LaFayette PR and Parrott WA (2014). Biolistic transformation of elite genotypes of switchgrass (Panicum virgatum L.). Plant Cell Rep. 33(2):313-22.

Kumar S, Barone P and Smith M (2016). Gene targeting and transgene stacking using intra genomic homologous recombination in plants. Plant Methods. 12(1):11.

Lin CY, Donohoe BS, Ahuja N, Garrity DM, Qu R, Tucker MP, Himmel ME and Wei H (2017). Evaluation of parameters affecting switchgrass tissue culture: toward a consolidated procedure for Agrobacterium-mediated transformation of switchgrass (Panicum virgatum). Plant Methods. 13(1):113.

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K and Shinozaki K (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain, separate two cellular signal transduction pathways in drought- and low temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell 10:1391–1406.

Liu W, Rudis MR, Peng Y, Mazarei M, Millwood RJ, Yang JP, Xu W, Chesnut JD and Stewart Jr CN (2014). Synthetic TAL effectors for targeted enhancement of transgene expression in plants. Plant Biotech J. 12(4):436-46.

Livak KJ and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 25(4):402-8.

Mann DG, LaFayette PR, Abercrombie LL, King ZR, Mazarei M, Halter MC, Poovaiah CR, Baxter H, Shen H, Dixon RA and Parrott WA (2012a). Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (Panicum virgatum L.) and other monocot species. Plant Biotech J. 10(2):226-36.

Mann DG, Abercrombie LL, Rudis MR, Millwood RJ, Dunlap JR and Stewart CN (2012b). Very bright orange fluorescent plants: endoplasmic reticulum targeting of orange fluorescent proteins as visual reporters in transgenic plants. BMC Biotech. 12(1):17.

Moehle EA, Rock JM, Lee YL, Jouvenot Y, DeKelver RC, Gregory PD, Urnov FD and Holmes MC (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. Proc. Natl. Acad. Sci. USA. 104(9):3055-60.

Nandy S and Srivastava V (2011). Site-specific gene integration in rice genome mediated by the FLP–FRT recombination system. Plant Biotech J. 9(6):713-21.

Nandy S, Zhao S, Pathak BP, Manoharan M and Srivastava V (2015). Gene stacking in plant cell using recombinases for gene integration and nucleases for marker gene deletion. BMC Biotech J. 15(1):93.

Nishimura A, Aichi I and Matsuoka M (2006). A protocol for Agrobacterium-mediated transformation in rice. Nat Prot. 1(6):2796.

Pabuayon IM, Yamamoto N, Trinidad JL, Longkumer T, Raorane ML and Kohli A (2016). Reference genes for accurate gene expression analyses across different tissues, developmental stages and genotypes in rice for drought tolerance. Rice. 9(1):32.

Papapetrou EP, Lee G, Malani N, Setty M, Riviere I, Tirunagari LM, Kadota K, Roth SL, Giardina P, Viale A and Leslie C (2011). Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. Nat Biotech. 29(1):73.

Pellegrineschi A, Reynolds M, Pacheco M, Brito RM, Almeraya R, Yamaguchi-Shinozaki K and Hoisington D (2004) Stress-induced expression in wheat of the *Arabidopsis* thaliana DREB1A gene delays water stress symptoms under greenhouse conditions. Genome 47:493–500

Petolino JF and Kumar S (2016). Transgenic trait deployment using designed nucleases. Plant Biotech J .14(2):503-9.

Ruijter JM, Pfaffl MW, Zhao S, Spiess AN, Boggy G, Blom J, Rutledge RG, Sisti D, Lievens A, De Preter K and Derveaux S (2013). Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications. Methods. 59(1):32-46.

Sacchetti A, Subramaniam V, Jovin TM and Alberti S (2002). Oligomerization of DsRed is required for the generation of a functional red fluorescent chromophore. FEBS Letters. 525(1-3):13-9.

Schnuetgen F, Stewart AF, von Melchner H and Anastassiadis K (2006). Engineering embryonic stem cells with recombinase systems. Meth Enzym. 420:100-36.

Shinozaki K and Yamaguchi-Shinozaki K (2000) Molecular response to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Curr Opin Plant Biol. 3:217–223.

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X and Choi VM (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature. 459(7245):437.

Sreedharan SP, Kumar A and Giridhar P. (2018) Primer design and amplification efficiencies are crucial for reliability of quantitative PCR studies of caffeine biosynthetic N-methyltransferases in coffee. 3 Biotech. 8(11):467.

Srivastava V, Ariza-Nieto M and Wilson AJ (2004). Cre-mediated site-specific gene integration for consistent transgene expression in rice. Plant Biotech J. 2(2):169-79.

Srivastava V and Ow DW (2001). Single-copy primary transformants of maize obtained through the co-introduction of a recombinase-expressing construct. Plant Mol Biol. 46(5):561-6.

Srivastava V and Thomson J (2016). Gene stacking by recombinases. Plant Biotech J. 14(2):471-82.

Stockinger EJ, Gilmour SJ and Thomashow MF (1997). *Arabidopsis* thaliana CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc. Natl. Acad. Sci. USA. 94(3):1035-40.

Thomashow MF (2001). So what's new in the field of plant cold acclimation? Lots! Plant Physiol 125:89–93

Wang Y, Yau YY, Perkins-Balding D andThomson JG (2011). Recombinase technology: applications and possibilities. Plant Cell Rep. 30(3):267-85.

Wright DA, Townsend JA, Winfrey Jr RJ, Irwin PA, Rajagopal J, Lonosky PM, Hall BD, Jondle MD and Voytas DF (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J. 44(4):693-705.

Xie K, Minkenberg B and Yang Y (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. USA. 112(11):3570-5.

Tables and Figures

Table 1: Characterization of SSI lines

T0 #	PCR			Copy # by Southern[1]				T1 segregation data[2]		Conclusion[3]
	Junction 1	Junction 2	Target site	GFP	RFP	GUS	DREB	GFP+	GFP -	
1	✓	×	✓	2	2	-	-	3	2	Imprecise
2[#]	✓	✓	✓	2	2	-	-	-	-	Truncated
3[#]	✓	✓	✓	2	2	-	-	-	-	Truncated
4	✓	✓	✓	1	1	-	-	-	-	SC/Monoallelic
5[*]	✓	✓	✓	2	3	-	-	8	2	MC/Monoallelic
6[*]	✓	✓	✓	2	3	-	-	-	-	Clonal to line 5
8	✓	✓	✓	2	2	0	0	-	-	Truncated
9	✓	✓	✓	1	1	1	1	36	11	SC/Monoallelic
10	✓	✓	✓	1	1	1	1	34	3	SC/Monoallelic
11	✓	✓	×	1	1	1	1	42	0	SC/Biallelic
12	✓	✓	✓	1	1	1	1	23	8	SC/Monoallelic
13	✓	✓	✓	2	2	0	0	-	-	Truncated
14	✓	✓	✓	2	2	1	1	16	6	MC/Monoallelic
15	✓	✓	✓	2	2	0	0	4	2	Truncated
16[$]	✓	✓	✓	2	4	1	1	23	3	MC/Monoallelic
17[$]	✓	✓	✓	2	4	1	1	2	1	Clonal to 16
18	✓	✓	✓	2	2	1	1	-	-	MC/Monoallelic
19	✓	✓	✓	2	3	1	1	13	2	MC/Monoallelic
20	✓	✓	×	1	1	1	1	23	0	SC/Biallelic
21	✓	✓	✓	1	1	1	1	23	5	SC/Monoallelic
22	✓	✓	×	1	1	1	1	-	-	SC/Biallelic
26	✓	✓	×	1	1	1	1	-	-	SC/Biallelic
27	✓	✓	✓	1	1	1	1	14	7	SC/Monoallelic
28	✓	✓	✓	2	2	0	0	-	-	Truncated
29	✓	✓	×	1	1	1	1	19	0	SC/Biallelic
30	✓	✓	✓	2	2	1	1	4	0	MC/Monoallelic
31	✓	✓	×	1	1	1	1	-	-	SC/Biallelic
32	✓	✓	✓	1	1	1	1	19	6	SC/Monoallelic

[#,*, $:] Considered clonal lines based on southern patterns.

[1]: lines 1-6 not studied in southern blots for *DREB* and *GUS* genes.

[2]: T1 data not studied due to plant sterility or low amount of seeds.

[3]: SC: Single copy; MC: Multicopy.

Table 2: Quantitative RT-PCR primer efficiency

Gene	Gene - GC content (%)	Primer Name	Sample type used for testing efficiency	Equation	R^2	Efficiency (%)	Used for expression analysis
AtDREB1A	50	qdrebf1-r1	Genomic DNA	-3.7818x + 24.439	0.987	84	No
		qdrebf2-r2	Genomic DNA	-3.4996x + 24.964	0.998	93	Yes
		qdrebf3-r3	Genomic DNA	-3.7173x + 25.155	0.987	86	No
pporRFP	47	qrfpf1-r1	Genomic DNA	-3.1186x + 25.724	1.000	109	No
		qrfpf2-r2	Genomic DNA	-3.1898x + 24.291	0.994	105	Yes
GFP	61	qGFPF2-R2	cDNA	-3.3908x + 18.492	0.955	97	Yes
		qGFPF3-R3	cDNA	-5.6011x + 17.115	0.965	58	No
NPT	60	qNPTF1-R1	cDNA	-3.19x + 21.472	0.975	108	No
		qNPTF2-R2	cDNA	-3.375x + 20.265	0.961	98	Yes
GUS	52	qGUSF1-gusR2	cDNA	-3.1925x + 21.025	0.997	106	Yes
		qGUSF2-qgusR2	cDNA	-2.9475x + 21.451	0.999	118	No

Table 3: Primers used in this study

Primer Name	Sequence (*5'-3'*)	Application
Ubi1960	GCTCACCCTGTTGTTTGGTG	Genotyping of pNS64 T0 and T1 lines
KanR	CTCGATGCGATGTTTCGCTT	
BamH1pporRFP-F	CGGGATCCATGGCTCTTTCAAAGC	
Cre2333	ATTGCTGTCACTTGGTCGTG	
Ubi	TCTACTTCTGTTCATGTTTGT	
RevcreATG	ACGGTCAGTAAATTGGACAT	
GusF982	ACCTCGCATTACCCTTACGC	
qDREB-F2	GGA GAC GTT GGT GGA GGC TA	*AtDREB1A* expression analysis
qDREB-R2	CGG ACG GAA GCG GCA AAA GCA	
qRFP-F2	GGCTCGATGGCGACTCTTTCAT	*pporRFP* expression analysis
qRFP-R2	CACCACACTCATACAGTCTCT	
qGFPF2-	GACCACTACCAGCAGAACAC	*GFP* expression analysis
qGFP-R2	CCATGTGATCGCGCTTCT	
qNPT-F2	CGTTGGCTACCCGTGATATT	*NPT* expression analysis
qNPT-R2	CTCGTCAAGAAGGCGATAGAAG	
qGUS-F1	CGACCTCGCAAGGCATATT	*GUS* expression analysis
GUS-R2	TCACCGAAGTTCATGCCAGT	
Q7Ubiq1445F	TGGTCAGTAATCAGCCAGTTTG	Reference genes for expression analysis
Q7Ubiq1520R	CAAATACTTGACGAACAGAGGC	
Ubiquitin-F	CGCAAGTACAACCAGGACAA	
Ubiquitin-R	GCTGTGACCACACTTCTTCTT	

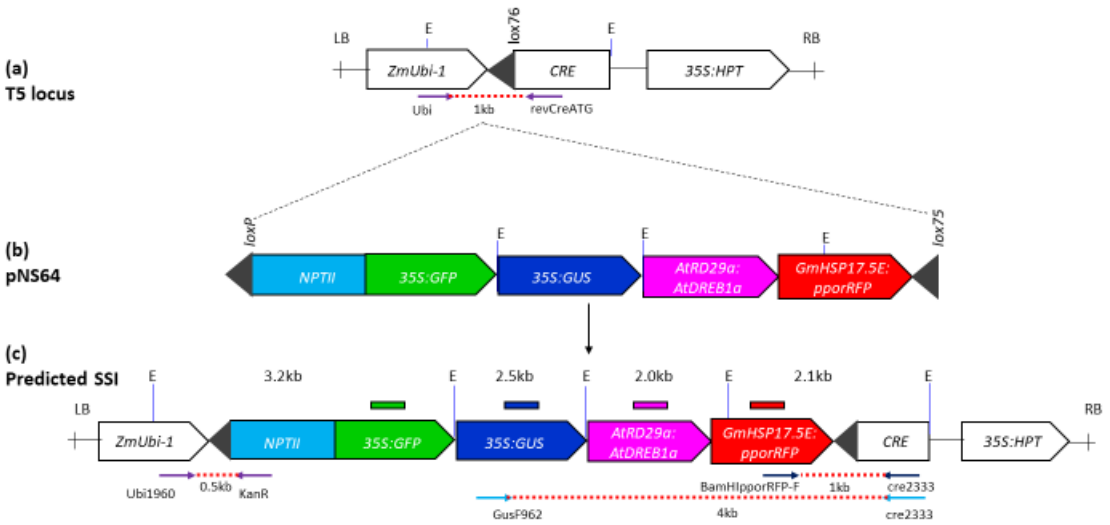


Figure 1: Molecular approach for site-specific integration (SSI) of a multigene fragment. **[a]** *T5* locus in cv. Taipei-309 containing a single-copy of T-DNA encoding Cre activity and the target *lox76* site (black triangle). The 35S:*HPT* gene serves as the selection marker. **[b]** Donor vector, pNS64, in pBluescript SK backbone (not shown) containing promoterless *NPTII* gene and four expression units (*GFP*, *GUS*, *AtDREB1A*, and *pporRFP*) between *loxP* and *lox75*. The *loxP x lox75* recombination will circularize the molecule, which will integrate into *T5* locus to generate the site-specific integration. The NPTII gene captures the maize ubiquitin-1 promoter (*ZmUbi-1*) at *T5* locus to make the event selectable and expresses four genes, two constitutive (*GFP* and *GUS*) and two inducible (*AtDREB1A* and *pporRFP*) genes. **[c]** Structure of the predicted site-specific integration locus that expresses a stack of four genes (*NPTII, GFP*, *GUS, AtDREB1A,* and *pporRFP*). *ZmUbi-1*: maize Ubiquitin-1 promoter; *HPT*: hygromycin phosphotransferase gene; *35S*: cauliflower mosaic virus *35S* promoter, *NPTII*: neomycin phosphotransferase II; *GFP*: green fluorescent protein; *GUS*: β-Glucuronidase; *AtRD29a*: *Arabidopsis thaliana RD29a* promoter; *AtDREB1A*: *Arabidopsis thaliana* dehydration responsive element 1A; *GmHSP17.5E*: soybean heat-shock 17.5E promoter; *pporRFP*: sea coral *Porites porites* red fluorescent protein; E: *Eco*RI; LB and RB: T-DNA left and right borders. Each gene carries a nopaline synthase *3'* transcription terminator (not shown). Fragment sizes in kb are indicated. The small rectangles are the probes used for southern hybridization and the primer names and positions (arrows) are shown along with their expected sizes.

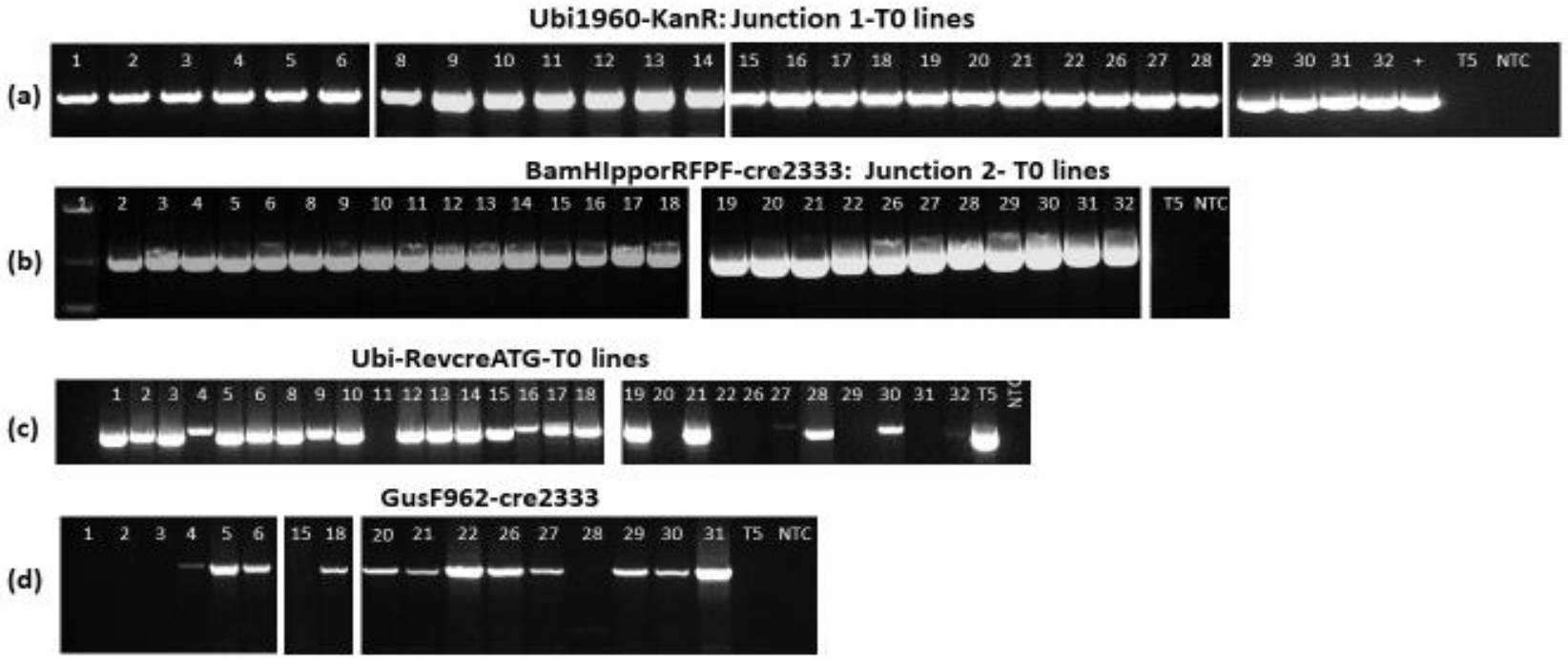
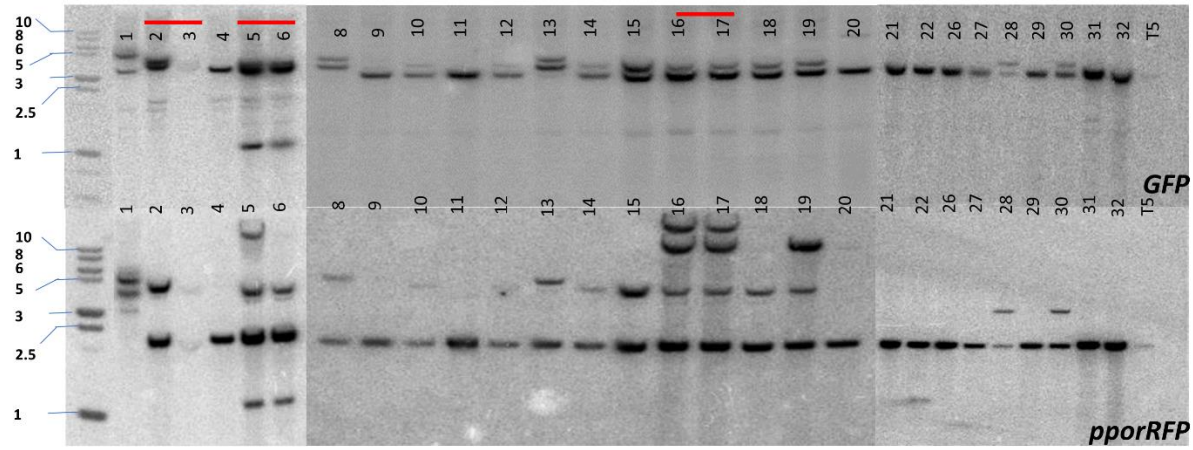
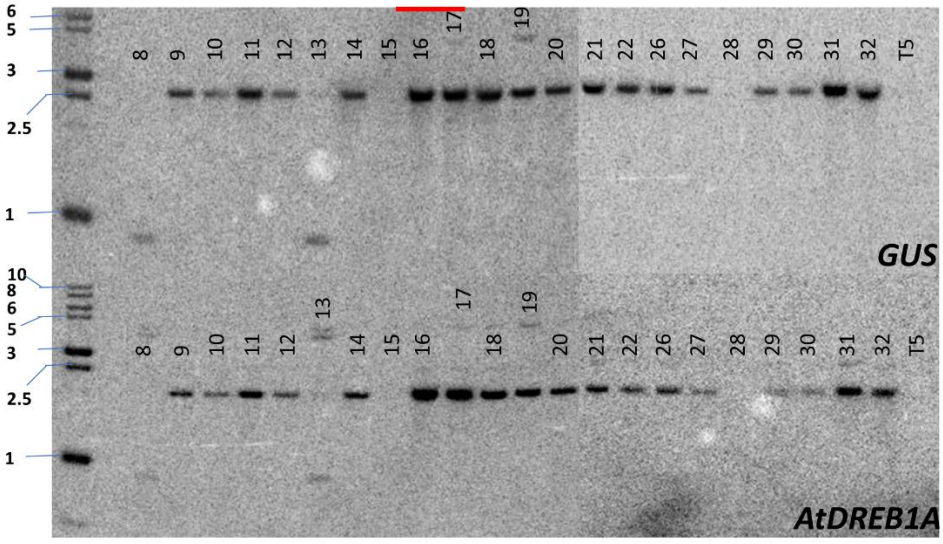


Figure 2: Verification of site-specification integration (SSI) by determining predicted junctions through PCR in the primary transgenic (T0) SSI lines. **[a]** PCR for the presence of the first SSI junction using Ubi1960 and KanR primers. **[b]** PCR for the second junction using BamHIpporRFPF and cre2333 primers. **[c]** PCR for the target site using Ubi and revcreATG primers. This PCR distinguishes monoallelic and biallelic integrations. **[d]** PCR for detecting full-length integration using GusF962 and cre2333 primers. Primer positions in the SSI and target sites are shown in Fig. 1. T5: Negative plant control; NTC: No template control.



(a)



(b)

Figure 3: Southern hybridization of *Eco*R1-digested genomic DNA of the primary transgenic (T0) site-specific integration (SSI) lines using *GFP* and *pporRFP* probes **[a],** and *GUS* and *AtDREB1A* probes **[b].** DNA ladder and sizes are indicated in kb.

.

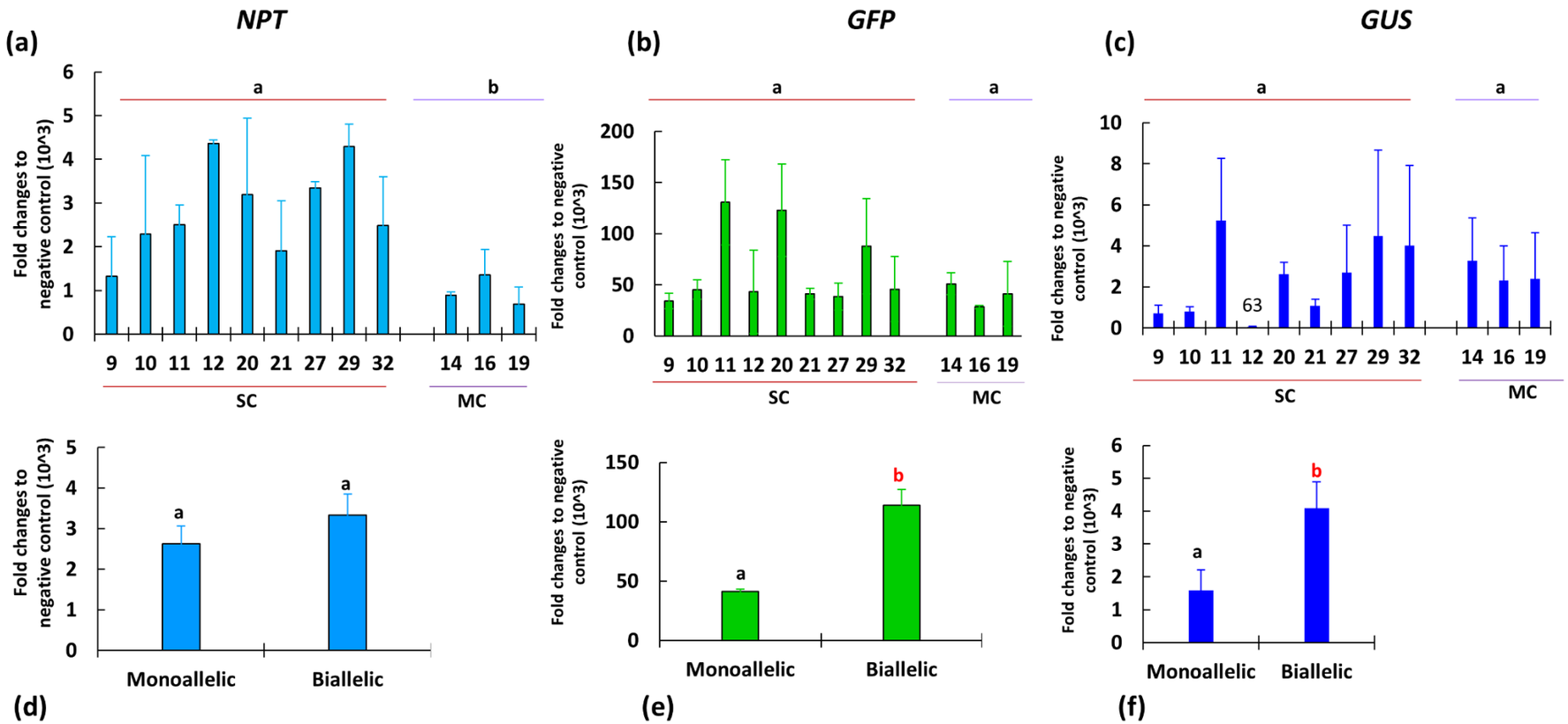


Figure 4: Expression analysis of constitutively expressed genes by real time quantitative PCR (RT-qPCR). **[a-c]** Relative expression of *NPT*, *GFP*, and *GUS* genes in the T1 plants of 9 single-copy (SC) and 3 multicopy (MC) lines**. [d-e]** Average of expression levels of *NPT*, *GFP*, and *GUS* in 6 monoallelic and 3 biallelic SC lines. Statistical differences, shown by the alphabets, were determined by student *t*-test at p=0.05. Error bars are the standard errors of 2 – 6 biological replicates.

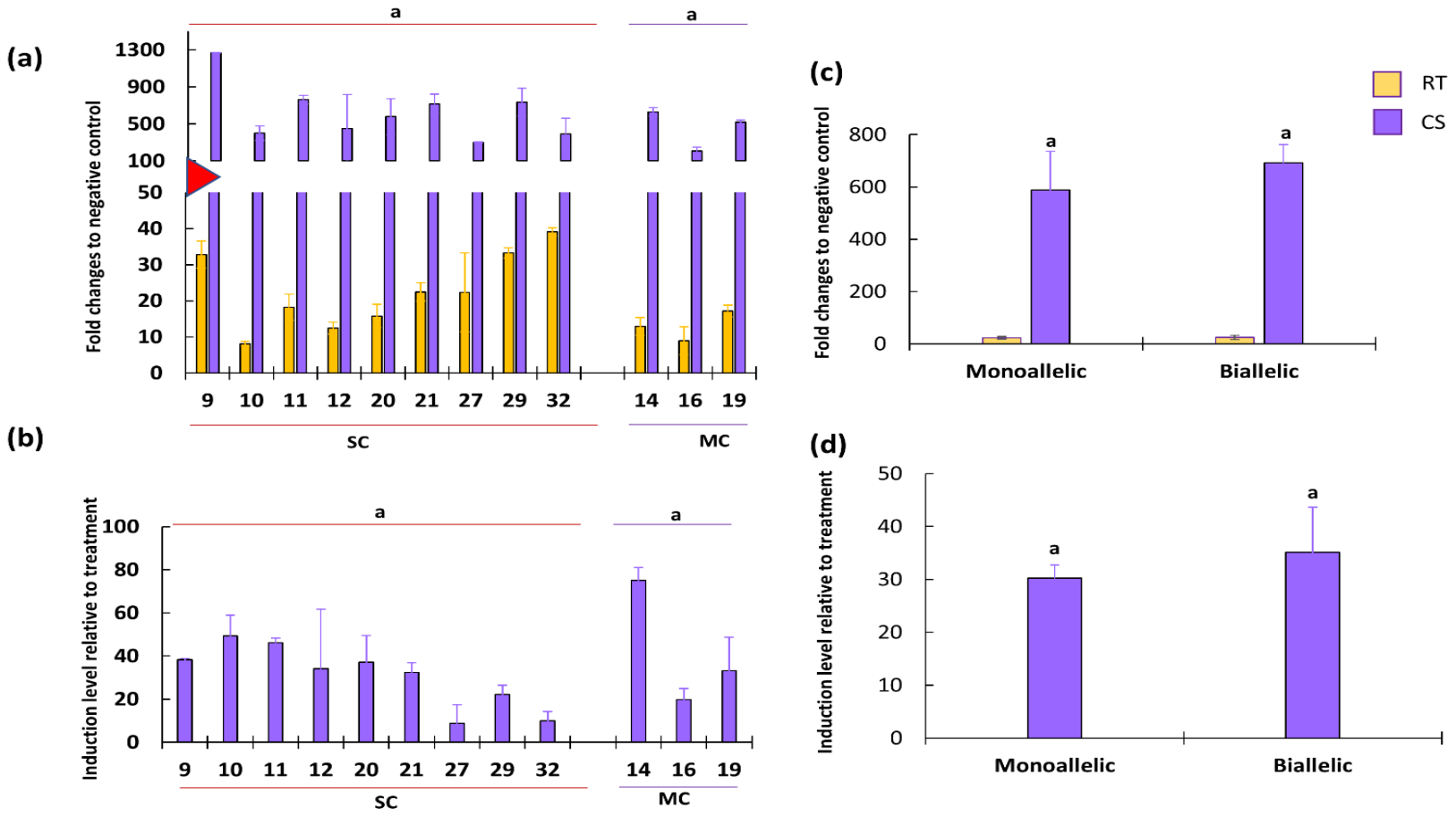


Figure 5: Expression analysis of the cold-inducible *AtDREB1A* gene by real time quantitative PCR (RT-qPCR). **[a]** *AtDREB1A* expression at room temperature or upon cold-induction (4ºC for 20 hours) relative to the T5 negative control in the T1 progeny. **[b]** Cold-induction levels of *AtDREB1A* in each line. The values in (a-b) are the average of 2 biological replicates. **[c-d]** Average of the expression levels in 6 monoallelic and 3 biallelic SC lines. Statistical differences, shown by the alphabets, were determined by student *t*-test at p=0.05. Error bars are the standard errors of 2 – 6 biological replicates. SC: Single copy, MC: Multicopy.

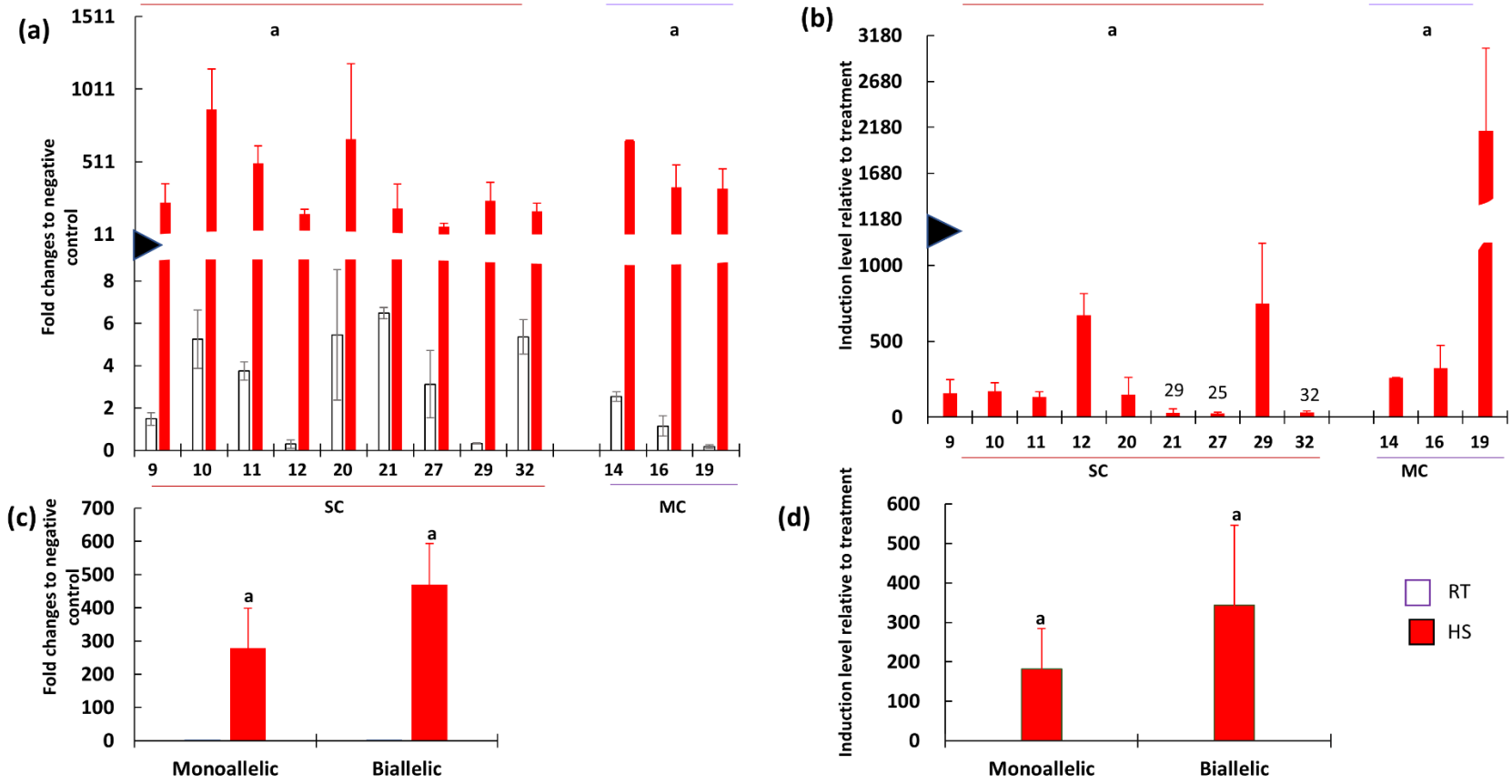


Figure 6: Expression analysis of heat-inducible *pporRFP* gene by real time quantitative PCR (RT-qPCR) in the T1 progeny plants of the site-specific integration (SSI) lines. **[a]** *pporRFP* expression at room temperature (white bars) or upon heat-induction (42ºC for 3 hours; red bars) relative to the T5 negative control. **[b]** Heat-induced levels of *pporRFP* in each line. The values in (a-b) are the average of 2 biological replicates. **[c-d]** Average of the expression levels in 6 monoallelic and 3 biallelic SC lines. Statistical differences, shown by the alphabets, were determined by student *t*-test at p=0.05. Error bars are the standard errors of 2 –6 biological replicates. SC: Single copy, MC: Multicopy.

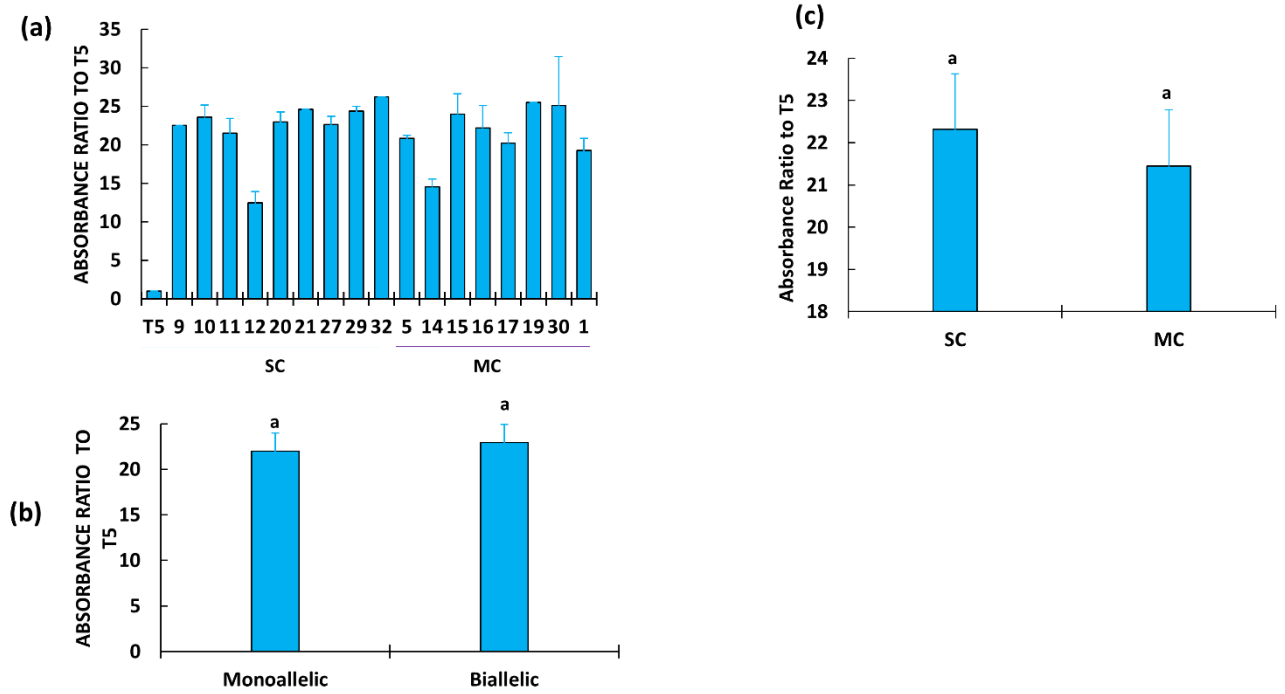


Figure 7: NPTII ELISA in T1 lines. **[a]** Absorbance ratio of the site-specific integration (SSI) lines relative to the T5 negative control. Each line represents the average of two-three biological replicates. **[b]** Average absorbance ratio of 6 monoallelic and 3 biallelic integrants of SC lines. **[c]** Average absorbance ratio of SC and MC lines. Statistical differences, shown by the alphabets, were determined by student *t*-test at p=0.05. Error bars are the standard errors. SC: Single copy, MC: Multicopy.

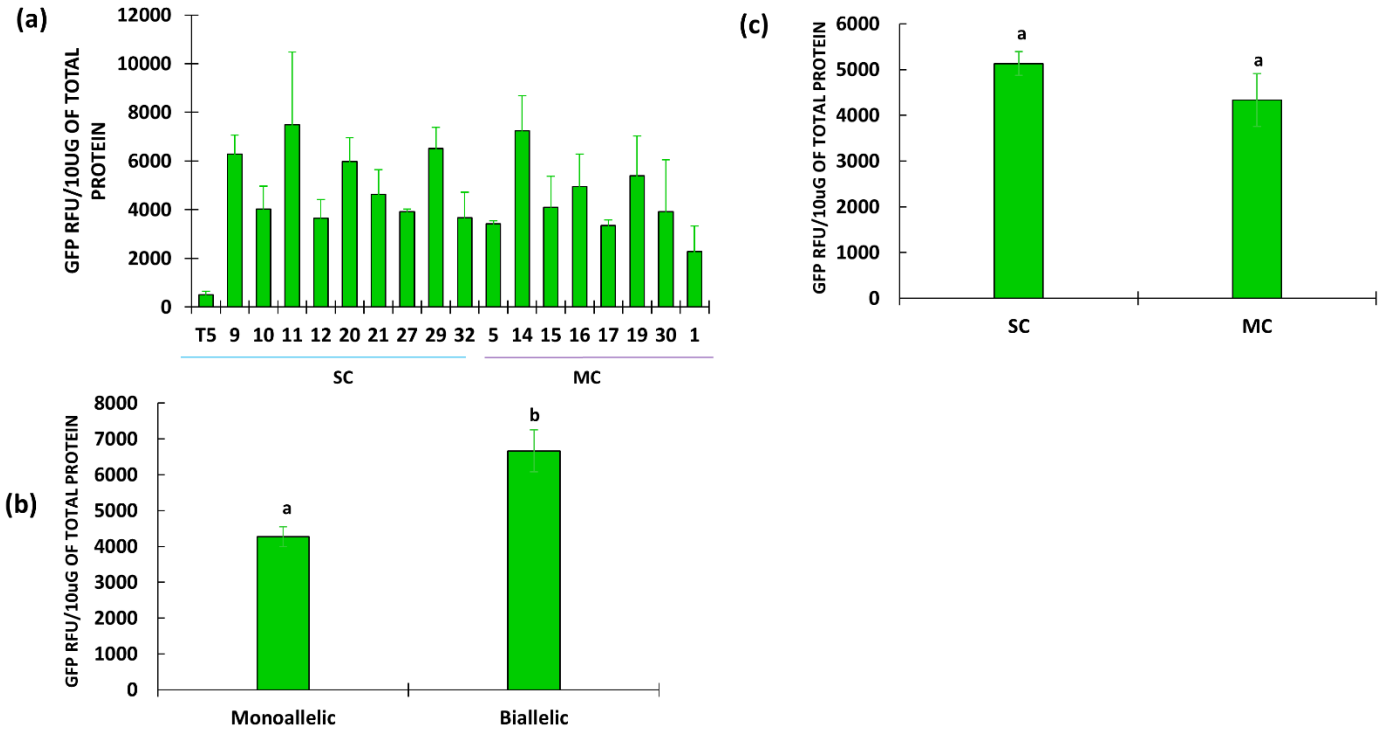


Figure 8: GFP quantification in the T1 plants of site-specific integration (SSI) lines by fluorometric assay (Relative Fluorescence Units, RFU). **[a]** RFU of SSI lines. Each line represents the average RFU of three biological replicates. **[b]** Average RFU of 6 monoallelic and 3 biallelic SC lines. **[c]** Average RFU of 9 SC and 8 MC lines. Statistical differences, shown by the alphabets, were Determined by student *t*-test at p=0.05. Error bars are the standard errors. SC: Single copy, MC: Multicopy.

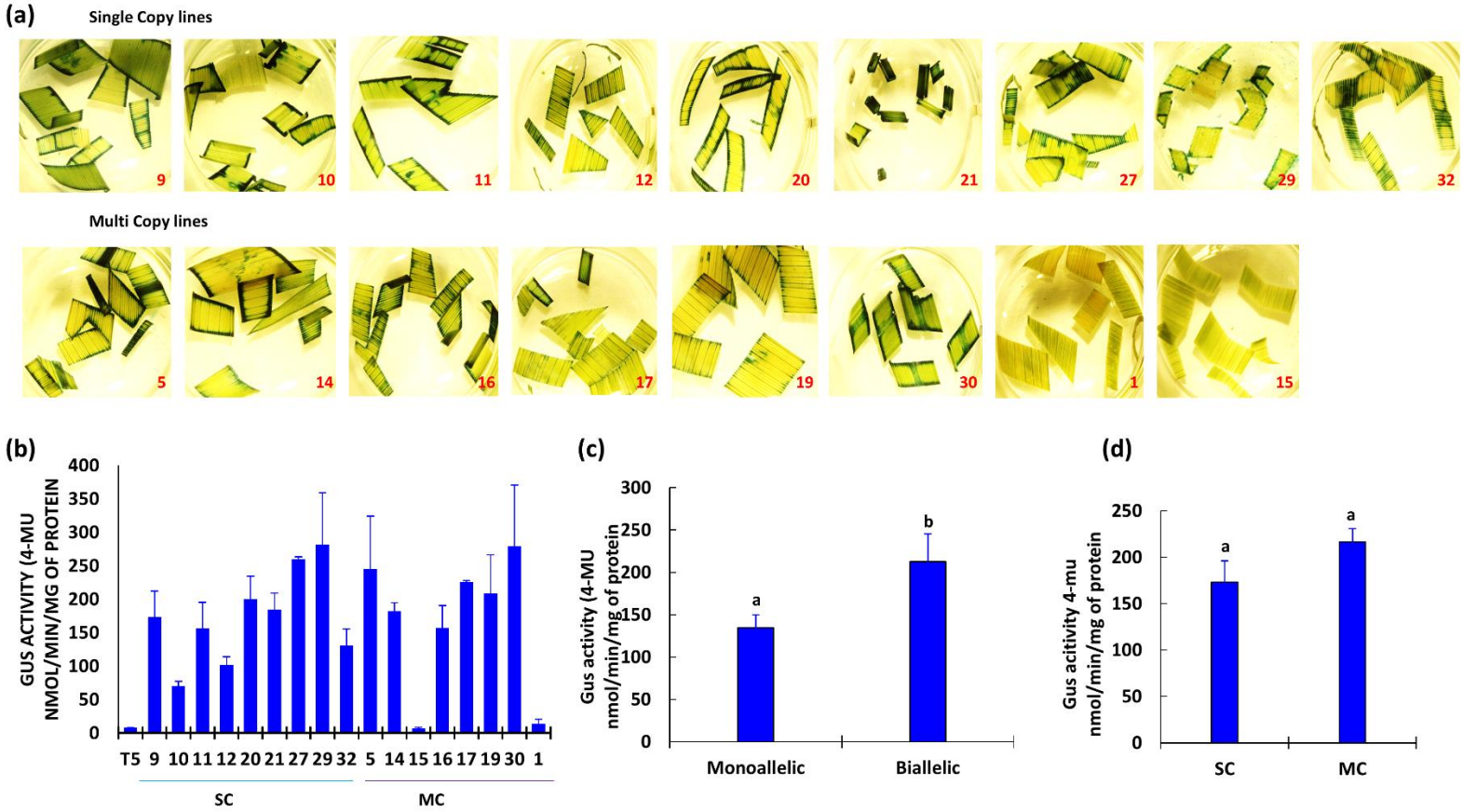


Figure 9: GUS activity in the T1 plants of the site-specific integration (SSI) lines. **[a]** Histochemical staining of the leaf cuttings of 9 SC and 8 MC lines. GUS activity is indicated by the dark blue staining. **[b]** Estimation of GUS activity in the T1 plants of SSI lines using fluorometric assay. Each line represents the average activity of three biological replicates. **[c]** Average GUS activity in the 6 monoallelic and 3 biallelic SC lines. **[d]** Average GUS activity of 9 SC and 6 MC lines. . Statistical differences, shown by the alphabets, were determined by student *t*-test at p=0.05. Error bars are the standard errors. SC: Single Copy, MC: Multicopy.

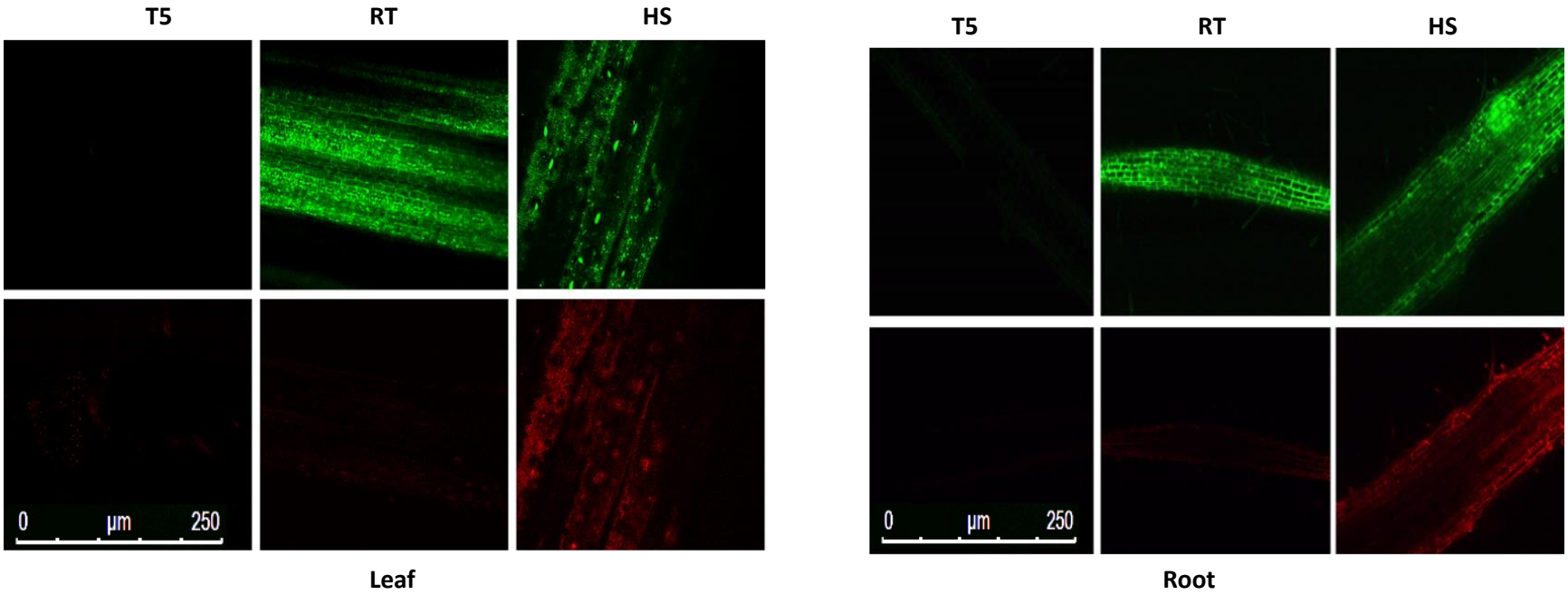


Figure 10: Confocal imaging of GFP (top) and *ppor*RFP (bottom) in roots and leaves of the T1 plant of SC line #9. All images were taken at 72 hours post heat-shock at 20x magnification. Image bar indicates the magnification, offset, and zoom used in all the images. T5: Negative control; RT: Room temperature; HS: Heat-shock.

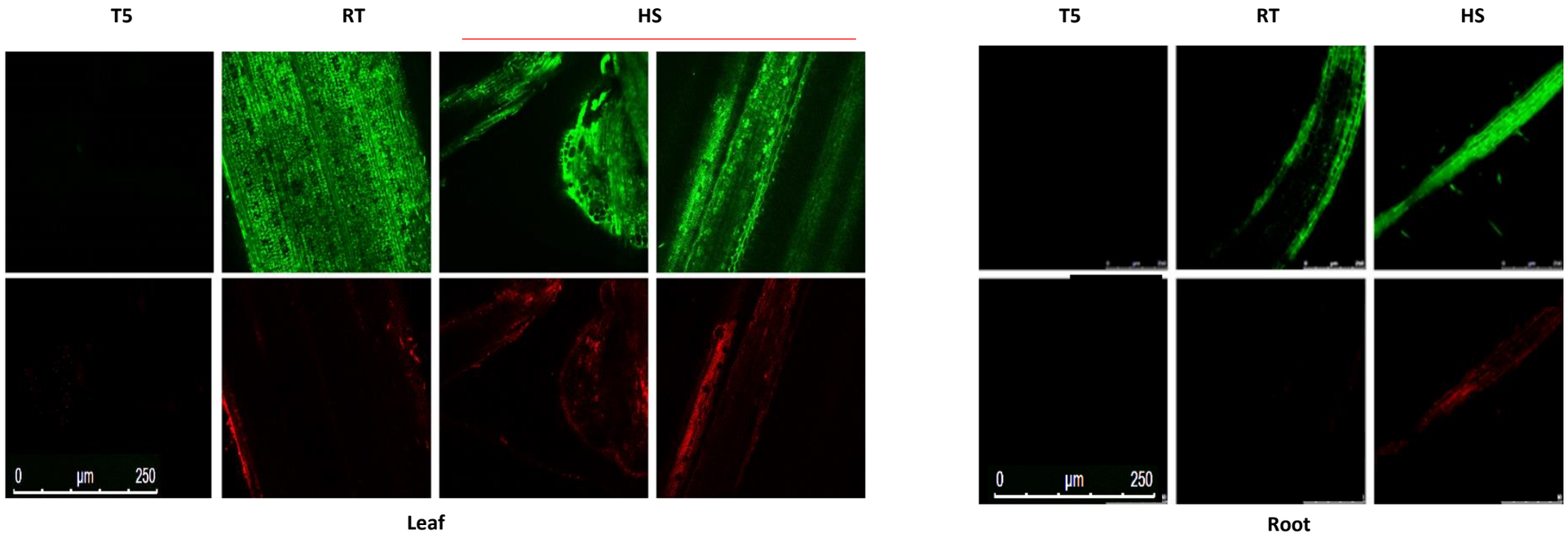


Figure 11: Confocal imaging of GFP (top) and *ppor*RFP (bottom) in roots and leaves of the T1 plant of SC line #10. All images were taken at 72 hours post heat shock at 20x magnification. Image bar indicates the magnification, offset, and zoom used in all the images. T5: Negative control; RT: Room temperature; HS: Heat-shock.

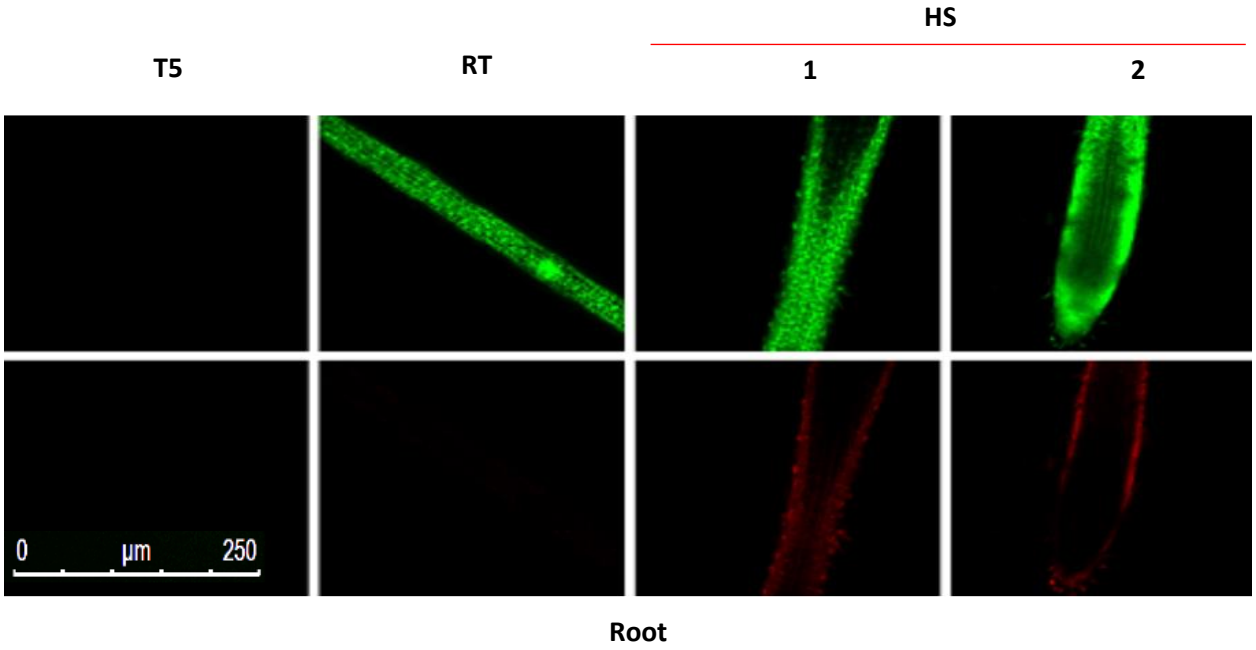


Figure 12: Confocal imaging of GFP (top) and *ppor*RFP (bottom) in the roots of the T1 plants of SC line #11. All images were taken at 72 hours post heat shock at 20x magnification. Image bar indicates the magnification, offset, and zoom used in all the images. T5: Negative control; RT: Room temperature; HS: Heat-shock; #1, #2 are the images from two different seedlings.

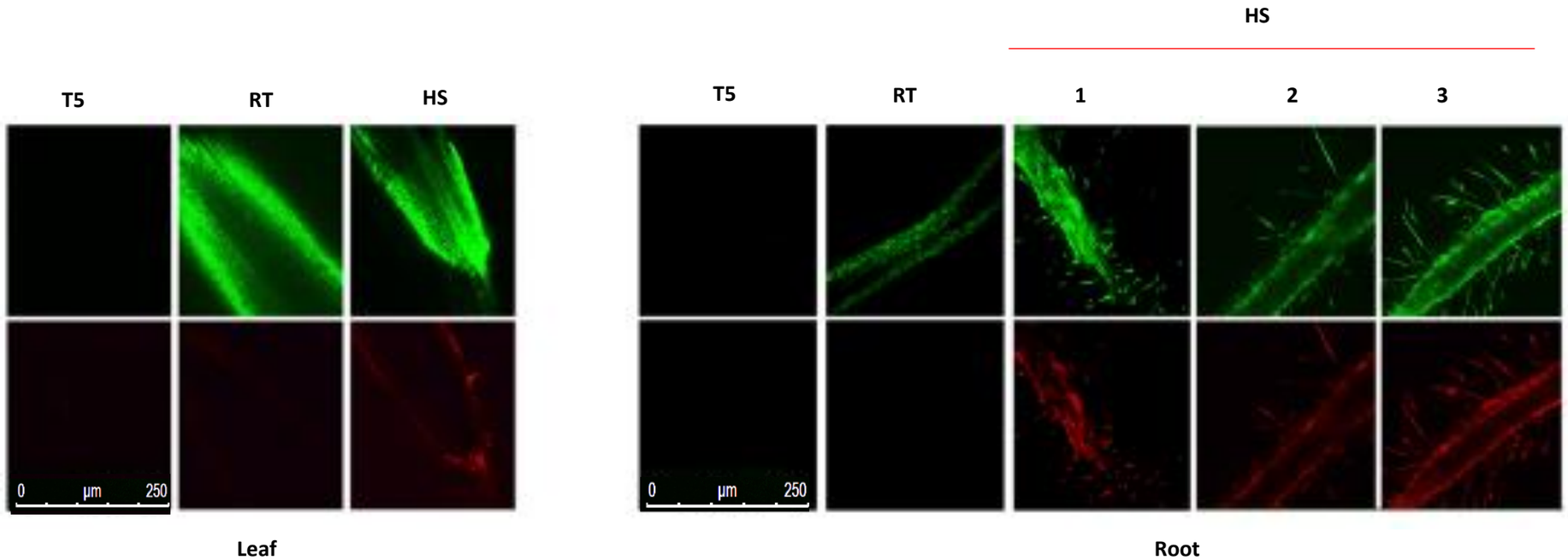
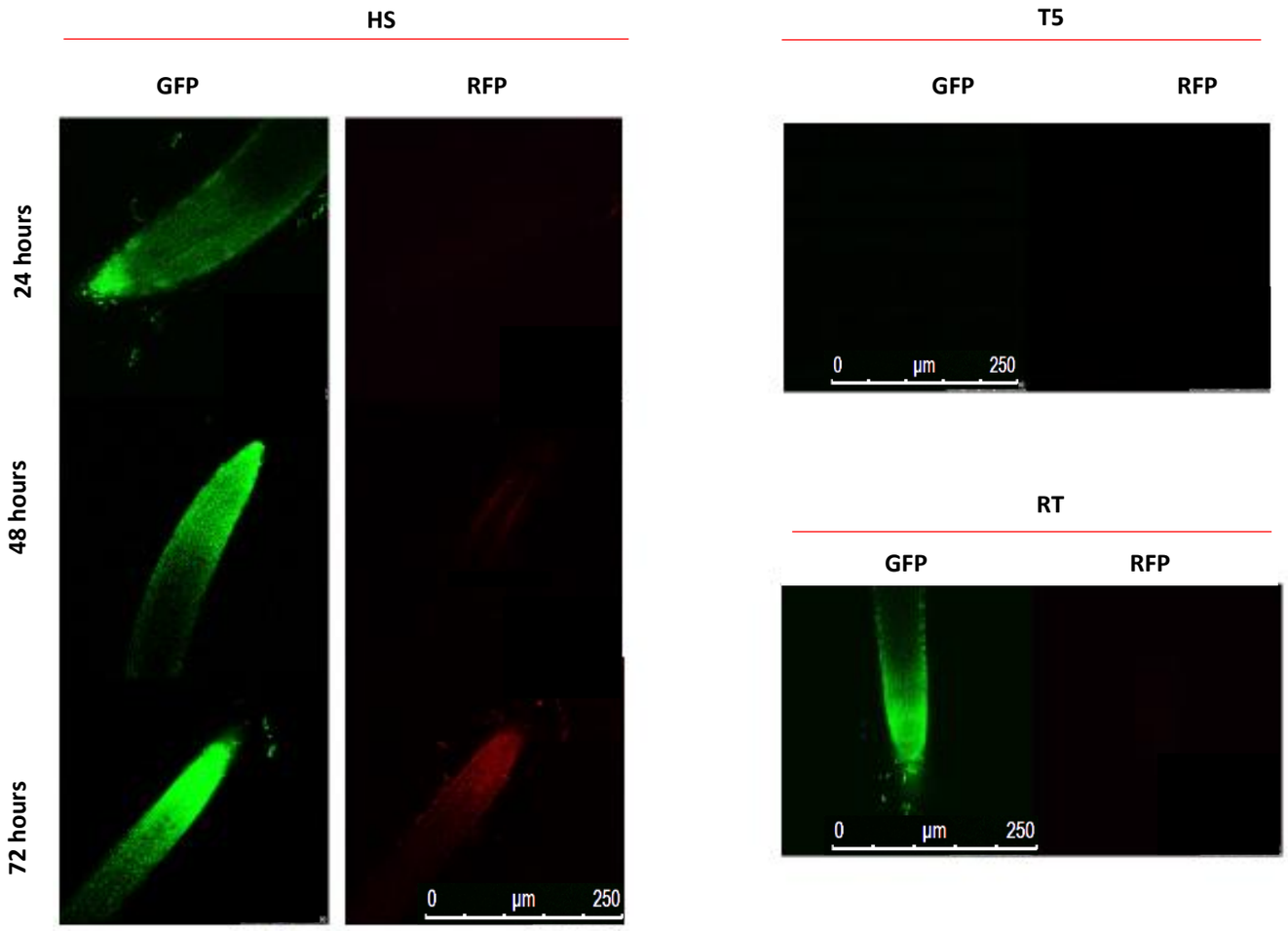
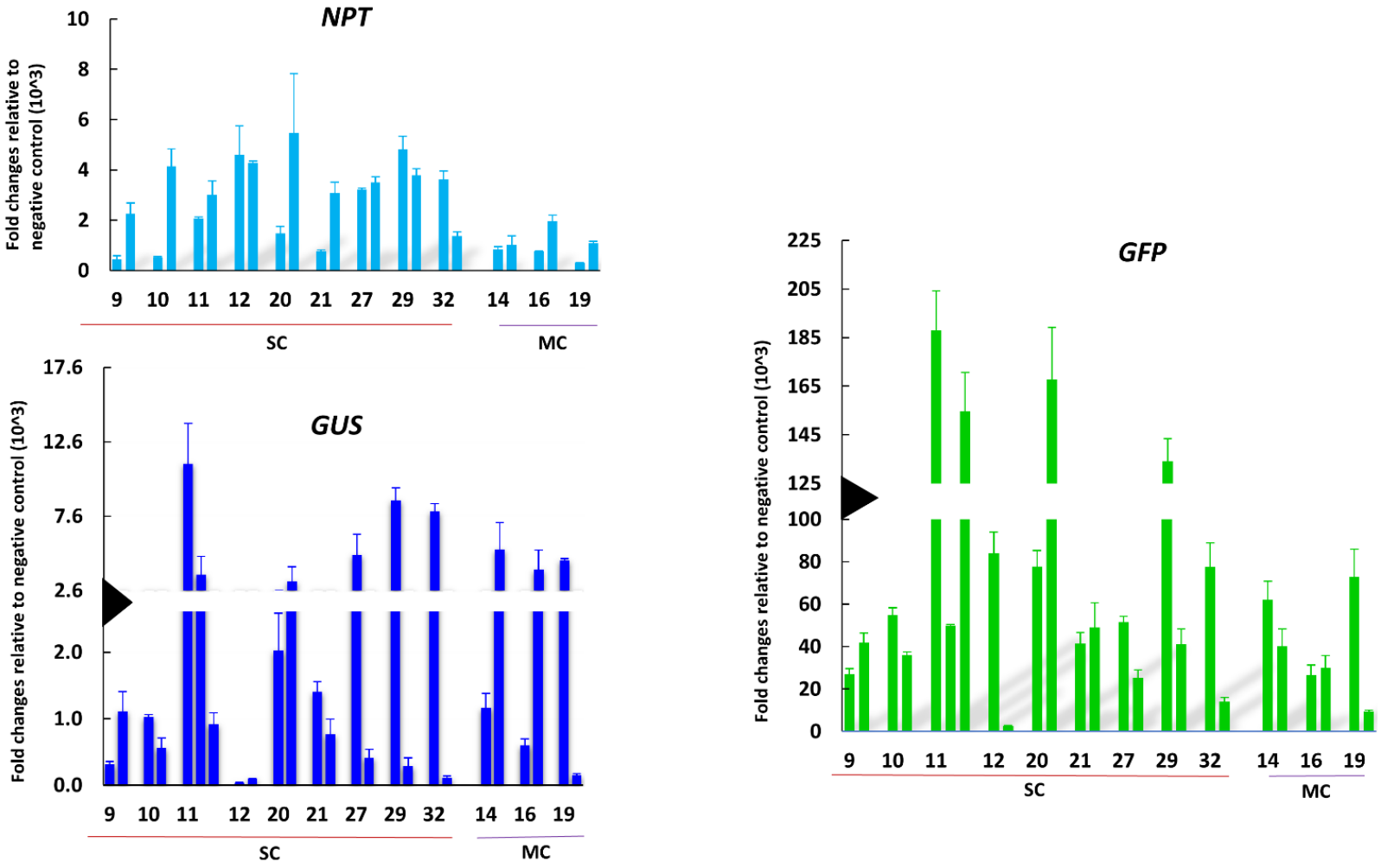


Figure 13: Confocal imaging of GFP (top) and *ppor*RFP (bottom) in roots and leaves of the T1 plants of the SC line #12**.** All images were taken at 72 hours post heat shock at 20x magnification. Image bar indicates the magnification, offset, and zoom used in all the images. T5: Negative control; RT: Room temperature; HS: Heat-shock: # 1, 2, and 3 are the images from three different seedlings.



Appendix figure 1: Time course confocal imaging of *ppor*RFP in 7 – 10 days old T1 seedlings of SC line #9 captured at 24, 48 and 72 hours post heat-shock treatment at 20x magnification. GFP imaging is included as an internal control. Image bar indicates the magnification, offset, and zoom used in all the images. T5: Negative control; RT: room temperature; HS: Heat-shock.



Appendix figure 2: Expression analysis of constitutively expressed genes by RT-qPCR on T1 plants of 12 SSI lines. Error bars are standard error of two technical replications. SC: Single copy; MC: Multicopy.

CHAPTER III

CHARACTERIZATION OF I-*Sce*I AND *CCR5*-ZFN NUCLEASES ACTIVITIES FOR TARGETED EXCISIONS IN RICE AND *ARABIDOPSIS*.

Abstract

Objectives

Removal of selection marker genes from transgenic plants is highly desirable for their regulatory approval and public acceptance. This study evaluated the use of two nucleases, the yeast homing endonuclease, I-*Sce*I, and the designed zinc finger nuclease, *CCR5*-ZFN, in excising marker genes from plants using rice and *Arabidopsis* as the models.

Results

In an *in vitro* culture assay, both nucleases were effective in precisely excising the DNA fragments marked by the nuclease target sites. However, rice cultures were found to be refractory to transformation with the I-*Sce*I and *CCR5*-ZFN overexpressing constructs. The inducible I-*Sce*I expression was also problematic in rice as the progeny of the transgenic lines expressing the heat-inducible I-*Sce*I did not inherit the functional gene. On the other hand, heat-inducible I-*Sce*I expression in *Arabidopsis* was effective in creating somatic excisions in transgenic plants but ineffective in generating heritable excisions. The inducible expression of *CCR5*-ZFN in rice, although transmitted stably to the progeny, appeared ineffective in creating detectable excisions. Therefore, toxicity of these nucleases in plant cells poses major bottleneck in their application in plant biotechnology, which could be avoided by expressing them transiently in cultures *in vitro*.

Introduction

Selection marker genes are indispensable tools in genetic engineering. Their presence in transgenic crops, however, could be detrimental [1], requiring methods for removing them from the plant. The most desirable outcome is to precisely delete the marker genes without creating off-target mutations. The Cre-*lox* site-specific recombination system is highly successful in achieving that goal [2, 3, 4], but it leaves a reactive footprint, the functional *lox* site, in the genome, rendering it non-reusable for the next round of transformation [5, 6].

The double-stranded break (DSB) repair mechanism has long been proposed as an alternative approach for excising marker genes, which can be repeatedly used in the same transgenic line as this mechanism destroys the target site by creating insertion-deletions (indels). Several nucleases, including meganuclease, ZFN, and CRISPR/Cas have been used for creating concomitant DSBs to achieve transgene deletions in the plant cells [7, 8, 9, 10, and 11]. However, their applications in generating marker-free plants need more investigation. This study evaluated the effectiveness of codon-optimized I-*Sce*I [12] and *CCR5*-ZFN [13] in excising genes in rice and *Arabidopsis* using overexpression and inducible expression approaches. These two nucleases were chosen because they have been successfully used in plant genome engineering [10, 14, 15, and 16].

In this study, the expression of I-*Sce*I and *CCR5*-ZFN appeared to be deleterious as indicated by the failure to transform rice with the overexpression constructs, indicating their activity on non-canonical target sites. The inducible expression was ineffective in creating excisions in plants and/or transmitting them to the progeny. Retransformation approach, on the other hand, was successful in creating targeted excision in cultures *in vitro*. Therefore, the use of nucleases in plants is hampered by their genotoxic property and lower efficiencies, but

retransformation of *in vitro* cultures could serve as a practical solution for creating targeted excisions, which could then be regenerated into plants. However, several 'excision events' will have to be screened for precise targeted excisions and the potential off-target mutations.

Main Text

Methods:

DNA constructs, plant transformation, and treatments:

All constructs were prepared using the standard molecular biology techniques. The synthetic coding sequences of I-*Sce*I and *CCR5*-ZFN were provided by Drs. Holger Puchta (Karlsruhe, Germany) and Joseph Petolino (Dow Agro Sciences, Inc.), respectively. *Agrobacterium*-mediated and biolistics- mediated rice (Nipponbare) transformations have been described earlier [9, 17]. *Arabidopsis* (Col-0) transformation was done using the floral-dip method [18]. Heat-shock treatments of rice *in vitro* cultures, cut leaves or the seedlings was done by placing the tissues in the petri-dish or wrapped in aluminum foil in an incubator maintained at 42ºC for 3 hours, followed by 72 hours of recovery before scarifying the tissue for DNA/RNA isolation. For *Arabidopsis*, seedlings in the germination media (MS media without sucrose) were placed in 40ºC for 3 hours followed by 48 hours of recovery.

Molecular analysis:

The PCR primers were designed using Primer Blast tool and verified in the IDT oligo-analyzer for the hairpin, self and heterodimer structures. They were also checked by BLAST to look for any potential non-specific sites in the rice and *Arabidopsis* genomes. Primers used in the present study are given in Additional File 1: Table S1. PCR was performed at 94ºC for 4 min followed by 40 cycles of 1 min at 58-60º C and 1-2 min at 72 ºC depending on the amplicon size (unless otherwise stated) using Emerald Amp PCR master mix (TaKaRa Inc.). All the PCR

assays included the non-transformed rice or *Arabidopsis* genomic DNA as the negative control to screen for any non-specific amplification. For gene expression analysis, total RNA isolated using RNaesy kit (Qiagen Inc.) was subjected to real-time PCR using Super Script III one step qRT-PCR kit (Invitrogen) using manufacturer's instructions. Relative expression was calculated against wild-type using $2^{\Delta\Delta Ct}$ method [19], and the Ct values were normalized against internal control, *Ubiquitin* or *Phytoene Desturase* genes. The purified PCR products were sequenced at Eurofin Genomics USA. Genomic DNA of selected lines was also analyzed on Southern blot using P32- labeled DNA probes.

Results

Expression of I-SceI and ZFN in rice

The overexpression constructs consisting of ZmUbi1 promoter for I-*Sce*I or ZFN expression (**Fig. 1a**) were co-bombarded with hygromycin resistance gene (hygR) on the scutellar callus of rice cv. Nipponbare. The hygR gene consisted of hygromycin phosphotransferase gene driven by CaMV 35S promoter. No selectable clones were obtained with I-*Sce*I overexpression construct in two different experiments, suggesting geno-toxicity of I-*Sce*I in rice. With ZFN overexpression construct, 11 hygR lines were generated that were PCR-positive for ZFN gene. However, only 3 of these set a low number of seeds (10-30 seeds/line), indicating high rate of sterility in ZFN rice plants. The PCR analysis of the T1 plants from these 3 lines revealed lack of inheritance of the ZFN gene (**Additional file 2: Figure S1**). Therefore, strong expression of ZFN also generated toxicity in rice cells that severely hampered inheritance of the ZFN gene. The BLASTn analysis, (using default parameters- input: 33 or 18 bp**;** e-value threshold: 10**;** match/mismatch score:1,-3**;** gapopen: -5 and gapextend: -3) of 18 bp I-*Sce*I and 33 bp *CCR5* sites did not reveal match in the rice or *Arabidopsis* genome. The online tools for

predicting off-target of I-*Sce*I are lacking, but five I-*Sce*I like sites [20] were also used in the BLASTn analysis, none of which found a 100% match in the rice or *Arabidopsis* genome. Off-target prediction of the *CCR5*-ZFN by Prognos tool [21] found 12 highly probable sites in the rice genome.

Next, inducible expression constructs consisting of *GmHSP17.5E* gene promoter expressing I-*Sce*I or ZFN (**Fig. 1b**) were co-transformed with hygR gene into Nipponbare callus. Seven I-*Sce*I and 8 ZFN lines were recovered, indicating curbed toxicity of the inducible I-*Sce*I and ZFN in rice. Expression analysis was conducted on heat-shock-treated (HS) cut leaves obtained from the greenhouse grown plants. Five HS-I*Sce*I lines and 7 HS-ZFN lines showed several fold increase in the expression with respect to the untreated control, confirming proper regulation of these nucleases in the rice plant (**Fig. 1c-d**). The HS-ZFN lines showed normal growth and fertility, and transmitted ZFN activity to the progeny. The HS-I*Sce*I lines, on the other hand, did not transmit I-*Sce*I gene to the progeny and showed poor growth and high sterility, indicating toxicity of the basal expression of the inducible I-*Sce*I gene to the somatic and germ cells.

Characterization of inducible ZFN activity in excising marker gene in rice plants

While the experiments with HS-I*Sce*I had to be discontinued due to problematic heritability of I-*Sce*I gene, HS-ZFN lines were cross-pollinated with *CCR5* target lines developed by transformation of Nipponbare rice with pBP5 that contains 3 gene cassettes, *GFP*, *HPT* and *NPT*, with a pair of 33 bp *CCR5* sites flanking the *HPT* cassette (**Fig. 2a**). Targeting of *CCR5* sites by ZFN could lead to the excision of *HPT* and fusion of the distal ends creating indels at the targeted sites (**Fig. 2b**). Five healthy F1 plants representing 3 different ZFN lines (lines #3, #6, #7; **Fig. 1b**) and two different *CCR5*-target lines (**Fig. 2c**) were heat-shocked and grown to

maturity in the greenhouse. All F1 plants expressed *GFP* and the HS-induced ZFN activity, confirming the presence of *CCR5* target and ZFN constructs; however, excision of the *HPT* cassette was undetectable by PCR across *CCR5* sites (data not shown). Several F2 seedlings that were positive for *GFP* and ZFN were also heat-shocked and sacrificed for DNA isolation, but none showed the excision site (≤1.3 kb) in the PCR, while the presence of intact target site (3.5 kb) was evident in a number of them (**Fig. 2d**). Hence, HS-induced ZFN activity appeared suboptimal in creating detectable excisions in rice. This observation corroborates with that of Lu et al. [22], who reported low frequency targeting by heat-inducible ZFN in poplar.

Targeted excisions by retransformation

The failure in scoring targeted excisions in the F1 hybrids and their progeny derived from the crosses between HS-ZFN and *CCR5*-target lines raised questions whether ZFN expression was sufficient and the target locus was accessible to ZFN activity. To address these questions, reciprocal transformations were done, *i.e.*, transformation of ZFN-expressing line with pBP5, and transformation of *CCR5*-target lines with pHS: ZFN. Retransformation of HS-ZFN line #7 with pBP5 generated 19 geneticin-resistant calli events that expressed *GFP*, indicating stable integration of the target construct in the genome. PCR across *CCR5* sites found that 17 of these lines showed both full-length *HPT* cassette (3.5 kb) and the excision site (≤1.3 kb) in the room temperature (RT) samples, 4 of which showed strong presence of excision site in the heat-shock (HS) samples (**Fig. 2e**). These data suggest that basal ZFN activity from HS: ZFN gene could induce targeting at *CCR5* sites but the targeting efficiency increased upon HS treatment. Four regenerated plants were obtained from these callus lines that also showed the ~1.3 kb excision site (**Fig. 2e**). Similarly, transformation of the *CCR5*-target lines with pHS: ZFN vector, produced 9 calli events, 4 of which showed ~1.3 kb excision band in HS-treated calli (**Fig. 2f**).

Sequencing of 5 excision sites (≤1.3 kb) from these experiments found complete or partial excision of HPT cassette with large indels (>1.5 kb) spreading into the adjacent sequences (**Fig. 2g**). In summary, HS-induced ZFN activity is capable of creating targeted excisions in rice cultures *in vitro*.

Inducible I-SceI mediated marker excision in Arabidopsis

Since I-*Sce*I expression was highly toxic in rice, further experiments with inducible I-*Sce*I were carried out in *Arabidopsis*. For this purpose, pEP4b construct was developed that contains a pair of I-*Sce*I target sites flanking the *GFP* cassette, the kanamycin resistance (*NPT*) cassette, and the HS-inducible I-*Sce*I expression cassette (**Fig. 3a**). The excision of the *GFP* cassette in this construct would result in fusion of I-*Sce*I and *NPT* cassette with indels in between (**Fig. 3b**). Transformation of *Arabidopsis* Col-0 with pEP4b generated 11 kanamycin resistant T1 lines that contained a full-length integration of pEP4b construct in the PCR assay (**Fig. 3c**). Fertility in these T1 plants was substantially low, indicating I-*Sce*I toxicity in the germline (≤10x lower compared to that of the healthy *Arabidopsis* plants). Germination of T2 seedlings on kanamycin-containing (50 mg/l) media displayed gradual lethality and receding GFP expression in all lines; however, seedlings could be rescued on a kanamycin-free medium and grown to maturity. This indicates that large indels possibly occurred at the target sites, eliminating *NPT* and *GFP* activity. The rescued T2 seedlings were analyzed by PCR to determine the target and excision sites, indicated by 3.0 and 1.2 kb products, respectively (**Fig. 3a-b**). The majority of T2 progeny either failed to show these PCR products or showed their weak presence, indicating large indels at the target site in the majority of the tissue. Two T2 lines showed strong presence of ~1.2 kb band (**Fig**. **3d**: white arrows), which was sequenced and found to contain the near-precise excision of GFP cassette with very small indels at the target sites (**Fig. 3e**). The analysis

of T3 seedlings, however, suggested that the observed excision site in the T2 parents was not transmitted to the progeny as none showed the 1.2 kb band (**Fig. 3d**). In summary, HS-I*Sce*I was able to generate targeted excisions in the *Arabidopsis* seedlings, but inheritance of the excision site was questionable.

Conclusions

Potential geno-toxicity of I-*Sce*I and *CCR5*-ZFN appears to be a major bottleneck in their application in plant biotechnology. However, retransformation of *in vitro* cultures could be used as an effective approach for excising of marker genes and regenerating the marker-free plants.

Limitations

The main limitation of this study is that rice and *Arabidopsis* genomes could contain off-target sites of I-*Sce*I and *CCR5*-ZFN nucleases that would prohibit the application of these nucleases in these plant species. A larger set of nucleases, *e.g.*, newly designed ZFNs or TALENs should be tested to determine if other nucleases can be used successfully in achieving marker excision in these plant species.

References:

1. Dale PJ, Clarke B, Fontes EM. Potential for the environmental impact of transgenic crops. Nat. Biotechnol. 2002; 20(6):567-74. doi: 10.1038/nbt0602-567.

2. Gidoni D, Srivastava V, Carmi N. Site-specific excisional recombination strategies for elimination of undesirable transgenes from crop plants. In Vitro Cell. Dev. Biol.– Plant. 2008; 44: 457-467. https://doi.org/10.1007/s11627-008-9140-3.

3. Gilbertson L. Cre-*lox* recombination: Cre-ative tools for plant biotechnology. Trends Biotechnol. 2003; 21(12):550-5.doi: 10.1016/j.tibtech.2003.09.011

4. Ow, DW. The long road to recombinase-mediated plant transformation. Plant Biotechnol. J. 2016;14: 441-447. doi:10.1111/pbi.12472.

5. Srivastava V, Thomson J. Gene stacking by recombinases. Plant Biotechnol. J. 2016; 14: 471–482. doi: 10.1111/pbi.12459

6. Srivastava V. Gene stacking in plants through the application of site-specific recombination and nuclease activity. Methods Mol. Biol. 2019; 1864:267-277. doi: 10.1007/978-1-4939-8778-8-18.

7. Antunes MS, Smith JJ, Jantz D, Medford JI. Targeted DNA excision in *Arabidopsis* by a re-engineered homing endonuclease. BMC Biotechnol. 2012; 12:86. doi: 10.1186/1472-6750-12-86.

8. Fauser F, Roth N, Pacher M, Ilg G, Sánchez-Fernández R, Biesgen C, Puchta H. *In planta* gene targeting. Proc. Natl. Acad. Sci. USA. 2012. 109(19):7535-40. doi: 10.1073/pnas.1202191109.

9. Nandy S, Zhao S, Pathak B, Manoharan M, Srivastava V. Gene stacking in plant cell using recombinases for gene integration and nucleases for marker gene deletion. BMC Biotechnol. 2015; 15:93 doi: 10.1186/s12896-015-0212-2.

10. Petolino JF, Worden A, Curlee K, Connell J, Strange Moynahan TL, Larsen C, Russell S. Zinc finger nuclease-mediated transgene deletion. Plant Mol. Biol. 2010; 73:617–628. doi: 10.1007/s11103-010-9641-4.

11. Srivastava V, Zhao S, Underwood J. Dual-targeting by CRISPR/Cas9 for precise excision of transgenes from rice genome. Plant Cell Tiss. Org. 2017; 129:153-160. doi: 10.1007/s11240-016-1166-3.

12. Hlubek A, Biesgen C, Höffken H-W. Chimeric endonucleases and uses therefore. 2011; Patent WO 2011/064750.

13. Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat. Biotechnol. 2008; 26(7):808–16. doi:10.1038/nbt1410.

14. Ainley WM, Sastry-Dent L, Welter ME, Murray MG, Zeitler B, Amora R, et al. Trait stacking via targeted genome editing. Plant Biotech J. 2013; 11:1126–34. doi: 10.1111/pbi.12107

15. D'Halluin K, Vanderstraeten C, Van Hulle J, Rosolowska J, Van Den Brande I, Pennewaert A, et al. Targeted molecular trait stacking in cotton through targeted double-strand break induction. Plant Biotechnol. J. 2013; 11:933–41. doi: 10.1111/pbi.12085.

16. Watanabe K, Breier U, Hensel G, Kumlehn J, Schubert I, Reiss B. Stable gene replacement in barley by targeted double-strand break induction. J. Exp. Bot. 2016; 67(5):1433-45. doi: 10.1093/jxb/erv537.

17. Nishimura A, Aichi I, Matsuoka M. A protocol for Agrobacterium-mediated transformation in rice. Nat. Protoc. 2006; 1(6):2796. doi: 10.1038/nprot.2006.469

18. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis* thaliana. Plant J. 1998;16(6):735-43.

19. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 2001; 25(4):402-8. doi: 10.1006/meth.2001.1262

20. Petek LM, Russell DW, Miller DG. Frequent endonuclease cleavage at off-target locations *in vivo.* Mol. Ther. 2010; 18(5):983-6. doi: 10.1038/mt.2010.35

21. Fine EJ, Cradick TJ, Zhao CL, Lin Y, Bao G. An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. Nucl. Acids Res. 2013; 42(6): e42. doi: 10.1093/nar/gkt1326.

22. Lu H, Klocko A, Dow M, Ma C, Amarasinghe V, Strauss SH. Low frequency of zinc-finger nuclease-induced mutagenesis in *Populus*. 2016; Mol. Breed. 36 (9):121. doi; 10.1007/s11032-016-0546-z.

Tables and Figures

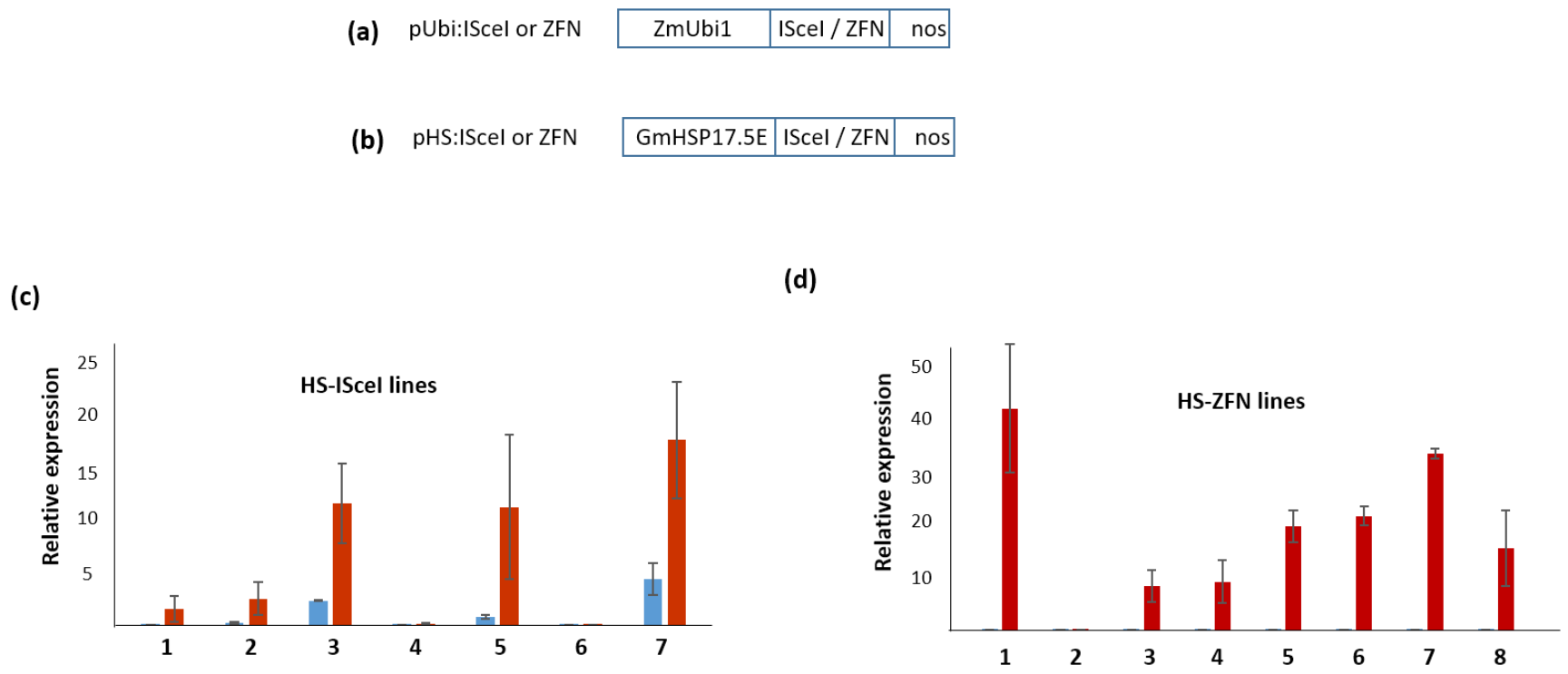


Fig. 1: Expression of I-*Sce*I and ZFN in rice. (**a, b**) Overexpression and inducible constructs of I-*Sce*I or ZFN contain ZmUbi1 for constitutive overexpression or GmHSP17.5E for HS-inducible expression with *nos 3*' as transcription termination sequence. (**c, d**) Real-time quantitative PCR analysis on total RNA isolated from the rice lines expressing HS inducible I-*Sce*I or ZFN gene. Relative expression against wild-type control is shown for each line. Bars show mean of 2 treatments with standard errors. Red and blue bars represent HS and room temperature (RT) samples, respectively. Note that ZFN expression at RT was close to the wild-type controls.

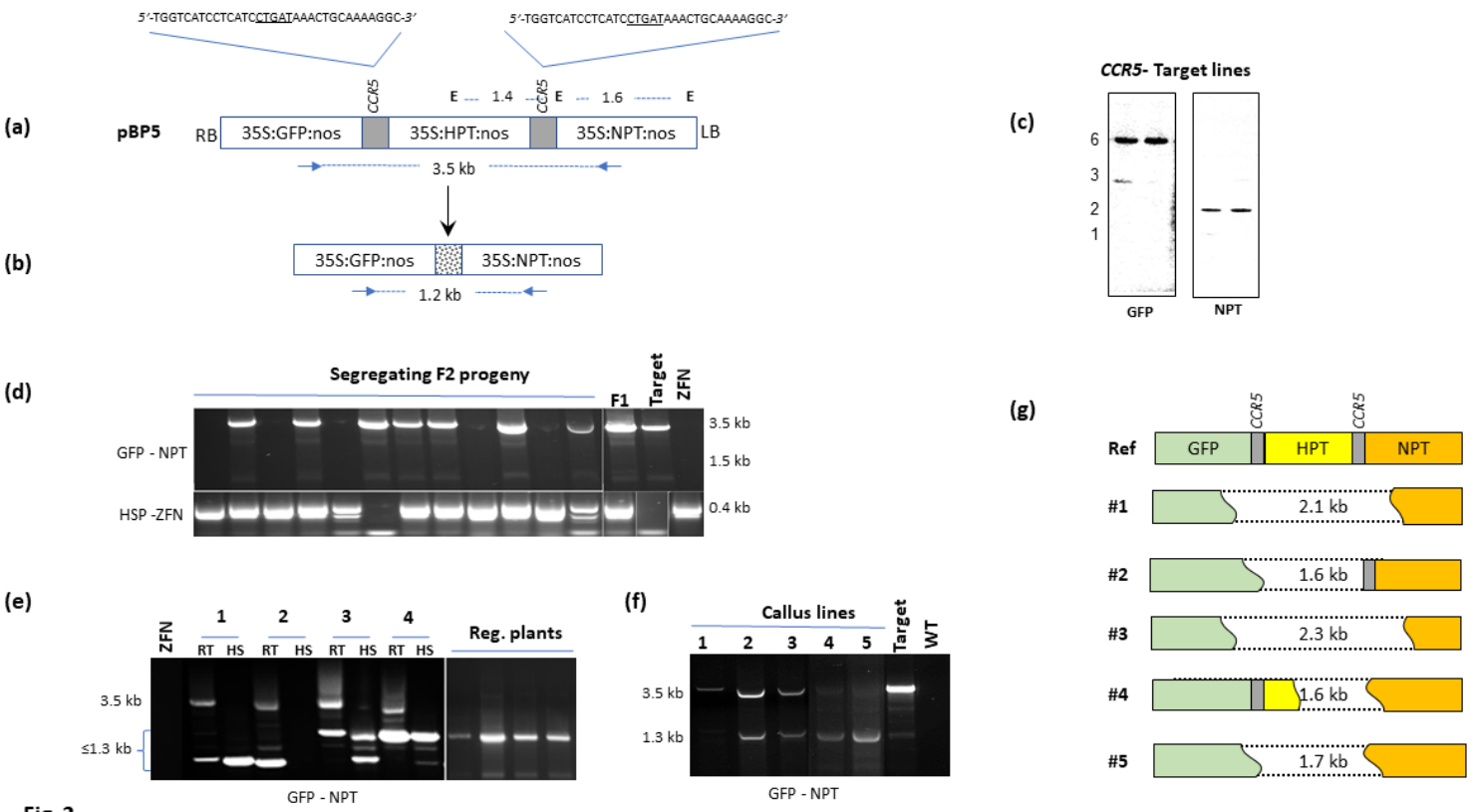


Fig. 2

Fig. 2: Characterization of HSP-ZFN in rice. (**a**) The *CCR5*-target construct in pPZP200 binary vector contains *GFP*, *HPT* and *NPT* genes. Each of which is controlled by *35S* promoter and *nos 3'* terminator. The *HPT* gene is flanked by 33 bp *CCR5* sequences (gray bars). Location of *Eco*R1 (E) sites and the fragment sizes are shown. (**b**) Predicted structure of ZFN-induced precise excision of *HPT* cassette with indels in between (dotted bar). PCR primer positions and predicted fragment sizes (in kb) are shown below each structure. (**c**) Southern blot analysis of rice lines transformed with pBP5. Genomic DNA was cut with *Eco*RI and hybridized with P^{32} labeled *GFP* or *NPT* probes. Fragment sizes are given in kb. (**d**) PCR analysis using primers located in *CCR5*-target sites (*GFP – NPT*) or ZFN gene (HSP – ZFN) on genomic DNA isolated from F2 plants derived from crosses between *CCR5*-target lines and HSP-ZFN lines. F1 parent, and *CCR5*-target and ZFN lines are also shown. (**e**) PCR across *CCR5* sites in the retransformed callus clones

Fig. 2 (Cont.)

and the regenerated plants obtained by retransformation of HS-ZFN line #7 (Fig. 1d) with pBP5. The room temperature (RT) or heat-shocked (HS) samples of the selected calli clones (1 – 4) are shown with the regenerated plants obtained from them. ZFN line #7 serves as the negative control. (**f**) PCR across *CCR5* sites in the retransformed clones derived from the retransformation of *CCR5*-target lines with pHSP: ZFN construct. Target line and wild-type (WT) are included as controls. (**g**) Depiction of indels created by targeting of the two *CCR5* sites in the target site as determined by aligning the DNA sequences of selected ≤1.3 kb bands with pBP5 reference. Deletions sizes are given in each diagram.

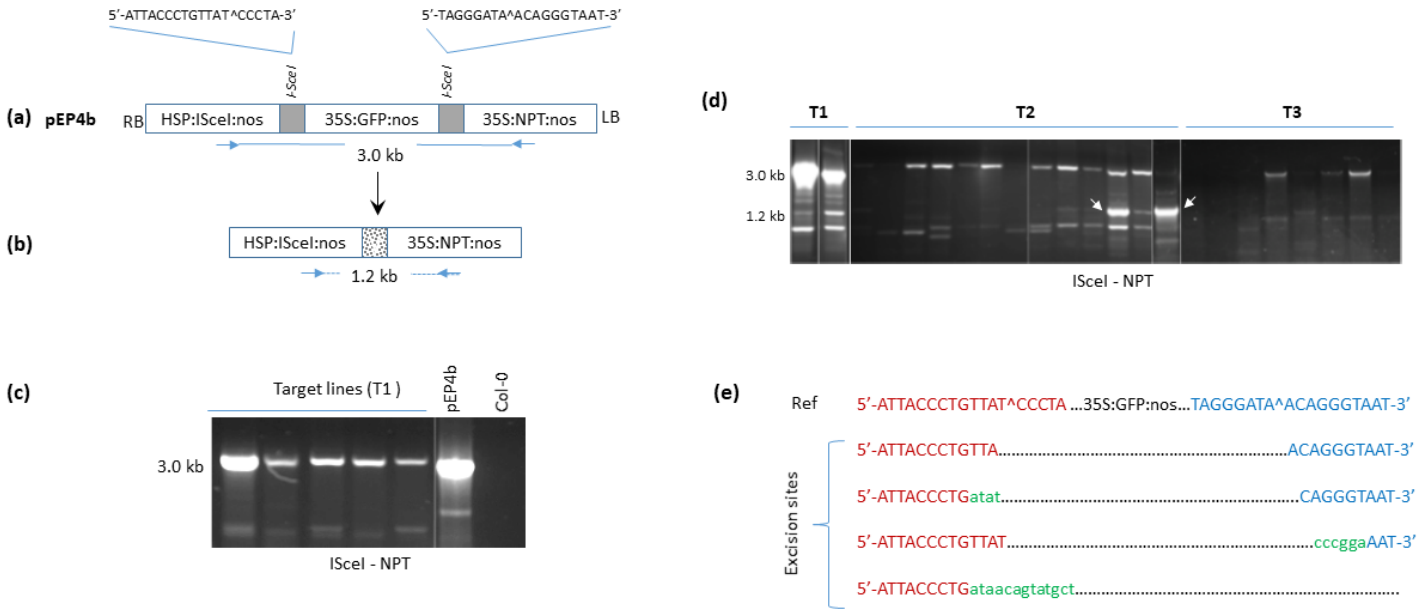


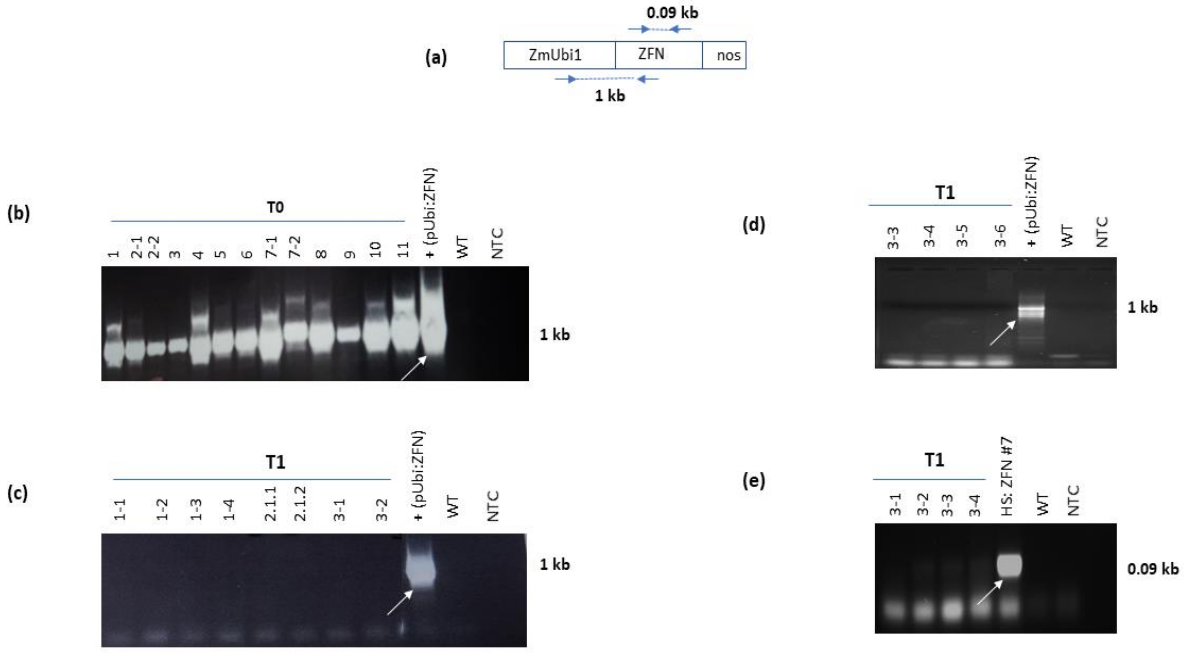
Fig. 3

Fig. 3: Characterization of HS-inducible I-*Sce*I in *Arabidopsis*. (**a**) I-*Sce*I target construct, pEP4b, in pPZP200 binary vector contains HS-inducible I-*Sce*I, *GFP*, and *NPT* expression units with 18 bp I-*Sce*I target sites (gray bars) flanking the *GFP* cassette. (**b**) Predicted structure of the target site upon precise excision of *GFP* cassette with indels at the targeted site (dotted bar). PCR primer positions and the fragment sizes are shown by blue arrows. (**c**) PCR analysis of the first generation transgenic (T1) lines using primers located in I-*Sce*I and *NPT* cassettes with pEP4b and wild-type Col-0 as controls. (**d**) PCR analysis of three generations: T1 parents, T2, and T3 progeny to detect excision of *GFP* cassette. White arrows indicate bands that were purified and subjected to Sanger sequencing. (**e**) DNA sequences of ~1.2 kb predicted excision bands were aligned with the pEP4b reference to determine indels at the targeted sites. Red and blue fonts represent the two I-*Sce*I sites with predicted breakpoints (^). Dotted lines indicate deletions and green small letters show insertions.

Additional File S1:

Table S1: List of the primers used in this study

Primers	**Sequence (*5' – 3'*)**	**Application**
pEP4b primers	TTCTCCACACCATGTACGCA	Genotyping *Arabidopsis* pEP4b lines
	GCATCGCCTTCTATCGCCTT	
pBP5 primers	AAGACCCCAACGAGAAGC	Genotyping rice pBP5 lines
	CTCGATGCGATGTTTCGCTT	
pHSP:ZFN primers	CCTTGCGTACATGGTGTGGA	Genotyping HS-ZFN lines
	TGCAGATTCGACACTGGAAG	
qZFN-F	TGAATGGTGGAAGGTGTATCC	Expression analysis of ZFN in rice
qZFN-R	AAGCTGTGCTTTGTAGTTACCC TTA	
qI-SceIF	GCTGTCTCCTCCTCACAAG	Expression analysis of I-*Sce*I in rice
qI-SceIR	GGGTCAGGTAGTTCTCCACC	
qUbi-F	CGCAAGTACAACCAGGACAA	Reference gene for expression analysis in rice
qUbi-R	GCTGTGACCACACTTCTTCTT	
qPDS-F	GCAGAGGAATGGGTTGGAC	Reference gene for expression analysis in rice
qPDS-R	GTGAACCTTGCCGACCTCT	



Additional File 2.ppt: Figure S1: Molecular analysis of rice lines transformed with ZFN overexpression construct. (**a**) ZFN overexpression construct containing maize Ubiquitin-1 (ZmUbi) promoter, ZFN coding region and nopaline synthase (nos) *3'* transcription terminator. Primer positions and their product size are shown. (**b**) PCR analysis of 13 primary transgenic plants (T0) representing 11 transgenic events. (**c**) PCR analysis of T1 progeny from three T0 plants # 1, 2-1 and 3. (**d, e**) PCR analysis of additional T1 progeny from line #3. Product sizes are shown. Arrows indicate expected products in each gel. The PCR conditions for Figures (b-d) are mentioned in the main text. The PCR for 0.09 kb product (Figure e) was performed at 95ºC for 3 min followed by 30 cycles of 95ºC for 30 sec, 60ºC for 30 sec, and 72ºC for 30sec.

CHAPTER IV

EVALUATION OF CRISPR/CAS9 IN GENERATING TARGETED MUTATIONS IN BY CONSTITUTIVE EXPRESSION OF CAS9 IN THE RICE GENOME

Abstract

The present study investigated the efficiency of CRISPR/Cas9 in creating genomic deletions as the basis of its application in removing selection marker genes or the intergenic regions. Three loci, representing a transgene and two rice genes, were targeted at two sites each, in separate experiments, and the deletion of the defined fragments was investigated by PCR and sequencing. Genomic deletions were found at a low rate among the transformed callus lines that could be isolated, cultured, and regenerated into plants harboring the deletion. However, randomly regenerated plants showed mixed genomic effects, and generally did not harbor heritable genomic deletions. To determine whether point-mutations occurred at each targeted site, a total of 114 plants consisting of primary transgenic lines and their progeny were analyzed. Ninety-three plants showed targeting, 60 of which were targeted at both sites. Presence of point-mutations at both sites was correlated with the guide RNA efficiency. In summary, genomic deletions through dual-targeting by the paired-guide RNAs were generally observed in callus, while *de novo* point-mutations at one or both sites occurred at high rates in transgenic plants and their progeny, generating a variety of insertion-deletions or single nucleotide variations. In this study, point-mutations were exceedingly favored over genomic deletions; therefore, for the recovery of plant lines harboring targeted deletions, identifying early transformed clones harboring the deletions, and isolating them for plant regeneration is recommended.

Introduction

Genome-editing effects are based on the creation of double-stranded breaks (DSB) in the target DNA that are repaired by the cell through non-homologous end-joining (NHEJ) or homology-directed repair (HDR) pathways (Jasin and Haber 2016; Waterworth et al. 2011). While HDR leads to predictable outcomes as determined by the DNA template, NHEJ ends up with insertions, deletions and/or substitutions (Puchta et al. 1996; Rouet et al. 1994; Szostak et al. 1983), leading to gene knockouts. The power of CRISPR/Cas9 lies in its efficiency in creating DSBs in genomic sequences containing NGG protospacer adjacent motif (PAM). The simplified version of CRISPR/Cas9 consists of a single-guide (sg) RNA bound to Cas9 (sgRNA: Cas9) that targets genomic sequences through RNA–DNA pairing. Although, sgRNA design is based on a relatively simple 5′-N(20)-NGG-3′ targeting rule (Cong et al. 2013; Jinek et al. 2012; Mali et al. 2013; Mojica et al. 2009), the efficiency of different sgRNAs could vary in the cell. Therefore, multiple sgRNAs are often used in creating targeted knockouts. As a result, targeted genomic deletions by CRISPR/Cas9 have been observed in numerous studies.

Dual-targeting by CRISPR/Cas9, based on the paired use of sgRNAs, could generate somatic and heritable deletions of genomic fragments. Short deletions of ~ 100 bp are frequently reported in plants (Brooks et al. 2014; Kapusi et al. 2017; Nekrasov et al. 2017; Ordon et al. 2017). Dual- targeting was also effective in deleting larger fragments (~ 0.5 kb, ~ 0.7 kb, and 1.6 kb) as reported in maize, kiwi fruit, and rice (Minkenberg et al. 2017; Shi et al. 2017; Srivastava et al. 2017; Wang et al. 2018). Fragments of 10–12 kb could be deleted in rice and *Arabidopsis* (Durr et al. 2018; Wang et al. 2017a), and even larger fragments of 170–245 kb were deleted by multiplex targeting in rice (Zhou et al. 2014). The efficiencies of genomic deletions varied greatly in these reports, but short deletions (~ 100 bp) were obtained

more readily than large deletions (Kapusi et al. 2017; Ordon et al. 2017; Wang et al. 2017a). However, compared to point mutagenesis (effect of a single sgRNA), genomic deletions (effect of paired sgRNAs) consistently occurred at much lower rate even when two or more sgRNAs of equal efficiencies were used (Minkenberg et al. 2017; Tian et al. 2017; Wang et al. 2017a, b).

The application of CRISPR/Cas9 in genome editing is limited by the DNA repair pathways of the host organism. In somatic cells of plants and other higher organisms, NHEJ is the major repair pathway (Puchta et al. 1996; Waterworth et al. 2011); therefore, targeted mutagenesis is the most successful application of CRISPR/Cas9. Another genomic effect that could be created by NHEJ is fragment deletion by a pair of sgRNAs to simultaneously create DSBs at two different sites on a segment of the genome (dual-simultaneous targeting). Ligations of the two distal ends through NHEJ would effectively delete the intervening fragment. Genomic deletions could serve as useful editing effects in functional genomics and biotechnology by targeting gene clusters, *cis*-regulatory elements or transgenes. However, current understanding of dual-targeting by CRISPR/Cas9 in creating genomic deletions is narrow. Many studies have reported genomic deletions, but little is known about the efficiency and success in recovering stable plants lines harboring the defined deletion.

The present study investigated the efficiency of obtaining defined genomic deletions of 240 bp, 945 bp, and 1637 bp from three different loci by dual-targeting in rice. Defined deletions were detected by PCR among transformed calli, and as expected, plants regenerated from these calli harbored the deletions and transmitted to their progeny. However, randomly regenerated plants harboring mixed genomic effects either did not show deletions or showed a low rate of somatic deletions. Furthermore, while targeting frequency of each sgRNA increased in the progeny, genomic deletions remained undetectable. Therefore, for ensuring the recovery of plant

lines harboring deletions defined by dual-targeting, it is recommended to screen early transgenic clones (calli) and isolate the characterized clones for plant regeneration. The recovery of *de novo* deletion lines through plant screening and progeny analysis, at least in rice, appears to be highly unlikely.

Materials and methods

DNA constructs and plant transformation

The sgRNA spacer sequences were selected using CRISPR RGEN tool (http://www.rgenome.net/cas-designer/; Park et al. 2015). Vector pRGE32 (Addgene#63159) was used for synthesizing the CRISPR/Cas9-targeting vectors pJU24, pJU34, and pJU46 against *GUS* (NCBI accession no. AF485783), *OsPDS* (Os03g08570), and *Chalk5* (Chromosome 5: 3,335,405–3,341,600) genes, respectively. The two sgRNAs targeting each gene were expressed as polycistronic tRNA–gRNA (PTG) genes, which was synthesized against pGTR (Addgene# 63143) using the protocol of Xie et al. (2015). The constructed PTG (tRNA–gRNA1–tRNA–gRNA2) was ligated to pRGE32 vector by FokI/BsaI digestions, and the resulting vectors were used for rice transformations. The gRNA oligos used for PTG construction are given in Table 4. For targeting *GUS*, B1 transgenic line (cv. Nipponbare) was used for trans- formation as described earlier by Srivastava et al. (2017), while Nipponbare was used for targeting rice genes, *OsPDS* and *OsChalk5*. The embryogenic callus from mature seeds was used for all transformations by the gene gun (PDS1000, Bio-Rad Inc.), in which pJU24, pJU34, or pJU46 DNA was co-bombarded with hygromycin phospho-transferase expressing vector, p35S:HPT. The transformed calli were isolated and regenerated on hygromycin (50 mg/l) containing media using the protocol of Nishimura et al. (2006).

Molecular analysis

Genomic DNA isolated from callus, regenerated plants or seedlings, was used for polymerase chain reaction (PCR) using primers spanning the target sites (Table 4). The PCR products were resolved on agarose gel and extracted using Geneclean Spin Kit (MP Biomedicals, CA, USA) for sequencing from both ends using forward and reverse primers by the Sanger Sequencing method at Eurofins Genomics USA. The sequences were viewed on Sequence Scanner 2 software (Applied Biosystems Inc.) and aligned with the reference sequences using CLUSTAL-Omega multiple sequence alignment tool. CRISPR-ID tool was used to separate superimposed overlapping spectrum in Sanger sequencing traces, characteristic of heterozygous or chimeric mutations (Dehairs et al. 2016). The type of indel was identified by cloning PCR amplicon into pCR2.1 vector using TA cloning kit (Thermo-Fisher Scientific, NY) as per manufacturer's instructions and sequencing individual colonies by Sanger sequencing.

Results

Experimental design

The efficiency of CRISPR/Cas9 in deleting genomic fragments was estimated on three loci, *GUS* transgene (AF485783), rice *PDS* (LOC_Os03g08570), and rice *Chalk5* (LOC_Os05g06480.1; Chromosome 5: 3,335,405–3,341,600) (Fig. 1a). Two sites in each locus were chosen based on 5′-N(20)-NGG-3′ rule (Cong et al. 2013; Jinek et al. 2012; Mali et al. 2013), with the goal of creating deletions through simultaneous targeting by a pair of sgRNAs (sg1 + sg2). While *GUS* and *PDS* sgRNAs targeted the genic regions, *Chalk5* sgRNAs targeted an intergenic region harboring *cis*-regulatory elements (Fig. 1a). To generate sg1 and sg2 from a single vector, oligonucleotides containing sgRNA spacers were cloned in pRGE32, which contains tRNA splicing mechanism to generate multiple sgRNAs from a single transcript

produced by the rice *U3* promoter (Xie et al. 2015). The resulting *GUS*-, *PDS*- or *Chalk5*-targeting vectors, pJU24, pJU34, and pJU46, respectively, were transformed into the B1 rice line, expressing the GUS gene, or the wild-type Nippon- bare rice. Line B1 that contains a single-copy of *GUS* gene has been described earlier (Nandy and Srivastava 2012). The resulting transgenic lines were screened by PCR to identify deletions in *GUS*, *PDS*, or *Chalk5* genes, indicated by amplification of fragments shorter by 1637 bp, 987 bp, and 240 bp, respectively (Fig. 1a). A representative PCR indicating genomic deletion in the three loci is shown in Fig. 1b. Targeted deletion of *GUS* in the callus lines has been described earlier (Srivastava et al. 2017). This work further investigated genomic deletions on two more loci, *PDS* and *Chalk5*, and analyzed plant lines to determine the rates of genomic deletions and point mutations through amplicon sequencing by the Sanger method.

Detection of genomic deletions in callus lines

Genomic deletions (Δ) in the callus lines transformed with pJU24, pJU34, or pJU46 were tested by PCR and indicated by the respective Δ amplicons observed in a PCR (Fig. 1b). As reported earlier, *GUS* deletion in pJU24- transformed lines occurred in 2 out of 113 callus lines (Srivastava et al. 2017). In the present study, genomic deletions in two additional loci, *PDS* and *Chalk5* loci, were determined in pJU34- and pJU46-transformed lines (Table 1). Genomic deletions at *PDS* locus was found in 2 out of 32 callus lines and at *Chalk5* locus in 4 out of 53 callus lines. Sequencing of the Δ amplicons indicated that the distal ends, created by the blunt DSBs, ligated without indels or with short indels to generate the Δ locus. The indels generally consisted of insertion or deletion of a single nucleotide or a few nucleotides (Fig. 1c), which is consistent with other studies that report single-nucleotide variations as most common outcome of CRISPR/Cas9 tar- geting (Mao et al. 2013, van Overbeek et al. 2016). One of the pJU46 lines

(Chalk5) showed an amplicon ~ 0.2 kb larger than the intact *Chalk5* amplicon. Sequencing of this amplicon showed insertion of 0.2 kb fragment of unknown source in one of the targeted sites (single-site targeting, data not shown). Overall, the efficiency of creating genomic deletions by dual-targeting was low and variable with the sgRNA pairs (sg1 + sg2). Targeted deletions by *GUS* sgRNA pairs were reported in only 1.7% of the transformed callus lines (Srivastava et al. 2017). The *PDS* and *Chalk5* sgRNA pairs, on the other hand, generated significantly higher rates of deletion at somewhat similar rates in the callus lines (Table 1). Nevertheless, these observations indicate that genomic deletions could be created through dual-targeting by CRISPR/Cas9, and as reported earlier, calli harboring Δ locus could be regenerated into plants (Srivastava et al. 2017). Plants regenerated from one of the callus lines (line#72) contained homozygous Δ locus, indicated by the presence of Δ1637 bp amplicon and absence of 1.8 kb amplicon in the PCR. As expected, the progeny of this plant inherited the stable Δ locus that independently segregated from Cas9 (Fig. 2a). The sequence of the Δ1637 bp in these plants was consistent with the creation of DSB at the predicted sites (3-bp upstream of PAM in each targeted site) followed by ligation of the distal ends without indels (Fig. 2b).

Targeting efficiency in plants

As described above, plant lines carrying the defined Δ locus could be regenerated from calli harboring the deletion. In the same experiment, a number of chimeric T0 plants were also regenerated that showed somatic deletions indicated by the presence of two amplicons, indicative of intact locus and Δ locus, in the same PCR reaction (Srivastava et al. 2017). However, when these chimeric plants were analyzed at a later stage of growth (flowering) in the greenhouse, the Δ1637 bp amplicon was undetectable, in spite of testing multiple tissue from different tillers of each plant. This observation suggests that the young regenerated plants harbored somatic

deletions that are unlikely to be transmitted to the progeny. Among *PDS* and *Chalk5* T0 plants, genomic deletions were undetectable by PCR at both early and late stages of growth (data not shown). To investigate the individual effect of each sgRNA, T0 plants were characterized for the presence of point mutations at each targeted site. A total of 50 T0 plants, representing *GUS*, *PDS*, or *Chalk5* targeting were analyzed by PCR and sequencing (Table 2). Some of these GUS plants selected for this analysis showed Δ1637 bp amplicon in the leaf tissue of the young regenerated plants (Srivastava et al. 2017). Twelve of the 21 GUS plants did not show mutations at either targeted sites. The remaining nine showed targeting but only at sg2 target. Of the 12 *PDS* lines, 3 lacked targeting, while 9 contained targeting at both sites. Finally, 6 out of 17 *Chalk5* lines lacked targeting, and the remaining contained targeting at both sites (Table 2). T0 plants were mostly chimeric for targeting, as 2 or more traces were observed in the characteristic superimposed overlapping peaks downstream of the DSB site in the sequencing spectra. Analysis of these traces revealed the types of mutations found at the DSB sites (Fig. 3). In summary, targeting efficiency of the two *GUS* sgRNAs was highly dissimilar, but the two *PDS* or *Chalk5* sgRNAs showed similar targeting efficiency (Table 2). Sequence alignments of the targeted sites revealed interesting observations: (1) the targeted GUS site in all 9 T0 plants contained only a single-nucleotide variation consisting of 1 bp insertion, deletion or substitution at the predicted DSB site; (2) the two targeted PDS sites contained short deletions ranging from 1 to 7 bp, with only one line containing a larger deletion; and (3) the targeted Chalk5 sites showed most diverse types of mutations with short indels and 1 bp insertions at the two DSB sites (Fig. 3). These observations suggest that possibly genomic context, target sequence, and sgRNA efficiency influence the outcome of CRISPR/Cas9 targeting. In support, a recent study in yeast showed that

types of indels generated by CRISPR/Cas9 depended on DNA sequence context and PAM orientation (Lemos et al. 2018).

Targeting in progeny plants

To investigate inheritance of CRISPR-induced deletions, 61 progeny seedlings derived from three GUS T0 plants were analyzed by PCR. None of the progeny, however, showed Δ1637 bp amplicon, indicative of stable genomic deletion. These plants were also stained for GUS activity, 34 of which were negative, indicating targeting at sg1 and/or sg2 sites. To determine the inheritance of point mutations, selected GUS-negative progeny derived from a single parent plant was analyzed and compared with the parent plant that contained chimeric targeting at sg2 site. In the parent plant, no targeting was evident in sg1 site, but three types of mutations were observed at the predicted sg2 DSB site: + 1 (A or C) and A-to-C substitution (Fig. 4a); however, + 1 C was the most commonly observed mutation in multi-sample analysis that likely rendered the plant GUS negative. None of the T1 plants showed Δ1637 bp amplicon; however, de novo targeting by sg1 was frequently observed. Eight of the 17 T1 plants showed chimeric targeting (≥ 2 types of sequences) at sg1 target. The most common type of mutation at sg1 target was 1 bp deletion; however, 1 bp insertion and longer deletions were also observed (Fig. 4a). The analysis of sg2 target among T1 plants revealed that all 17 plants contained monoallelic or biallelic mutations (Table 3). Biallelic mutations were either identical on each allele (homozygous) or different (heterozygous). The alignment of sequences revealed that all observed mutations were also present in the parent. Four T1 plants (T1–7, 9, 12, 15) had segregated from Cas9 gene, confirming inheritance of the mutation (Fig. 4a). In summary, while targeting at both sites was observed in T1 plants, de novo genomic deletions were undetectable.

Next, T2 progeny derived from three T1 plants (T1–2, T1–3, and T1–4) were analyzed by PCR and sequencing. Once again, no genomic deletion was detected in any of the T2 plants. The three T1 parents all contained identical mutation at sg2 site (+ 1 C), but differed at sg1 site. T1–2 contained 7 bp deletion at sg1 site, but its progeny completely lacked mutations at sg1 sites and contained de novo single-nucleotide variation (+ 1 A) at sg2 site, indicating that mutations observed in the parent were not heritable and de novo mutations were introduced. T1–3 lacked mutations at sg1 site and contained C insertion at sg2 site. Its T2 progeny showed de novo mutations at sg1 site: single bp variation (insertion/deletion/substitution) and 6 bp deletion, whereas at sg2 site, both inheritances of + 1 C insertion and de novo single-base variations were observed. T1–4 contained − 1 T in sg1 site and + 1 C at sg2 target. Its T2 progeny, one of which lacked Cas9, inherited these mutations; however, new mutations were also observed: + 1 A and A–C substitution (Fig. 4b). All of these mutations were observed in the T1 parents; therefore, mutations at sg2 target were likely inherited, but de novo mutations were also created. Inheritance of mutation was confirmed in one T2 plant that contained – 1 and + 1 at the sg1 and sg2 sites, respectively (Fig. 4b). In summary, while genomic deletions remained undetectable, increased rate of point mutations (effect of single sgRNA) was observed in T1 and T2 progeny with single-base variation as the common type of mutation at the targeted site. We also investigated whether single-base variations frequently found at sg2 site could alone confer GUS negative phenotype as observed in T0 parent plant. We found that A–C substitution did not change the protein sequence, but + 1 A and + 1 C generated frame shift and early stop codon (data not shown), mutating the C-terminal catalytic domain of β-glucuronidase (GUS) enzyme (Wallace et al. 2010), leading to inactivation of GUS activity.

We also analyzed T1 progeny of Chalk5 T0 plants that showed chimeric effects at sg1 and sg2 sites by superimposed overlapping peaks downstream of the DSB site in the sequencing spectra. The analysis of the spectra by CRISP-ID tool identified short deletions at sg1 site and 1 bp insertions (+ 1) at sg2 site (Fig. 5). Thirty T1 plants from this chimeric parent were analyzed by PCR and sequencing. No deletion was evident, but point mutations at each site were found as homozygous or heterozygous mutation (Table 3; Fig. 5). Furthermore, at least one of the mutations identified in the parent plant (− 3 at sg1 and + 1 at sg2) was transmitted to the progeny at high rates.

Same mutation pattern from different targeting events

We frequently observed − 1 and/or + 1 mutations at GUS sg1 and sg2 sites in the targeted lines. To investigate whether the same type of mutation arises from different targeting events, we compared *GUS* sg1 and sg2 sites in 23 different lines obtained from 3 different experiments. At sg1 site, deletion of a single nucleotide (− 1) at the DSB site was observed 13 times (Fig. 6a), whereas at sg2, insertion of a single nucleotide (+ 1) at the DSB site was observed 12 times (Fig. 6b). The next most frequent type of mutation was single-base substitution (s1), which either occurred at the DSB site or in the PAM (Fig. 6a, b). Other types of mutations at the two sites included short deletions or single-nucleotide variations, which were generally observed once in the population. In summary, the repair of sg1 and sg2 DSB sites led to a predictable mutation pattern of − 1 or + 1 in ~ 50% of the transformed lines generated within the experiment or between experiments.

Discussion

Plant genome engineering involves a variety of genomic modifications including gene insertion, replacement, inactivation, or deletion. Creating predictable genetic variation is highly

desirable, but often defeated by the host repair processes that ignore DNA homologies and generate unpredictable mutations in higher plants (Jasin and Haber 2016; Puchta et al. 1996; Waterworth et al. 2011). As a result, targeted knockout is the most common outcome of genome editing. Genomic deletions, however, do not rely on homology-based DNA repair and, therefore, should be possible to create by standard gene-editing methods.

One of the applications of targeted genomic deletion is transgene excision to rid transgenic plant of antibiotic-resistance marker genes. While effective methods of transgene removal are available, they require specialized vector constructions, e.g., adding recombination sites or separating marker gene from the gene-of-interest in two T-DNAs (Gidoni et al. 2008; Komari et al. 1996; Wang et al. 2011). On the other hand, CRISPR/Cas9 can target loci by virtue of the cloned sgRNA spacers (Cong et al. 2013; Jinek et al. 2012; Mali et al. 2013), thereby, giving more flexibility to the user. Genomic deletion could also be pursued to create null mutations to allow detection by standard PCR, while screening of small indels would require mismatch cleavage assay, DNA sequencing, quantitative, or digital PCR (Belhaj et al. 2013; Falabella et al. 2017; Kim et al. 2009; Voytas 2013; Xie and Yang 2013). Genomic deletions could also create useful traits. The natural variant of rice *DEP1* harbors Δ625 bp that confers erect panicles and increased grain yield (Huang et al. 2009), and the spontaneous deletions in maize *WAXY* gene alter starch composition of the grains (Wessler et al. 1990). Genomic deletions also play major roles in plant evolution (De Smet et al. 2017; Soltis et al. 2014). Divergence in the function of the duplicated genes could occur upon deletions in the genes (Haberer et al. 2004; Liu et al. 2011). For example, deletions in the intergenic regions could either remove or change the position of *cis*-elements leading to altered tissue specificity and neo-

functionalization of the gene (Arsovski et al. 2015; De Smet and Van de Peer 2012). Thus, targeted genomic deletions could serve as useful effects in plant genome engineering. CRISPR/Cas9 has emerged as the dominant gene-editing tool that holds a great promise for genome engineering in plants and animals. This study evaluated the practical application of CRISPR/Cas9 in creating targeted genomic deletions in three loci in the rice genome. Previously, we reported successful deletion of *GUS* gene through dual-targeting by CRISPR/Cas9, which was accomplished by PCR screening and regeneration of the selected clones (Srivastava et al. 2017). Zhou et al. (2014) also reported chromosomal deletions in rice calli that were subjected to regeneration to recover plant lines. Similarly, in the present study, dual-targeting was successful in creating genomic deletion in trans- formed callus lines that mostly correlated with the efficiency of the sgRNA pairs. However, genomic deletions were rarely detected among plants transformed with Cas9: sgRNA constructs, and recovery of stable deletion lines was unsuccessful unless they were derived from calli harboring the deletion. This is somewhat surprising as point mutations by each sgRNA employed in dual-targeting occurred at high frequency, and the efficiency of the two sgRNAs used on two rice loci (*PDS* and *Chalk5*) was comparable. Furthermore, rate of point mutations in the two sites increased dramatically in the progeny, yet targeted deletions remained undetectable. Consistent with our study, others have also reported a much lower rate of genomic deletions by multiplex sgRNAs that is generally one order of magnitude lower than targeted point mutagenesis at two or more sites in the segment of the genome (Durr et al. 2018; Ordon et al. 2017). At the outset, these observations suggest that multiplex targeting by CRISPR/Cas9 occurs through non-concurrent activity on different sites as a result of dissimilar sgRNA efficiencies. Low rate of deletions in *GUS*, as observed in this study, could be based on dissimilar sg1 and sg2 efficiencies. However, genomic deletions in *PDS*

and *Chalk5* that were targeted by equally efficient sgRNA pairs were not proportionately increased. Therefore, understanding of the kinetics of Cas9-generated DSB could lend a mechanistic explanation. The Cas9: sgRNA complex stays bound to the broken termini of the DNA (Jiang and Doudna 2017; Stern- berg et al. 2014), which may prevent the free-fragment from being physically removed from the site. Subsequently, the free-fragment could participate in the NHEJ process and eventually be glued back to the genome. Thus, simultaneous DSBs end up with point mutations at each site rather than fragment deletion. Our dual-targeting data on three loci with highly variable efficiencies of sgRNA suggest that although sgRNA efficiency and Cas9 expression are important for the success of targeting, above a threshold, these parameters are unlikely to improve the rate of genomic deletions. Furthermore, DNA repair mechanisms in plants could affect the targeting outcome and enforce DSB repair by preserving broken termini and introducing only small indels, the most commonly observed effect of CRISPR/ Cas9 targeting in plants (Mao et al. 2013). Nevertheless, heritability of genomic deletions and other editing effects could be improved by expressing Cas9 by germline promoters (Durr et al. 2018; Feng et al. 2018). Finally, the survey of mutations in multiple transformed lines obtained from different experiments showed that the same type of mutation occurred frequently in the DSB sites. While sg1 site mostly lost a nucleotide (− 1), the sg2 site gained one (+ 1). The mechanistic explanation of this curious observation is not clear, but it implicates the role of target site and/or genomic context. More analysis with additional sgRNAs is needed to better understand the frequency of a given type of mutation in CRISPR/Cas9 targeting; however, similar observations have been made by Jacobs et al. (2015), who found identical mutation in multiple soybean lines. In a separate study based on targeting 10 loci in rice, + 1 was found to be the most common mutation (> 50%), followed by − 1 (Zhang et al. 2014). However, our data

suggest that a target site could also have the preference for either an insertion (+ 1) or a deletion (− 1).

In summary, consistent with a previous report on CRISPR/Cas9 targeting in rice (Jang et al. 2016), this study found that primary regenerated plants mostly harbor chimeric mutational effects. However, since the observed effects are generally not heritable, PCR screening at an early stage of callus growth, and isolation of the calli harboring the deletions will be an important step in recovering stable deletion lines. In addition, this study found that the types of mutations induced at a specific site by CRISPR/Cas9 are not highly variable, and frequently, the same type of mutation is observed from different targeting events. This observation suggests that DSB repair is highly dependent on the target sequence.

References

Arsovski AA, Pradinuk J, Guo XQ, Wang S, Adams KL (2015) Evolution of cis-regulatory elements and regulatory networks in duplicated genes of *Arabidopsis*. Plant Physiol. 169:2982-2991. DOI: 10.1104/pp.15.00717.

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V (2013) Plant genome editing made easy: targeted mutagenesis in modeland crop plants using the CRISPR/Cas system. Plant Methods. 9:39. DOI: 10.1186/1746-4811-9-39.

Brooks C, Nekrasov V, Lippman ZB, Van Eck J (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated 9 system. Plant Physiol. 166:1292-7. DOI: 10.1104/pp.114.247577.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science. 339:819-23. DOI: 10.1126/science.1231143.

De Smet R, Van de Peer Y (2012) Redundancy and rewiring of genetic networks following genome-wide duplication events. Curr. Opin. Plant Biol. 15:168–176. DOI: 10.1016/j.pbi.2012.01.003.

De Smet R, Sabaghian E, Li Z, Saeys Y, Van de Peer Y (2017) Functional divergence of duplicates in *Arabidopsis*. Plant Cell. DOI:10.1105/tpc.17.00531.

Dehairs J, Talebi A, Cherifi Y, Swinnen JV (2016) CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. Sci. Rep. 6: 28973. DOI: 10.1038/srep28973.

Durr J, Papareddy R, Nakajima K, Gutierrez-Marcos J (2018) Highly efficient heritable targeted deletions of gene clusters and non-coding regulatory regions in *Arabidopsis* using CRISPR/Cas9. Sci. Rep. 8:4443. DOI: 10.1038/s41598-018-22667-1.

Falabella M, Sun L, Barr J, Pena AZ, Kershaw EE, Gingras S, Goncharova EA, Kaufman BA (2017) Single-step qPCR and dPCR detection of diverse CRISPR-Cas9 gene editing events in vivo. G3 7:3533-3542. DOI:10.1534/g3.117.300123.

Feng C, Su H, Bai H, Wang R, Liu Y, Guo X, Liu C, Zhang J, Yuan J, Birchler JA, Han F (2018) High-efficiency genome editing using a dmc1 promoter-controlled CRISPR/Cas9 system in maize. Plant Biotechnol J. DOI: 10.1111/pbi.12920.

Gidoni D, Srivastava V, Carmi N (2008) Site-specific excisional recombination strategies for elimination of undesirable transgenes from crop plants. In Vitro Cell. Dev. Biol.-Plant 44: 457-467. https://doi.org/10.1007/s11627-008-9140-3.

Haberer G, Hindemitt T, Meyers BC, Mayer KF (2004) Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*. Plant Physiol. 136:3009-22. DOI: https://doi.org/10.1104/pp.104.046466.

Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009) Natural variation at the DEP1 locus enhances grain yield in rice. Nat. Genet. 41:494-7. DOI: 10.1038/ng.352.

Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA (2015) Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnol. 15:16. doi: 10.1186/s12896-015-0131-2.

Jang G, Lee S, Um TY, Chang SH, Lee HY, Chung PJ, Kim J-K, and Choi YD (2016) Genetic chimerism of CRISPR/Cas9-mediated rice mutants. Plant Biotechnol. Rep. 10: 425-435. https://doi.org/10.1007/s11816-016-0414-7.

Jasin M, Haber JE (2016) The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. DNA Repair 44: 6-16. https://doi.org/10.1016/j.dnarep.2016.05.001.

Jiang F, Doudna JA (2017) CRISPR–Cas9 structures and mechanisms. Ann. Rev. Biophys. 46:505-529. DOI: 10.1146/annurev-biophys-062215-010822.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA guided DNA endonuclease in adaptive bacterial immunity. Science 337:816-21. DOI: 10.1126/science.1225829.

Kapusi E, Corcuera-Gómez M, Melnik S, Stoger E (2017) Heritable genomic fragment deletions and small indels in the putative ENGase gene induced by CRISPR/Cas9 in barley. Front. Plant Sci. DOI: 10.3389/fpls.2017.00540.

Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res. 19:1279–1288. DOI: 10.1101/gr.089417.108.

Komari T, Hiei Y, Saito Y, Murai N, Kumashiro T (1996) Vectors carrying two separate T-DNAs for co-transformation of higher plants mediated by Agrobacterium tumefaciens and segregation of transformants free from selection markers. Plant J. 10:165-74. https://doi.org/10.1046/j.1365-313X.1996.10010165.x

Lemos BR, Kaplan AC, Bae JE, Ferrazzoli AE, Kuo J, Anand RP, Waterman DP, Haber JE (2018) CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. Proc. Natl. Acad. Sci. DOI:10.1073/pnas.1716855115

Liu S-L, Baute GJ, Adams KL (2011) Organ and cell type–Specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in *Arabidopsis* thaliana. Genom. Biol. Evol. 3:1419–1436.https://doi.org/10.1093/gbe/evr114.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA guided human genome engineering via Cas9. Science. 339:823-6. DOI:10.1126/science.1232033.

Mao Y, Zhang H, Xu N, Zhang B, Gou F, Zhu J-K (2013) Application of the CRISPR–Cas system for efficient genome engineering in plants. Mol. Plant. 6:2008-2011. DOI:10.1093/mp/sst121.

Minkenberg B, Xie K, Yang Y (2017) Discovery of rice essential genes by characterizing a CRISPR-edited mutation of closely related rice MAP kinase genes. Plant J. 89:636-48. DOI: 10.1111/tpj.13399.

Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiol.155:733-40. DOI: 10.1099/mic.0.023960-0.

Nandy S, Srivastava V (2012) Marker-free site-specific gene integration in rice based on the use of two recombination systems. Plant Biotech J. 10:904–912. DOI:10.1111/j.1467-7652.2012.00715.x.

Nekrasov V, Wang C, Win J, Lanz C, Weigel D, Kamoun S (2017) Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. Sci. Rep. 7:482. DOI: 10.1038/s41598-017-00578x.

Nishimura A, Aichi I, Matsuoka M (2006) A protocol for Agrobacterium mediated transformation in rice. Nat. Protoc 1:2796–2802. DOI:10.1038/nprot.2006.469.

Ordon J, Gantner J, Kemna J, Schwalgun L, Reschke M, Streubel J, Boch J, Stuttmann J (2017) Generation of chromosomal deletions in dicotyledonous plants employing a user-friendly genome editing toolkit. Plant J. DOI:10.1111/tpj.13319.

Park J, Bae S, Kim J-S (2015) Cas-Designer: A web-based tool for choice of CRISPR-Cas9 target sites. Bioinformatics. 31:4014-4016. DOI: 10.1093/bioinformatics/btv537.

Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc. Natl. Acad Sci USA. 93:5055-60. https://doi.org/10.1073/pnas.93.10.5055.

Rouet P, Smih F, Jasin M (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. Biol. 14:8096–8106.

Shi J, Gao H, Wang H, Lafitte HR, Archibald RL, Yang M, Hakimi SM, Mo H, Habben JE (2017) ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. Plant Biotech. J. 15:207-216. DOI: 10.1111/pbi.12603.

Soltis DE, Visger CJ, Soltis PS (2014) The polyploidy revolution then…and now: Stebbins revisited. Am. J. Bot. 101:1057-1078. https://doi.org/10.3732/ajb.1400178.

Srivastava V, Underwood JL, Zhao S (2017) Dual-targeting by CRISPR/Cas9 for precise excision of transgene from rice genome. Plant Cell Tiss. Organ Cul. 129:153 – 160. DOI:10.1007/s11240-016-1166-3.

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 507:62-7. DOI: 10.1038/nature13011.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33: 25-35.

Tian S, Jiang L, Gao Q, Zhang J, Zong M, Zhang H, Ren Y, Guo S, Gong G, Liu F, Xu Y (2017) Efficient CRISPR/Cas9-based gene knockout in watermelon. Plant Cell Rep 36: 399. https://doi.org/10.1007/s00299-016-2089-5.

van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, Reece-Hoyes JS, Nye C, Gradia S, Vidal B, Zheng J, Hoffman GR, Fuller CK, May AP (2016) DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. Mol Cell. 63:633-646. doi: 10.1016/j.molcel.2016.06.037.

Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. Annu. Rev. Plant Biol. 64:327-350. DOI: 10.1146/annurev-arplant-042811-105552.

Wallace BD, Wang H, Lane KT, Scott JE, Orans J, Koo JS, Venkatesh M, Jobin C, Yeh LA, Mani S, Redinbo MR (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. Science.330:831-35. DOI: 10.1126/science.1191175

Wang Y, Yau Y-Y, Perkins-Balding D, Thomson JG (2011) Recombinase technology: applications and possibilities. Plant Cell Rep. 30:267-285. DOI:10.1007/s00299-010-0938-1.

Wang Y, Geng L, Yuan M, Wei J, Jin C, Li M, Yu K, Zhang Y, Jin H, Wang E, Chai Z (2017a) Deletion of a target gene in Indica rice via CRISPR/Cas9. Plant Cell Rep. 36:1333-43. DOI: 10.1007/s00299-017-2158-4.

Wang X, Tu M, Wang D, Liu J, Li Y, Li Z, Wang Y, Wang X (2017b) CRISPR/Cas9-mediated efficient targeted mutagenesis in grape in the first generation. Plant Biotech. J. 16:844-855. DOI: 10.1111/pbi.12832.

Wang Z, Wang S, Li D, Zhang Q, Li L, Zhong C, Liu Y, Huang H (2018) Optimized paired-sgRNA/Cas9 cloning and expression cassette triggers high-efficiency multiplex genome editing in kiwifruit. Plant Biotech. J. DOI: 10.1111/pbi.12884.

Waterworth WM, Drury GE, Bray CM, West CE (2011) Repairing breaks in the plant genome: the importance of keeping it together. New Phytol. 192:805-22. DOI: 10.1111/j.1469-8137.2011.03926.x.

Wessler S, Tarpley A, Purugganan M, Spell M, Okagaki R (1990) Filler DNA is associated with spontaneous deletions in maize. Proc. Natl. Acad Sci USA 87:8731-5.

Xie K, Yang Y (2013) RNA-guided genome editing in plants using a CRISPR-Cas system. Mol. Plant 6:1975-1983. DOI:10.1093/mp/sst119.

Xie K, Minkenberg B, Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. USA 112:3570–3575. DOI:10.1073/pnas.1420294112.

Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, Mao Y, Yang L, Zhang H, Xu N, Zhu JK (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol J. 2014 Aug;12(6):797-807. doi: 10.1111/pbi.12200

Zhou H, Liu B, Weeks DP, Spalding MH, Yang B (2014) Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucl. Acids Res. 42:10903–10914. DOI:10.1093/nar/gku806.

Tables and Figures

Table 1: Genomic deletion by dual-targeting in callus lines

Exp.	Target Gene	Vector	Predicted Δ size (bp)	Total lines	PCR detection	DNA sequencing		[1]Eff. (%)
						(-) InDel	(+) InDel	
1	*OsPDS*	pJU34	985	32	2	-	2	6.2
2	*OsChalk5*	pJU46	240	53	4	2	2	7.5

[1]Percent events showing genomic deletion by PCR as shown in Fig. 1. *GUS* deletion data is given in Srivastava et al. 2017.

Table 2: Point-mutations in primary transgenic (T0) plants

Exp.	Target	Total no. of plants	Non-targeted	[1]No. of plants targeted		[2]Eff. (%)	
				sg1 site	sg2 site	sg1	sg2
1	*GUS*	21	12	0	9	-	42
2	*OsPDS*	12	3	9	9	75	75
3	*OsChalk5*	17	6	11	11	64	64

[1]Generally chimeric mutations observed. Types of mutations shown in Fig. 3
[2]Percent plants harboring mostly chimeric mutations at predicted DSB sites.

Table 3: Point-mutations in *GUS*-CRISPR/Cas9 progeny

Locus	Gener-ation	No. of plants tested	[1]sg1 mutations				[1]sg2 mutations			
			Non-targeted	Mono-allelic	[2]Bi-allelic	[3]Chimer.	Non-targeted	Mono-allelic	[2]Bi-allelic	[3]Chimer.
GUS	T1	17	9	6	-	2	-	12	5	-
GUS	T2	17	8	7	1	1	-	10	7	-
[4]Chalk5	T1	30	0	0	30	-	7	8	15	-

[1]Types of mutations shown in Fig. 4-5.
[2]Heterozygous or homozygous
[3]Presence of >2 overlapping traces downstream of DSB site in the sequencing spectra
[4]T1 plants of *Chalk5* from potentially same transgenic event but different T0 plants

Table 4: Point-mutations in *Chalk5*-CRISPR/Cas9 progeny

Generation	[1]No. of plants tested	[2]sg1 mutations			[2]sg2 mutations		
		Non-targeted	Mono-allelic	[3]Bi-allelic	Non-targeted	Mono-allelic	[3]Bi-allelic
T1	30	0	0	30	7	8	15

[1]T1 plants from potentially same transgenic event but different T0 plants.
[2]Types of mutations shown in Fig. 5
[3]Heterozygous or homozygous

Table 5: Primers used in the study

Primer	Sequence (*5' – 3'*)	Application
gGus1F	TAGGTCTCCTGATCAGCGTTGGgttttagagctagaa	Construction of sgRNA1 *(GUS*): *5*'-GTGGAATTGATCAGCGTTGG-*3'*
gGus1R	CGGGTCTCAATCAATTCCACtgcaccagccggg	
gGus2F	TAGGTCTCCCCGCAAACCGAAGTgttttagagctagaa	Construction of sgRNA2 (*GUS*): *5'*-ACGCGACCGCAAACCGAAGT-*3*'
gGus2R	CGGGTCTCAGCGGTCGCGTtgcaccagccggg	
gPDS1F	TAGGTCTCCCAGGAGAATTCAGCgttttagagctagaa	Construction of sgRNA1 (*OsPDS*): *5'*-ACAAGCCAGGAGAATTCAGC-*3'*
gPDS1R	CGGGTCTCACCTGGCTTGTtgcaccagccggg	
gPDS2F	TAGGTCTCCATGGATAACTCATCgttttagagctagaa	Construction of sgRNA2 (*OsPDS*): *5*'-CACTGCATGGATAACTCATC-*3'*
gPDS2R	CGGGTCTCACCATGCAGTGtgcaccagccggg	
gChalk1F	TAGGTCTCCTCATGTAGATCTTgttttagagctagaa	Construction of sgRNA1 (*OsChalk5*): *5'*-CTCTGGGTCATGTAGATCTT-*3'*
gChalk1R	CGGGTCTCAATGACCCAGAGtgcaccagccggg	
gChalk2F	TAGGTCTCCGACGGTGCCGTTTGTAGgttttagagctagaa	Construction of sgRNA2 (*OsChalk5*): *5'*-GATTCTAGACGGTGCCGTTTGTAG-*3'*
gChalk2R	CGGGTCTCACGTCTAGAATCtgcaccagccggg	
Ubi1812	TCTAACCTTGAGTACCTATCTATTA	Forward primer in B1 locus
NosR2	GCGGGACTCTAATCATAAAAACCC	Reverse primer in B1 locus
PDSF	GGTAGAAATGCCATGCGGGA	Forward primer in *OsPDS*
PDSR	GTGGTGAGGTTCGGCTGAAT	Reverse primer in *OsPDS*
Chalk5F	ACAAGGCTAGCAAGTTGGC	Forward primer in *OsChalk5*
Chalk5R	CACTCGCTCGTCTTCTCCTC	Reverse primer in *OsChalk5*
Cas9F	AAAGACCGAGGTGCAGACAG	Forward primer in Cas9
Cas9R	ACCAGCACAGAATAGGCCAC	Reverse primer in Cas9

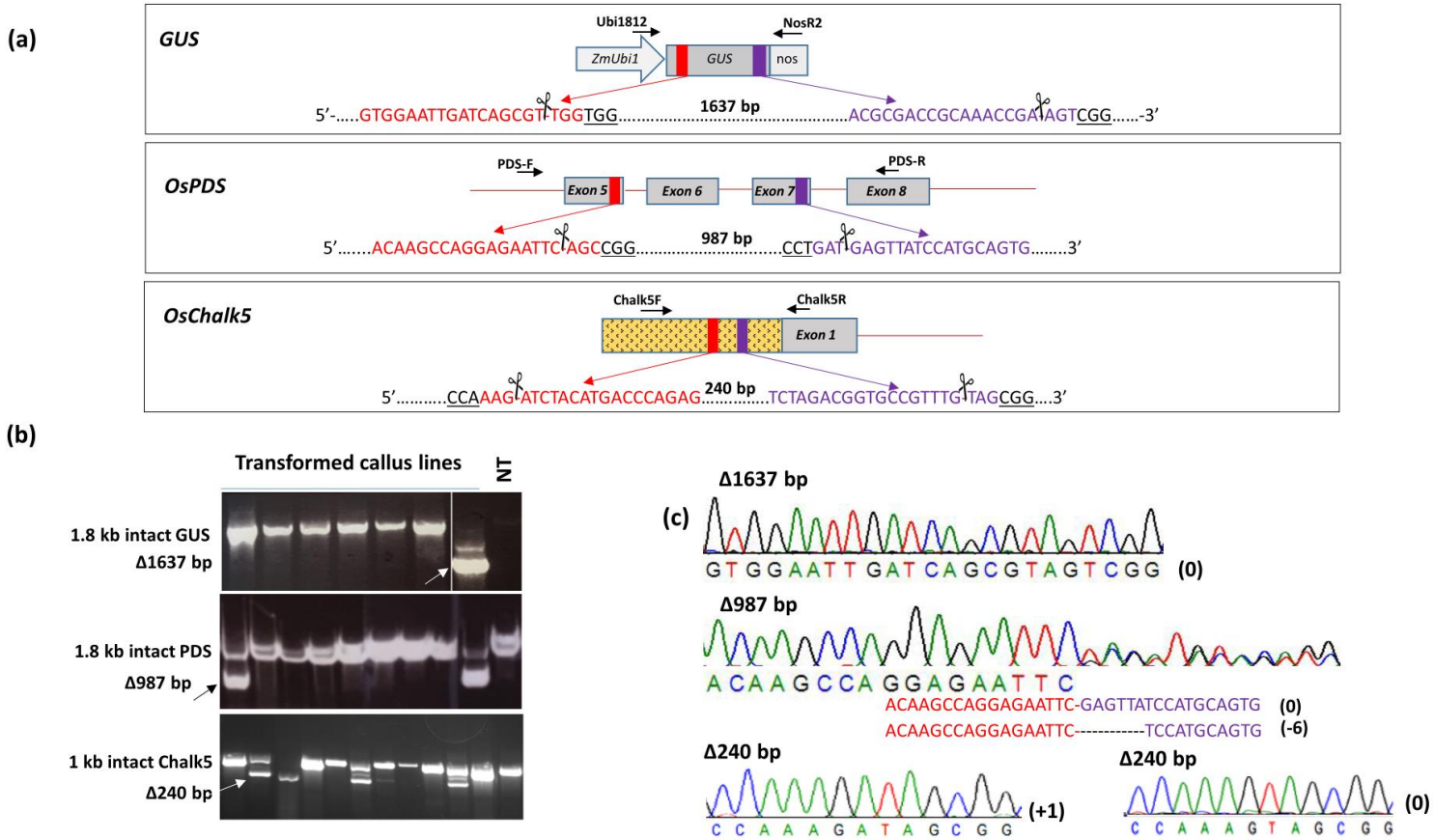


Fig. 1: Dual-targeting by CRISPR/Cas9 for fragment deletions. **a** Paired sgRNAs for targeting three genes, transgene *GUS* and native genes, *OsPDS* and Os*Chalk5*, in rice. Full structure of *GUS* gene and partial structures of *OsPDS* and *OsChalk5* genes are shown with sgRNA (red and purple boxes) and primer (arrows) locations. sgRNA spacer 1 (red) or sgRNA spacer 2 (purple) for each locus are shown with protospacer adjacent motif (PAM) (underlined). The positions of double-stranded break (DSB) sites are shown by scissors that defined deletion sizes given in base pairs (bp). *ZmUbi* refers to maize Ubiquitin-1 promoter and *nos* to nopaline synthase *3'* transcription terminator. *GUS* and *OsPDS* genes are targeted in the genic regions (exons), while *OsChalk5* in the intergenic region, upstream of promoter harboring *cis*-elements (white box). b PCR screening of callus clones using forward and reverse primers spanning targeted sites (see Table 1; a). Representative callus lines are shown with non-transgenic controls (NT; cv.

Figure 1 (contd…)

Nipponbare). The intact and the deletion fragments (Δ) are indicated; c Sequences of the representative deletion fragments of GUS (Δ1637 bp), PDS (Δ987 bp), and Chalk5 (Δ240 bp) loci. The number of bases representing insertion–deletions (indels) is given in parentheses.

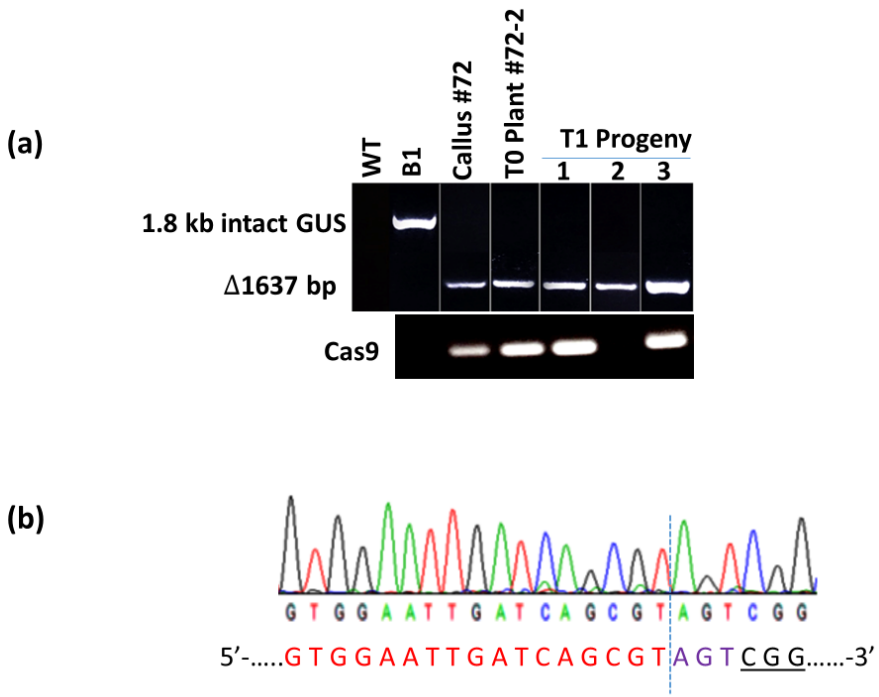


Fig. 2: Recovery of stable plant lines harboring Δ1637 bp *GUS* deletion. **a** PCR analysis to detect *GUS* and Cas9 in the callus, primary transgenic plant (T0), and the progeny (T1). WT, wild-type Nipponbare; B1, transgenic *GUS* line; **b** DNA sequencing spectrum ofΔ1637 bp fragment in T0 plant#72-2 generated by the paired used of sgRNAs. The observed sequence matches the predicted deletion site derived from joining of distal ends without indels. Dashed vertical line indicates blunt DSB ligation.

Fig. 3: Types of mutations observed in T0 plants. Sequence alignments of *GUS*, *PDS* and *Chalk5* sequences at sg1 and sg2 targeted sites (yellow highlights). PAM sequences are underlined, and DSB site is shown as (−) in each reference sequence. Insertion/deletions/ substitutions for each site are shown on the right. Deletions are shown as red dashes, insertions as small red letters, and substitutions as large blue letters.

(a)

Fig. 4a: For legends, please see page 124

(b)

Fig. 4b: Genotyping of progeny plants derived from the T0 parent expressing *GUS*-targeting vector. **a** T1 progeny, and **b** T2 progeny. The mutation types in sg1 and sg2 targets are shown, see Fig. 3 for notations. Bold T1/T2 lines are Cas9-negative. Parent plants are underlined with their representative progeny given below.

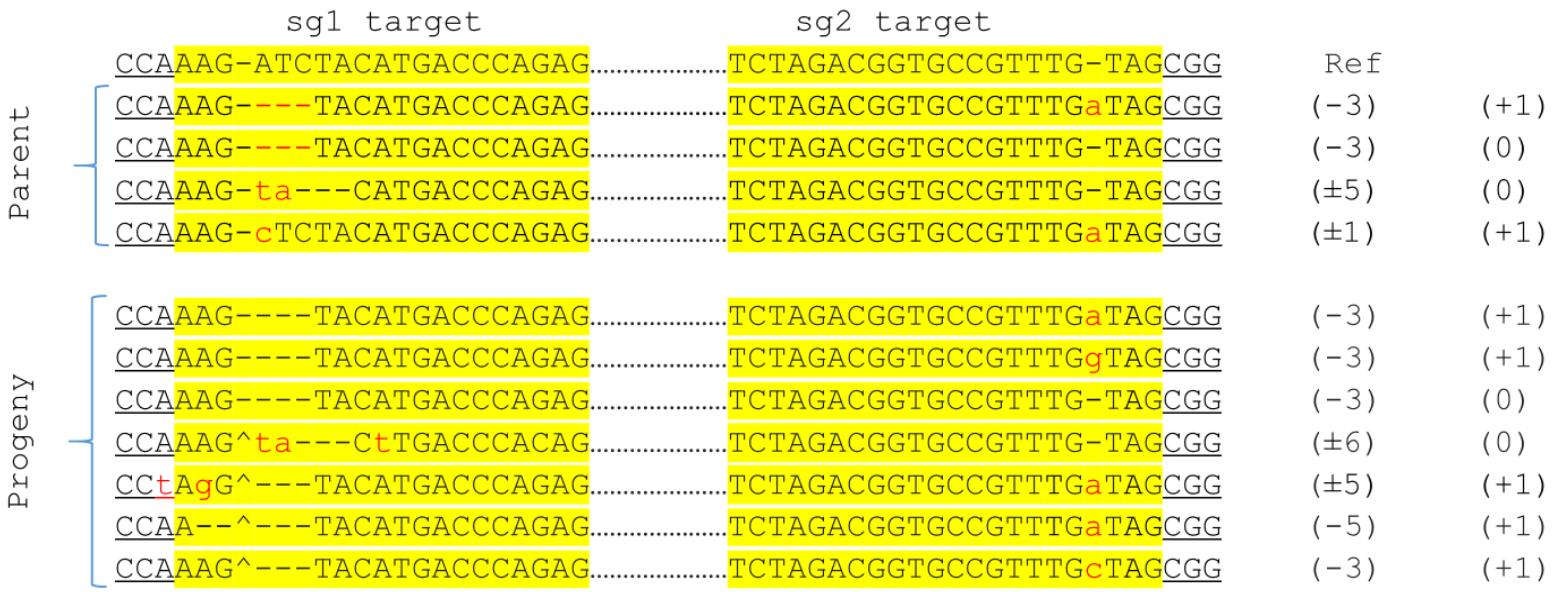


Fig. 5: Genotyping of progeny plants derived from the T0 parent expressing *Chalk5*-targeting vector. The mutation types in sg1 and sg2 targets in the parent and progeny plants are aligned with the reference, see Fig. 3 for notations.

Fig.6: Frequency of mutations observed at *GUS* targets as determined by Sanger sequencing of the sg1 target (**a**) and sg2 target (**b**). The reference sequences with PAM (underlined) and DSB site (−) are shown on the top. Insertions (+) and deletions (−) are shown in red and substitutions (s) in blue fonts. s1 refers to single-nucleotide substitution at or near DSB site. Frequency refers to number of times a mutation type observed among the 23 lines. Boxed numbers indicate most common mutation types (− 1 or + 1) and their frequency.

CHAPTER V

EVALUATION OF CRISPR/CAS9 IN GENERATING TARGETED MUTATIONS BY INDUCIBLE EXPRESSION OF CAS9 IN THE RICE GENOME

Summary

Transient expression of CRISPR/Cas9 is an effective approach for limiting its activities and improving its precision in genome editing. Here, we describe the heat-shock- inducible CRISPR/Cas9 for controlled genome editing, and demonstrate its efficiency in the model crop, rice. Using the soybean heat-shock protein gene promoter and the rice *U3* promoter to express Cas9 and sgRNA, respectively, we developed the heat-shock (HS) -inducible CRISPR/Cas9 system, and tested its efficacy in targeted mutagenesis. Two loci were targeted in rice, and the presence of targeted mutations was determined before and after the HS treatment. Only a low rate of targeted mutagenesis was detected before HS (~16%), but an increased rate of mutagenesis was observed after the HS treatment among the transgenic lines (50–63%). Analysis of regenerated plants harboring HS-CRISPR/Cas9 revealed that targeted mutagenesis was suppressed in the plants but induced by HS, which was detectable by Sanger sequencing after a few weeks of HS treatments. Most importantly, the HS-induced mutations were transmitted to the progeny at a high rate, generating monoallelic and biallelic mutations that independently segregated from the *Cas9* gene. Additionally, off-target mutations were either undetectable or found at a lower rate in HS-CRISPR/ Cas9 lines as compared to the constitutive-overexpression CRISPR/Cas9 lines. Taken together, this work shows that HS-CRISPR/Cas9 is a controlled and reasonably efficient platform for genome editing, and therefore, a promising tool for limiting genome-wide off-target effects and improving the precision of genome editing.

Introduction

The CRISPR/Cas9 system is an efficient tool for genome editing that is gaining popularity in both agricultural and medical biotechnology. It consists of two components: the Cas9 nuclease and a single-guide RNA (sgRNA) that forms a complex (sgRNA:Cas9) and targets sequences complementary to ~20 nt spacer sequence in sgRNA, provided the NGG protospacer adjacent motif (PAM) is located at the *3′* end of the target sequence. Successful targeting by Cas9 results in a blunt double-stranded break (DSB), 3-nt upstream of the NGG motif (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013; Mojica, Díez-Villaseñor, García-Martínez, & Almendros, 2009), the repair of which by the cell leads to gene editing effects such as insertion-deletions (indels) and gene replacement (Jasin & Haber, 2016; Puchta, Dujon, & Hohn, 1996; Rouet, Smih, & Jasin, 1994; Szostak, Orr-Weaver, Rothstein, & Stahl, 1983; Waterworth, Drury, Bray, & West, 2011). Similarly, CRISPR/Cas12a, an alternative gene editing tool, can be deployed on sequences ending with TTTN motifs (Endo, Masafumi, Kaya, & Toki, 2016; Schindele, Wolter, & Puchta, 2018; Wang, Mao, Lu, Tao, & Zhu, 2017; Zetsche et al., 2015).

To improve the gene editing efficiency, different approaches including sgRNA designs or Cas9 expression systems have been described that mostly include developmental and constitutive gene promoters (Feng et al., 2018; Hu, Meng, Liu, Li, & Wang, 2018; Ma, Zhu, Chen, & Liu, 2016; Miki, Zhang, Zeng, Feng, & Zhu, 2018; Wang et al., 2015). In monocots, rice and maize ubiquitin promoters for Cas9 expression and the *U3* or *U6* promoter for sgRNA expression are quite successful in creating targeted effects in the primary transformed (T0) plants (Lee et al., 2018; Wang et al., 2014; Xie & Yang, 2013). Previous studies have also shown that CRISPR/Cas9 effects could occur at a high rate during tissue culture or regeneration phases,

leading to edited T0 lines that efficiently transmit the mutations to the next generation (Mikami, Toki, & Endo, 2015; Srivastava, Underwood, & Zhao, 2017; Zhang, Zhang, Wei, et al., 2014). However, in these approaches, the strong doses of sgRNA: Cas9 could persist far beyond the incidence of targeted gene editing, and provide a wider opportunity to mutagenize the genome-wide off-target sites. Accordingly, off-targeting was found to be higher with the higher doses of sgRNA:Cas9 in human cells, and ~100× higher with constitutive-Cas9 as compared to the transient-Cas9 in maize cells, as well as in the rice plants expressing constitutive-Cas9 (Hsu et al., 2013; Hu et al., 2018; Pattanayak et al., 2013; Svitashev et al., 2015). The dose of the sgRNA:Cas9 complex determines targeting efficiency; however, since mismatches between the sgRNA spacer sequence and the target genomic sites are allowed at the PAM-distal end (Fu et al., 2013; Jinek et al., 2012; Lin et al., 2014; Liu et al., 2016), each sgRNA could potentially target numerous off-sites in the genome. Although, off-sites would generally be targeted at lower rates than the bona fide target site, constitutive or tissue-specific expression systems would be more permissive to the off-site mutations by providing strong doses of Cas9 for a longer than necessary period of time.

Off-target effects of CRISPR/Cas9 are topic of intense investigation as it can induce high-frequency mutations at unintended off-target sites. Although, genetic segregation is an option for removing such mutations in many plant species, curbing off-target effects will be a better approach for developing high-quality edited lines. Restricted expression of the Cas9 can minimize the off-target effects while inducing high-efficiency on-target mutations. Several approaches for improving the precision of gene editing have been described, for example, high fidelity Cas9, split-Cas9, and ribonucleoprotein (RNP) Cas9 (Kleinstiver et al., 2016; Liang et al., 2017; Murovec, Guček, Bohanec, Avbelj, & Jerala, 2018; Senturk et al., 2017; Svitashev,

Schwartz, Lenderts, Young, & Cigan, 2016; Wright et al., 2015). The use of RNPs has additional benefits in plant biotechnology as this DNA-free approach generates targeted mutations without incorporating the foreign genes (Wolt, Wang, Sashital, & Lawrence-Dill, 2016; Wolter & Puchta, 2017). However, RNP approach in plants is faced with the difficulty of delivering the reagent in the cell wall bound compartments, and recovering the edited lines without selection in the tissue culture.

Here, we describe the use of the inducible expression system for controlling CRISPR/Cas9 mutagenesis. Our rationale is to generate short phases of Cas9 expression in the tissue culture or the regenerated plants for allowing targeted genome editing but keeping the Cas9 suppressed at most other times until genetic segregation. In addition to helping reduce off-target effects, this temporal control on Cas9 could improve gene editing efficiencies by inducing Cas9 in the phases conducive to gene editing, for example, plant regeneration phase in the tissue culture (Srivastava et al., 2017; Zhang, Zhang, Wei, et al., 2014), and enable conditional targeting to avoid lethal effects of mutations.

Using the heat-shock-inducible promoter to express Cas9 and the rice *U3* promoter for sgRNAs, we developed transformed lines of rice that essentially contained heat-shock (HS)-controlled CRISPR/ Cas9 system. By targeting genomic loci with a paired sgRNA, we determined the efficacy and efficiency of HS-CRISPR/Cas9 system in rice. Our analysis indicates that HS-CRISPR/Cas9 rarely induced mutations at the ambient room temperatures but efficiently created mutations upon the heat-shock treatment in the callus and the regenerated plants. Notably, targeted mutations were transmitted to the progeny at a high rate and segregated independently from the *Cas9* gene. In comparison with strong constitutive expression system consisting of the rice Ubiquitin promoter (RUBI) to express Cas9 (Xie, Minkenberg, & Yang,

2015), HS-CRISPR/Cas9 created mutations at ≥50% rate. More importantly, a comparative analysis of the predicated off-target sites of the designed sgRNAs using the Sanger sequencing showed a higher rate of off-targeting under constitutive expression system (RUBI), and undetectable and or a lower rate of off-targeting in the inducible expression system (HS). Overall, this study shows that HS-CRISPR/Cas9 is a more precise and efficient system for creating targeted mutagenesis, and therefore, a promising platform of improving gene editing that would be less permissive to off-target effects.

Materials and Methods

DNA constructs and plant transformation

The *Cas9* coding sequence was PCR amplified from pRGE32 (Addgene #63159) using primers (Table S8) laced with specific restriction enzyme sites and cloned between the soybean *HSP17.5E* gene promoter (GenBank accession no. M28070) and the nopaline synthase terminator (*nos 3′*) in the pUC19 vector backbone. The sgRNA vectors were made in pRGE32 backbone using the protocol of Xie et al. (2015) and the sgRNA spacer sequences were selected using the CRISPR RGEN tool (http://www.rgenome.net/cas-de- signer; Park, Bae, & Kim, 2015). The resulting *GUS* (GenBank accession no. AF485783) and *OsPDS* (Os03g08570) sgRNA constructs were PCR amplified with primers shown in Table S8 and cloned into a vector harboring the 35S promoter driven hygromycin phospho- transferase (*HPT*) gene. All vectors were verified by sequencing. The B1 transgenic line (*cv*. Nipponbare), which has been described by Nandy and Srivastava (2012) or wild type Nipponabare was used for transformation. B1 contains a single-copy of *GUS* gene controlled by the maize ubiquitin-1 gene promoter. The GUS activity was verified by staining endosperms using the GUS staining solution described by Jefferson (1987). The embryogenic callus obtained from the mature seeds of the homozygous B1

line was used for all transformations. All transformations were done by the gene gun (PDS1000, Bio-Rad Inc.)-based DNA delivery of the Cas9 and the sgRNA vectors (Fig. 1a). The transformed calli were isolated on the hygromycin (50 mg/L) containing media. All tissue culture and regeneration in this study were done using the method of Nishimura, Aichi, and Matsuoka (2006).

Heat-shock treatments

Freshly plated calli, rooted regenerated plants in the glass tubes or ~1-week-old seedlings on MS/2 plates were subjected to the heat-shock (HS) treatment by transferring them to preheated 42°C incubator. The Petri dishes containing the calli or germinating seedlings were laid on their sides between the preheated metal plates, whereas, regenerated plants in the glass tubes were submerged in 42°C water bath. After 3 h, plates or tubes were returned to the tissue culture chamber set at 25°C for further growth. Tissues were harvested after a few days for genotyping by PCR and sequencing.

DNA extraction, PCR, and sequencing

Genomic DNA isolated from callus, regenerated plants or seedlings was used for the polymerase chain reaction (PCR) using primers spanning the target sites (Table S8) or the predicted off- target sites (Table S9). PCR products were resolved on the agarose gel and extracted using GeneJET Gel Extraction Kit (Thermo Scientific, USA) for sequencing from both ends using the forward and the reverse primers by the Sanger Sequencing method at Eurofins Genomics USA (www.eurofinsgenomics.com). Selected PCR amplicons were cloned into pCR2.1 vector using the TA cloning kit (Thermo-Fisher Scientific, NY) as per the manufacturer's instructions. Randomly picked 15 to 20 colonies were verified for the insert by PCR using the amplicon-specific primers and sequenced at Eurofins Genomics USA. The

sequence traces (ABI files) were analyzed on the Sequence Scanner 2 software (Applied Biosystems Inc.) and aligned with the reference sequences using the CLUSTAL-Omega multiple sequence alignment tool. The over- lapping sequence traces arising from heterozygous alleles or chimeric samples were separated using the CRISP-ID tool (Dehairs, Talebi, Cherifi, & Swinnen, 2016).

Gene expression analysis

Young developing leaves were collected from the same tiller and incubated at the room temperature (25°C) or 42°C for 3 h for the control and the heat-shock treatments, respectively. The total RNA was isolated from 100 mg samples using the QIAGEN RNeasy plant mini kit (Qiagen, Valencia, CA), and treated with RNase-Free RQ1 DNase (Promega, San Luis Obispo, CA), and quantified using NanoDrop 2000 (Thermo Fisher Scientific, NY). The expression analysis on Cas9 and sgRNAs was performed on 25 ng of RNA using Superscript III Platinum SYBR green one step qRT-PCR (Life Technologies, Grand Island, NY) in the CFX96 Real-Time PCR Detection system (Bio-Rad, Hercules, CA). The values were normalized against the rice ubiquitin gene, and the relative expression to the non-transgenic control was calculated using the 2ΔΔCt (Livak & Schmittgen, 2001) method. Standard errors of two to six biological replicates were calculated. Each biological replicate was repeated two times for the analysis. Student *t*-test (unpaired) was used to determine the p-value. Primers used in qRT-PCR are given in Table S8.

Off-target analysis

Potential off-target sites (OT) for the designed sgRNAs of *GUS* and *PDS* genes were searched using the GGGenome (https://gggenome. dbcls.jp/, Naito, Hino, Bono, & Ui-Tei, 2015) and the CCTOP (https:// crispr.cos.uni-heidelberg.de/; Stemmer, Thumberger, del Sol Keyer, Wittbrodt, & Mateo, 2015) programs with the search queries of 20nt, 12nt seed sequences and ≤4

mismatches. A total of 26 sites for the *GUS* and 30 sites for the *PDS* were shortlisted. The BLAST analysis on all of the 56 sites was performed in the Plant Ensembl and NCBI against *Oryza sativa* Japonica IRGSP 1.0 to verify the sequences and locate their positions (*i.e.* intergenic or genic). Based on (i) the sequence homology across the genome and (ii) the presence/absence of SNPs and/or indels at the off-target and its surrounding primer designing area; 14 sites for *GUS* and 15 sites for *PDS* sgRNAs were selected for the analysis. The primers flanking the off-target sites were designed using the Primer Quest tool (https:// www.idtdna.com/PrimerQuest/). The primer sequences are shown in Table S9. The PCR was first performed on the negative controls; the WT Nipponbare (*PDS*) and the B1 line (Nipponbare) (*GUS*) and were sequenced by the Sanger method. All the samples were sequenced at Eurofins Genomics USA. The sequence traces were analyzed on Sequence Scanner 2 and aligned with the negative control sequences and the chromosomal reference using the Clustal Omega and t-coffee multiple sequence alignment tools. The overlapping sequences arising from the heterozygous or chimeric samples were separated using the CRISP-ID (Dehairs et al., 2016) and Polypeak Parser tools (Hill et al., 2014).

Results

Heat-shock-induced CRISPR/Cas9 mutagenesis in the rice in vitro tissue

We used the soybean heat-shock protein 17.5E (*HSP17.5E*) gene promoter to express the humanized *Streptococcus pyogenes* Cas9 (SpCas9), and the tRNA-processing system to express two sgRNAs by the rice snoRNA *U3* promoter (Czarnecka, Ingersoll, & Gurley, 1992; Xie et al., 2015; Fig. 1a,b). The motivation to use *HSP17.5E* promoter was based on its observed efficacy in controlling the Cre-*lox* recombination in the tissue culture-derived rice plants and seedlings. Earlier, we showed that a simple heat treatment of 42°C for 3 h led to efficient Cre-*lox*-mediated

excision of the marker gene in rice seedlings and inheritance of the marker-free locus by their progeny (Nandy & Srivastava, 2012). We chose previously tested target loci and sgRNAs for this study that include rice *Phytoene Desaturase* gene (*OsPDS*) and the *β-Glucuronidase* transgene inserted in the rice genome (Srivastava et al., 2017). For *GUS* targeting, a well-characterized transgenic line, B1 (cv. Nipponbare), that harbors a single-copy of the *GUS* gene driven by the maize ubiquitin promoter (*Ubi*), and for *PDS* targeting, non-transgenic Nipponbare was transformed. The resulting hygromycin-resistant calli were maintained and regenerated at the ambient room temperature. For testing HS-CRISPR/Cas9 activity, randomly sampled calli were either kept at the room temperature (pre-HS) or transferred to the fresh media plate for heat-shock treatment, and analyzed 5–7 days later (post-HS). A total of 23 PDS and 12 GUS calli were screened for mutations at the two sgRNA sites (Table 1). Two out of the 12 pre-HS PDS calli were found to contain the targeted mutations, one of which contained monoallelic mutation at both sg sites, while the other showed biallelic heterozygous mutation at the sg2 site (Table S1). Similarly, one of the 6 pre-HS GUS samples showed mutations (monoallelic) at the sg1 target (Table 1; Table S2). The pre-HS mutations could be derived from the leaky HS-Cas9 activity and established early in the selection of the transformed clones. Accordingly, characteristic overlapping dual traces were observed in the pre-HS samples, representing heterozygous or chimeric clones (Figs 1c, d, 2a, b). Next, the calli were subjected to heat-shock (HS) treatment for 3 h and returned to ambient room temperature for further growth. After 5–7 days (post-HS), freshly grown tissue from each callus culture was analyzed. Since calli could contain multiple independent mutations, HS-induced targeting could contain multiple overlapping traces in the Sanger sequencing spectra downstream of the predicted DSB sites (Fig. 1c, d). Further, if induced mutations are rare in the post-HS samples, they would appear as the

minor trace in the sequencing spectra (Fig. 2a, b). Accordingly, overlapping and/or minor traces in the sequencing spectra were found in 7 PDS and 3 GUS calli, indicating mosaic pattern of mutations due to HS-CRISPR/Cas9 activity (Table 1; Tables S1–S2). Mosaic pattern was observed at *PDS* sg1 site in 3 samples and at PDS sg2 site in 7 samples (Table S1). Similarly, mosaic pattern in GUS samples occurred once in the sg1 site and three times in the GUS sg2 site (Table S2). In summary, HS-CRISPR/Cas9 was effective in creating targeted mutations with a higher rate of targeting in post-HS calli (50–63%) as compared to the pre-HS calli (16%) of rice (Table 1). To verify these mutations, traces were separated using the CRISP-ID tool or subjected to TA cloning and colony sequencing. These analyses revealed indels at the predicted DSB sites, indicating CRISPR/Cas9 mediated mutagenesis (Fig. 1e–f, 2c, d). In conclusion, *HSP17.5E*-Cas9 is effective in creating induced targeted mutations in the rice calli. With the paired sgRNAs, HS-CRISPR/Cas9 generated HS-induced mutations in ≥50% of the transformants (Table 1). All callus cultures were subjected to plant regeneration; however, PDS cultures mostly appeared non-embryogenic, while GUS cultures regenerated plants. Therefore, all subsequent work was done with HS-CRISPR/Cas9 targeting the GUS transgene.

Heat-shock-induced targeting in T0 plants

Twenty regenerated plants (T0) expressing HS-CRISPR/Cas9 against the *GUS* gene were obtained from two experiments. At the rooting stage, 1–3 leaf samples from each were subjected to PCR and Sanger sequencing at the targeted sites. Two of the T0 plants (#9 and #12) were found to harbor homozygous or heterozygous mutations at the sg2 target, indicating leaky pre-HS Cas9 expression in these plants (Fig. 3). The rest did not show mutations at either site (Table 2). Next, T0 plants were given two rounds of HS treatment by transferring them to 42°C incubator for 3 h and repeating the treatment after ~20 h of rest at the room temperature. The HS

plants were subsequently transplanted in the soil and grown in the greenhouse. After ~4 weeks of HS treatment, at the young vegetative stage, target site analysis by PCR and sequencing was conducted in 2–3 leaf samples. No detectable targeting was found in any of the samples except those derived from T0#9 and #12; although, a baseline secondary sequence was detected in the sequencing spectra of a few lines, indicating a low rate of HS-induced mutations (Table 2). T0#1 and #3 showed a clear WT sg1 target in the young plants but minor targeting, indicated by the secondary baseline sequence trace, in the flowering plant. At the sg2 target, on the other hand, these plants showed minor targeting in the young plants, but monoallelic targeting in the flowering plants (Fig. 4a,b). Similar mixed traces were observed in the other post-HS samples of different T0 plants (Fig. S1). These observations corroborated with histochemical GUS staining as these plants progressively lost GUS activity. For example, T0#1 showed strong GUS staining in the leaf cuttings taken from the young vegetative plant but diminished staining in the leaves collected from the flowering plant (Fig. 5a; Table 2). Similarly, T0#3 progressively lost GUS activity, while T0#2 that lacked detectable mutations continued to show strong GUS staining, and T0#9 and #12 that harbored biallelic mutations also did not display GUS staining in the leaves derived from the vegetative or flowering stages of the plant (Table 2; Fig. S2). These observations are analogous to our work with HS Cre-*lox* system, in which, rice seedlings harboring HS Cre showed progressive recombination in the heat-shocked plants, and transmitted the recombined locus to the next generation (Nandy & Srivastava, 2012). Taken together, HS-induced gene editing effects likely occurred in the early cell lineages and established in the plant through cell division.

T0 plants # 1, # 2, and # 3 flowered and set seeds. These plants were analyzed at the flowering stage (>12 weeks post-HS) for the presence of mutations at the target sites. As shown

in Fig. 4a, b, T0 #1 and #3 showed rare targeting at the sg1 site but a clear monoallelic targeting at the sg2 site. Since, a low rate of mutagenesis at sg2 was detected in these plants at the young vegetative stage (baseline minor trace in the spectra) (Fig. 4a,b), these monoallelic mutations were likely induced early in the plant. Both plants contained a characteristic + 1 mutation at the predicted DSB site. T0#2, however, did not show mutations in any of analyzed tissue, and later was found to contain a silenced Cas9 gene (described below).

The Cas9 expression was analyzed in a subset of T0 plants and compared with non-transgenic wild-type and the constitutive Cas9 lines using the real-time quantitative PCR. Of 12 plants, nine showed an increase in the Cas9 expression (2–84×) upon HS over their respective room-temperature (RT) values (Fig. 6a; Table 2). Two T0 plants (#2, #10) appeared to be silenced as the relative Cas9 expression did not increase by the HS treatment in these plants, whereas #14 showed equally high expression at RT and HS (Table 2). Three constitutive-Cas9 lines expressing RUBI-CRISPR/Cas9 (RUBI-1,2, 3) were included in the analysis, each of which showed strong relative expression, and one of them (RUBI-1) harbored targeted mutations in the GUS gene (Table 2). In comparison to these RUBI-Cas9 lines, the Cas9 expression was three orders of magnitude lower in HS- Cas9 lines, which could be induced ~34-fold by HS (Fig. 6b; Table 2).

Inheritance of targeted mutations by the progeny

T0#1 and #3 were selected for the progeny analysis. These plants, at the young vegetative stages, showed strong GUS activity but diminished activity in the flowering stages, presumably due to multiplication of cells harboring mutations in the *GUS* gene (Fig. 5a; S2; Table 2). Sequencing of the sg1 and sg2 sites in these plants at the flowering stage detected a rare targeted mutagenesis in the sg1 site and a monoallelic mutation at the sg2 site (Fig. 4a, b). Twenty-four

seeds derived from T0#1 parent and 30 seeds from T0#3 parent were germinated for the progeny analysis. When their coleoptiles were fully emerged, seedlings were subjected to 2–3 rounds of HS treatment. Therefore, de novo targeting could occur in the Cas9+ lines. Histochemical GUS staining of these seedlings (~2 weeks after germination) showed strong (+) or diminished (−) GUS staining (Fig. 5b; Tables S3, S4). As expected, *Cas9* independently segregated in the population, and a few null-segregants were identified (Table 3). A subset of 16 T1 plants derived from T0#1 was subjected to PCR/sequencing at sg1 and/or sg2 sites. At the sg1 site, 11 contained monoallelic (68.7%) and one biallelic mutations (6.2%), while at sg2 site, nine contained monoallelic (56.2%) and one biallelic (6.2%) mutations (Table 3). Analysis of 25 T0#3 progeny, on the other hand, revealed monoallelic and biallelic mutations at the sg1 site in 18 (72%) and two (8%), respectively, while at sg2 only monoallelic mutations (96%) were found (Table 3). The remaining inherited the WT allele. The analysis of mutant reads revealed 4–5 types of mutations among T0#1 progeny but only one type at each site among T0#3 progeny (Fig. 7a-b). The abundance of one type of mutation in each population indicates a high rate of inheritance, which was confirmed by three Cas9 null-segregant in each population that harbored mutations at the sg1 and/or sg2 sites (Fig. 7c, d). The detection of only one type of mutation among T0#3 progeny raises the question whether this line is derived from *HS-Cas9* activity induced by the tissue culture. However, since the analysis of three different leaf samples of T0#3 plant detected only the WT sg1 site (Fig. 4b), the observed mutations are likely established in the germline at a later stage, possibly after the HS treatment of this plant.

Reduced rate of off-targeting in HS-CRISPR/ Cas9 lines

A total of 29 off-target (OT) sites with significant matches to the four designed sgRNAs against *GUS* or *PDS* genes were selected for PCR-sequencing analysis (Table S5, S6). However,

six GUS-OTs could not be validated by sequencing in the parental controls, and therefore, removed from the analysis. The remaining 23 OTs, representing eight GUS-OTs and 15 PDS-OTs, were analyzed in their respective transgenic lines. In order to compare the rates of off-targeting between the inducible (HS-Cas9) and the constitutive (RUBI-Cas9) expression systems, RUBI-CRISPR/Cas9 lines targeting *PDS* and *GUS* were included in this analysis (Table S7). The only difference between the RUBI- and HS-CRISPR/Cas9 lines used in this study is the promoter of Cas9, while both expressed the same sgRNAs by the rice *U3* promoter.

Four of the 23 OTs, representing the intergenic or intronic regions, were found to be targeted in one or more lines, whereas, targeting in the remaining 19 OTs was undetectable in both RUBI- and HS-Cas9 lines analyzed in this study (Tables S5, S6). Off-targeting by Cas9 was defined as insertion-deletions (indels) at the predicted DSB site; although, other effects such as base substitution, and the occasional single base insertion in the seed sequences were also observed (Fig. S3). Only one line showed 3-nt insertion near PAM but away from DSB of GUS OT-11. This variation was called as "other effects" since it did not occur at the predicted DSB site. Tissue culture is widely known to induce somaclonal variations, including transitions and transversions in the intergenic and intronic regions at high rates (Tang et al., 2018; Zhang, Wang, et al., 2014). Therefore, the observed single-nucleotide variations in the seed sequences or PAM that did not fall in the DSB site were called as non-Cas9, possibly tissue culture effects (Fig. S3). Of the four OTs that were evidently targeted by Cas9, PDS- OT2 was targeted in five of eight RUBI-Cas9 lines (~62%), showing indels at the predicted DSB site. The remaining three, all of which were GUS- OTs, were targeted in 1–7 RUBI-Cas9 lines (~4–30%) (Fig. 8a, Table 4). Off-targeting in HS-CRISPR/Cas9 lines was analyzed in 22 PDS (see Table S1) and 27 GUS samples (see Tables 2, S3, S4), representing pre-HS or post-HS samples. Only PDS-OT2 was found to be

targeted among HS-CRISPR/ Cas9 lines, whereas no off-target mutations were found in GUS-OT2, 3 or 11 in any of the HS-CRISPR/Cas9 lines. Three pre-HS samples and two post-HS samples showed off-target mutations in PDS-OT2 (Fig. 8b). Mutations in the pre-HS sample could arise from a high background Cas9 activity or a high transient activity in the progenitor cells during the DNA delivery process. These pre-HS samples did not contain the on-target mutations (Table S1). Off-targeting in the clones lacking on-target mutations has been reported by others (Aryal, Wasylishen, & Lozano, 2018). In summary, RUBI-Cas9 was found to be much more active in creating insertion-deletions in four different off-target sites, while a reduced rate of off-targeting was observed in the HS-Cas9 lines tested in this study.

Discussion

The CRISPR/Cas9 system shows high efficiency targeting in plants and animals, and is often described as a precise system that generates limited or undetectable off-target effects in plants (Feng et al., 2018; Lee et al., 2018; Tang et al., 2018). However, since the mechanism of targeting is based on a short-stretch of sequence complementarity and presence of a trinucleotide PAM (NGG) (Jinek et al., 2012), and since mismatches are tolerated at the PAM-distal end, numerous sites in a complex genome could potentially fall within the scope of CRISPR/Cas9 targeting. Further, sequences ending with non-canonical PAMs such as NAG can also be targeted by Cas9 (Zhang et al., 2014c), and while chromatin structure plays a marginal role in targeting, the secondary structures in the target DNA and the sgRNA could allow significant pairing, in spite of the mismatches at the PAM end (Lin et al., 2014). In both mammalian and plant cells, higher concentrations or the constitutive expression of sgRNA:Cas9 reportedly induced a high rate of off-target mutations (Hsu et al., 2013; Hu et al., 2018; Pattanayak et al., 2013; Svitashev et al., 2015). In plants, ribonucleoprotein Cas9 (RNP) has been used as an

effective transient expression system (Liang et al., 2017; Svitashev et al., 2016). However, the efficiency of the RNP in plant cells is impacted by the difficulty in delivering it into the cell wall-bounded compartments and isolating the edited lines in the selection-free transformation system (Yin, Gao, & Qiu, 2017). Inducible expression systems can be argued as more versatile transient expression systems, provided they generate low or undetectable background expression and a high-induced expression. Heat-shock promoters meet these criteria as they have been successfully used in applications where their proper regulation was critical, for example, controlling the Cre-*lox* recombination or the nuclease activity for marker excision (Khattri, Nandy, & Srivastava, 2011; Lloyd, Plaisier, Carroll, & Drews, 2005; Nandy & Srivastava, 2012; Nandy, Zhao, Pathak, Manoharan, & Srivastava, 2015; Zhang et al., 2003). Here, we describe the use of the heat-shock (HS) -CRISPR/Cas9 system consisting of the HS-inducible expression of the Cas9 and the standard *U3* promoter for sgRNA expression. We found that HS-CRISPR/ Cas9 at the room temperature was suppressed in rice tissue culture and the regenerated plants as mutations in the targeted sites occurred at a low rate in this study (16%). However, upon HS treatment, the characteristic CRISPR/Cas9 mutations were found in ≥ 50% of calli at the targeted sites (Table 1). It is well known that targeting efficiency varies between the genomic sites. However, constitutive CRISPR/Cas9 is often reported to generate ≥80% targeting (Ma et al., 2015; Zhou, Liu, Weeks, Spalding, & Yang, 2014). Therefore, the relative targeting efficiency of HS-Cas9 with one or two rounds of HS treatments appears to be lower than that of the constitutive-Cas9. Whether this efficiency could be further improved by additional HS treatments is yet to be determined. The two Cas9 expression systems could not be compared in T0 plants, in this study, as HS-induced mutations in the plants are evident only as rare or chimeric mutations, indicated by the baseline secondary trace in the sequence spectra (Fig. S1).

However, in plants, inheritance rate is the most important criteria of the gene editing efficiency. We show that the HS-induced mutations in T0 plants were transmitted to the progeny at a high rate and segregated independently from Cas9 (Table 3). Further, our data reflect on the efficiency of HS-CRISPR/ Cas9 is inducing mutations in the meristem, leading to the mutant cell lineage in the somatic tissue and the germline, which explains the high frequency of one type of mutation observed in the progeny, especially, in the T1 progeny of T0#3 parent (Fig. 7a,b).

Drug-inducible gene editing systems have been described for the human cells (Dow et al., 2015; Nihongaki, Otabe, & Sato, 2018), but heat-inducible Cas9 has so far been used only in *Caenorhabditis elegans* (Li, Yi, & Ou, 2015; Liu et al., 2014). In addition to their potential in curbing off-target effects, inducible expression systems could confer spatio-temporal control on gene editing, which can simplify editing of essential genes, avoid lethality by activating Cas9 at specific developmental stage, and improve gene editing efficiency by inducing Cas9 in the repair-competent cells. Use of the heat-inducible expression system could also leverage improved CRISPR/Cas9 activity by heat-shock, leading to higher rates of mutagenesis (LeBlanc et al., 2018). Additionally, heat-shock was found to enhance the sgRNA levels (Fig. S4), which could improve gene editing efficiency, if the sgRNA is limiting. Although, the molecular basis of heat-induction of sgRNAs is not clear, a similar observation was made in *Arabidopsis* by LeBlanc et al. (2018). Finally, HS-CRISPR/ Cas9 was found to be more precise as it generated either undetectable or a lower rate of off-target activity on the predicted off-target sites (Table 4). Of 28 OTs screened in this study, four OTs (PDS-OT2, GUS-OT2, 3, 11) were found to be targeted in the constitutive (RUBI-Cas9) CRISPR/Cas9 lines. Irrespective of the OT site, a higher percentage of off-targeting was observed in the constitutive RUBI-Cas9 lines. PDS-OT2 was targeted in ~62% of RUBI-Cas9 lines, and GUS-OTs were targeted in 4–30% of the RUBI-

Cas9 lines. HS-Cas9 lines, on the other hand, did not show off-targeting at GUS-OTs and showed a reduced rate (~22%) of off-targeting at PDS-OT2 (Table 4). Since the analysis was based on the Sanger sequencing, off-targeting in every other line cannot be ruled out; however, this study showed a clear difference in the rates of off-targeting in the inducible and constitutive CRISPR/Cas9 systems. Finally, as all the clones were derived from tissue culture, base substitutions in the target sites were observed in both HS- and RUBI-CRISPR/Cas9 lines.

In summary, we demonstrate HS-inducible CRISPR/Cas9 system is generally suppressed at the ambient room temperature in rice, and activated by the heat-shock treatment. The heat-shock-induced genome editing is efficient at producing heritable targeted mutations, while curbing the off-target mutations. Targeting of more loci and a deeper analysis of off-targeting will be needed to affirm the precision of the HS-CRISPR/Cas9 system for wider applications in plant biotechnology. However, this pilot study shows that HS-CRISPR/ Cas9 is a promising genome editing tool that can provide temporal control toward improving the precision of the CRISPR/Cas9 activities. This expression platform could also be used for the temporal control of other gene editing tools such as CRISPR/Cas12a.

References:

Aryal NK, Wasylishen AR, Lozano G (2018).CRISPR/Cas9 can mediate high-efficiency off-target mutations in mice in vivo. Cell Death Dis. 9(11):1099. doi: 10.1038/s41419-018-1146-0.

Bae S, Park J and Kim J-S (2014). Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30:1473-1475. https://doi.org/10.1093/bioinformatics/btu048.

Brinkman EK, Chen T, de Haas M, Holland HA, Akhtar W and van Steensel B (2018). Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. Molecular Cell. 2018 70:801-813. Doi: 10.1016/j.molcel.2018.04.016.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA and Zhang F (2013).Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823. doi:10.1126/science.1231143.

Czarnecka E, Ingersoll JC and Gurley WB (1992). AT-rich promoter elements of soybean heat shock gene Gmhsp17.5E bind two distinct sets of nuclear proteins in vitro. Plant Mol. Biol. 19:985–1000.

Dehairs J, Talebi A, Cherifi Y and Swinnen JV (2016). CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. Scientific Rep. 6:28973. doi: 10.1038/srep28973.

de Solis, CA, Ho A, Holehonnur R and Ploski JE (2016). The Development of a Viral Mediated CRISPR/Cas9 System with Doxycycline Dependent gRNA Expression for Inducible *In vitro* and *In vivo* Genome Editing. Front. Mol. Neurosci. 9, 70. https://doi.org/10.3389/fnmol.2016.00070

Dow LE, Fisher J, O'Rourke KP, Muley A, Kastenhuber, ER, Livshits G, Tschaharganeh DF, Socci ND and Lowe SW (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nature Biotech. 33:390−394. doi:10.1038/nbt.3155.

Endo A, Masafumi M, Kaya H and Toki S (2016).Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. Scientific Rep. 6:38169. doi: 10.1038/srep38169.

Feng C, Su H, Bai H, Wang R, Liu Y, Guo X, Liu C, Zhang J, Yuan J, Birchler JA and Han F (2018). High-efficiency genome editing using a dmc1 promoter-controlled CRISPR/Cas9 system in maize. Plant Biotech. J.1-10. doi: org/10.1111/pbi.12920.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK and Sander JD (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotech. 13:9. doi:10.1038/nbt.2623.

Hill JT, Demarest BL, Bisgrove BW, Su YC, Smith M and Yost HJ. (2014). Poly Peak Parser: Method and software for identification of unknown indels using Sanger Sequencing of PCR products. Dev. Dyn. 243:1632-6. doi: 10.1002/dvdy.24183.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G and Zhang F (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotech. 31:827-32. doi: 10.1038/nbt.2647.

Hu X, Meng X, Liu Q, Li J, Wang K (2018). Increasing the efficiency of CRISPR-Cas9-VQR precise genome editing in rice. Plant Biotechnol J. 16:292-297. doi: 10.1111/pbi.12771.

Jasin M and Haber JE (2016).The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. DNA Repair 44:6–16. doi: 10.1016/j.dnarep.2016.05.001.

Jefferson RA (1987). Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol. Biol. Rep. 5:387-405.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA and Charpentier E (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816-821. doi: 10.1126/science.1225829.

Khattri A, Nandy S and Srivastava V (2011). Heat-inducible Cre-*lox* system for marker excision in transgenic rice. Journal Biosci. 36:37–42. doi: 10.1007/s12038-011-9010.

Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z and Joung JK (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529:490-495. doi:10.1038/nature16526.

LeBlanc C, Zhang F, Mendez J, Lozano Y, Chatpar K, Irish VF and Jacob Y (2018).Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. Plant J. 93:377-386. doi: 10.1111/tpj.13782.

Lee K, Zhang Y, Kleinstiver BP, Guo JA, Aryee MJ, Miller J, Malzahn A, Zarecor S, Lawrence-Dill CJ, Joung JK, Qi Y and Wang K (2018). Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. Plant Biotech. J. 1-11. doi: 10.1111/pbi.12982.

Li W, Yi P and Ou G (2015). Somatic CRISPR–Cas9-induced mutations reveal roles of embryonically essential dynein chains in *Caenorhabditis elegans* cilia. J. Cell Biol. 208:683–692. doi: 10.1083/jcb.201411041.

Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q, Liu J, Zhang H, Liu C, Ran Y and Gao C (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nature Comm. 8:14261. doi: 10.1038/ncomms14261.

Lin Y, Cradick TJ, Brown MT, Deshmukh H, Ranjan R, Sarode N, Wile BM, Vertino PM, Stewart FJ and Bao G (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res. 42:7473-7485. doi:10.1093/nar/gku402.

Liu P, Long L, Xiong K, Yu B, Chang N, Xiong JW, Zhu Z and Liu D (2014). Heritable/conditional genome editing in *C. elegans* using a CRISPR-Cas9 feeding system. Cell Res. 24:886-889. doi:10.1038/cr.2014.73.

Liu X, Homma A, Sayadi J, Yang S, Ohashi J and Takumi T (2016). Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. Scientific Rep. 6:19675. doi:10.1038/srep19675.

Livak KJ and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25: 402-408. doi.org/10.1006/meth.2001.1262.

Lloyd A, Plaisier CL, Carroll D and Drews GN (2005). Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 102:2232–2237. doi: 10.1073/pnas.0409339102.

Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, Wang B, Yang Z, Li H, Lin Y, Xie Y, Shen R, Chen S, Wang Z, Chen Y, Guo J, Chen L, Zhao X, Dong Z, Liu YG (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol Plant. 8(8):1274-84. doi: 10.1016/j.molp.2015.04.007

Ma X, Zhu Q, Chen Y and Liu YG (2016). CRISPR/Cas9 platforms for genome editing in plants: developments and applications. Mol. Plant 9:961–974. doi: 10.1016/j.molp.2016.04.009.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE and Church GM (2013). RNA-guided human genome engineering via Cas9. Science 339:823–826. doi: 10.1126/science.1232033.

Mikami M, Toki S and Endo M (2015). Parameters affecting frequency of CRISPR/Cas9 mediated targeted mutagenesis in rice. Plant Cell Rep. 10:1807-1815. doi: 10.1007/s00299-015-1826-5.

Miki D, Zhang W, Zeng W, Feng Z and Zhu JK (2018). CRISPR/Cas9-mediated gene targeting in *Arabidopsis* using sequential transformation. Nature Comm. 9:1967. doi: 10.1038/s41467-018-04416-0.

Mojica FJ, Díez-Villaseñor C, García-Martínez J and Almendros C (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155:733-40. doi: 10.1099/mic.0.023960-0.

Murovec Jana, Guček Katja, Bohanec Borut, Avbelj Monika, Jerala Roman (2018). DNA-Free Genome Editing of Brassica oleracea and B. rapa Protoplasts Using CRISPR-Cas9 Ribonucleoprotein Complexes. Front. Plant Sci. 9:1594. DOI=10.3389/fpls.2018.01594.

Nandy S and Srivastava V (2012). Marker-free site-specific gene integration in rice based on the use of two recombination systems. Plant Biotech. J. 10:904–912. doi:10.1111/j.1467-7652.2012.00715.x.

Nandy S, Zhao S, Pathak BP, Manoharan M and Srivastava V (2015). Gene stacking in plant cell using recombinases for gene integration and nucleases for marker gene deletion. BMC Biotech. 15:93. doi 10.1186/s12896-015-0212-2.

Nihongaki Y, Otabe T and Sato M (2018). Emerging Approaches for Spatiotemporal Control of Targeted Genome with Inducible CRISPR-Cas9. Anal Chem. 90:429-439. doi: 10.1021/acs.analchem.7b04757.

Nishimura A, Aichi I and Matsuoka M (2006). A protocol for *Agrobacterium* mediated transformation in rice. Nature Protocols 1:2796–2802. doi:10.1038/nprot.2006.469.

Park J, Bae S and Kim JS (2015).Cas-Designer: A web-based tool for choice of CRISPR-Cas9 target sites. Bioinformatics 31:4014-4016. doi: 10.1093/bioinformatics/btv537.

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA and Liu DR (2013). High through put profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nature Biotech. 31:839-43. doi: 10.1038/nbt.2673.

Puchta H, Dujon B and Hohn B (1996). Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc. Natl. Acad. Sci. USA 14:5055-60.

Rouet P, Smih F and Jasin M (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc. Natl. Acad. Sci. USA 21:6064-6068.

Schindele P, Wolter F and Puchta H (2018). Transforming plant biology and breeding with CRISPR/Cas9, Cas12 and Cas13. FEBS Lett. 592:1954-1967. doi: 10.1002/1873-3468.13073.

Senturk S, Shirole NH, Nowak DG, Corbo V, Pal D, Vaughan A, Tuveson DA, Trotman LC, Kinney JB and Sordella R (2017). Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. Nature Comm. 22:14370. doi: 10.1038/ncomms14370.

Srivastava V, Underwood JL and Zhao S (2017). Dual-targeting by CRISPR/Cas9 for precise excision of transgene from rice genome. Plant Cell Tiss. Org. Cul. 129:153–160. doi:10.1007/s11240-016-1166-3.

Stemmer M, Thumberger T, del Sol Keyer M, Wittbrodt J and Mateo JL (2015). CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. PloS one Apr 24;10(4):e0124633. doi: 10.1371/journal.pone.0124633

Svitashev S, Young JK, Schwartz C, Gao H, Falco SC and Cigan AM (2015). Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. Plant Physiol. 2:931-45. doi: 10.1104/pp.15.00793.

Svitashev S, Schwartz C, Lenderts B, Young JK and Cigan AM (2016). Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. Nature Comm. 16:13274. doi: 10.1038/ncomms13274.

Szostak JW, Orr-Weaver TL, Rothstein RJ and Stahl FW (1983). The double-strand-break repair model for recombination. Cell 33:25-35.

Tang X, Liu G, Zhou J, Ren Q, You Q, Tian L, Xin X, Zhong Z, Liu B, Zheng X, Zhang D, Malzahn A, Gong Z, Qi Y, Zhang T and Zhang Y (2018). A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. Genome Biol. 4:84. doi: 10.1186/s13059-018-1458-5.

Wang M, Mao Y, Lu Y, Tao X and Zhu JK (2017). Multiplex gene editing in rice using the CRISPR-Cpf1 system. Mol. Plant 10:1011-1013. doi: 10.1016/j.molp.2017.03.001.

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

Wang ZP, Xing HL, Dong L, Zhang HY, Han CY, Wang XC and Chen QJ (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. Genome Biol. 16:144. doi: 10.1186/s13059-015-0715-0.

Waterworth WM, Drury GE, Bray CM and West CE (2011). Repairing breaks in the plant genome: the importance of keeping it together. New Phytol. 192:805–822. doi: org/10.1111/j.1469-8137.2011.03926.x.

Wolter F and Puchta H (2017). Knocking out consumer concerns and regulator's rules: efficient use of CRISPR/Cas ribonucleoprotein complexes for genome editing in cereals. Genome Biol. 18:43. doi: 10.1186/s13059-017-1179-1.

Wolt JD, Wang K, Sashital D and Lawrence-Dill CJ (2016). Achieving plant CRISPR targeting that limits off-target effects. Plant Geno. 9:3. doi: 10.3835/plantgenome2016.05.0047.

Wright AV, Sternberg SH, Taylor DW, Staahl BT, Bardales JA, Kornfeld JE and Doudna JA (2015). Rational design of a split-Cas9 enzyme complex. Proc. Natl. Acad. Sci. USA 12:2984-2989. doi: org/10.1073/pnas.1501698112.

Xie K and Yang Y (2013). RNA-guided genome editing in plants using a CRISPR-Cas system. Molecular Plant 6:1975-83. doi: 10.1093/mp/sst119.

Xie K, Minkenberg B and Yang Y (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. USA 112:3570–3575. doi: 10.1073/pnas.1420294112.

Xu H, Xiao T, Chen CH, Li W, Meyer CA, Wu Q, Wu D, Cong L, Zhang F and Liu JS (2015). Sequence determinants of improved CRISPR sgRNA design. Genome Res. 25: 1147–1157

Yin H, Gao P, Liu C, Yang J, Liu Z and Luo D (2013). SUI-family genes encode phosphatidylserine synthases and regulate stem development in rice. Planta. 237(1):15-27. https://doi.org/10.1007/s00425-012-1736-5

Yin K, Gao C and Qiu JL (2017). Progress and prospects in plant genome editing. Nature Plant 3:17107. doi: 10.1038/nplants.2017.107.

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV and Zhang F (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163:759-71. doi: 10.1016/j.cell.2015.09.038.

Zhang D, Wang Z, Wang N, Gao Y, Liu Y, Wu Y, Bai Y, Zhang Z, Lin X, Dong Y, Ou X (2014). Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. PLoS One. 9(5):e96879. doi.org/10.1371/journal.pone.0096879

Zhang W, Subbarao S, Addae P, Shen A, Armstrong C, Peschke V and Gilbertson L (2003). Cre/*lox*-mediated marker gene excision in transgenic maize (*Zea mays* L.) plants. Theor. Appl. Genet. 107:1157-1168. doi: 10.1007/s00122-003-1368-z.

Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, Mao Y, Yang L, Xu N and Zhu JK (2014a). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotech. J. 12:797–807. doi: org/10.1111/pbi.12200.

Zhang Y, Ge X, Yang F, Zhang L, Zheng J, Tan X, Jin ZB, Qu J and Gu F (2014b). Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Sci. Rep. 4:5405. doi:10.1038/srep05405.

Zhang D, Wang Z, Wang N, Gao Y, Liu Y, Wu Y, Bai Y, Zhang Z, Lin X, Dong Y, Ou X (2014c). Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. PLoS One. 9(5):e96879. doi: 10.1371/journal.pone.0096879

Zhou H, Liu B, Weeks DP, Spalding MH, and Yang B (2014). Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Res. 42(17):10903-14. doi: 10.1093/nar/gku806.

Tables and Figures

Table 1: HS-CRISPR/Cas9 activity in rice callus

Exp.	Target	Total no. of calli	Pre-HS calli[1]			Post-HS calli[1]		
			Total no.	Targeted[2]	Eff.[3]	Total no.	Targeted[2]	Eff.[3]
1	*PDS*	23	12	2	16	11	7	63.6
2	*GUS*	12	6	1	16	6	3	50.0

[1]Number of room temperature (pre-HS) or heat-shocked (post-HS) calli showing mutations at the two (sg1, sg2) target sites
[2]Indels at DSB sites of sg1 or sg2 targets
[3]Percent calli showing targeted mutations at one or both targets. See Table S1 and S2 for description of each line analyzed

Table 2: Characterization of T0 Plants transformed with HS-CRISPR/Cas9 targeting GUS gene

Line	GUS staining[#]		Cas9 expression		Sg1	Sg2	Off target studied
	Y	O	Fold-induction by HS	% RUBI-Cas9*			
1	+	-	7.0	0.03	WT[¶]	WT[¶]	Yes
2	+	+	0.35[†]	0.07	WT	WT	Yes
3	+	-	2.5	0.13	WT[¶]	WT[¶]	Yes
4	+	+	10	0.02	WT	WT	-
5	+	+	84	0.03	WT	WT	Yes
6	+	+	-	-	WT	WT	-
7	+	-	-	-	WT[¶]	WT[¶]	-
8	+	+	-	-	WT	WT	-
9	-	-	-	-	WT	Biallelic	-
10	+	+	0.45[†]	0.2	WT	WT	Yes
11	+	+	-	-	WT	WT	-
12	-	-	63	5.96	WT	Biallelic	Yes
13	+	-	-		WT	WT[¶]	-
14	+	+	1[‡]	16.96	WT	WT	-
15	+	+	2.2	-	WT	WT	-
16	+	+	-	-	WT	WT	-
17	+	+	-	-	WT	WT	-
18	+	-	6.9	0.09	WT	WT	Yes
19	+	-	9.2	0.02	WT[¶]	WT	Yes
20	+	+	3.1	0.03	WT	WT	Yes
RUBI-1	-	-	-	100	Biallelic	Biallelic	Yes
RUBI-2	+	+	-	100	-	-	-
RUBI-3	+	+	-	50	-	-	-

[#]Histochemical staining of leaf cuttings from young vegetative (Y) or older flowering (O) plants.
*Non-induced (room temp) expression value in HS-Cas9 compared to RUBI-Cas9 expression values
[†]Silenced Cas9 lines
[‡]Overexpression Cas9 lines
[¶]Baseline secondary sequence trace in the sequencing spectra (see Fig. S1).

Table 3: Inheritance of HS-CRISPR/Cas9 induced mutations by the progeny

Parent	No. of T1 plants analyzed	Cas9 (+)	Cas9 (-)	GUS staining[1]		% Mutants at Sg1		% Mutants at Sg2	
				+	-	Mono-allelic	Bi-allelic	Mono-allelic	Bi-allelic
T0#1	24	18	6	4	20	68.7	6.2	56.2	6.2
T0#3	30	25	5	-	30	72	8	96	-

[1]Histochemical staining of leaf cuttings showing strong (+) or weak/no (-) staining

Table 4: Comparative analysis of off-targeting by the inducible (HS) and the constitutiv (RUBI) CRISPR/Cas9 systems

	[a]Off-Targets (OT)	RUBI-CRISPR/Cas9			HS-CRISPR/Cas9			
		Total no. of samples	[b]Samples showing off-target mutation	%Off-targeting[c]	Total no. of samples	[b]Samples showing off-target mutations		%Off-targeting[c]
						[d]Pre-HS	[e]Post-HS	
1	PDS-OT2	8	5	62.5	22	3	2	22.7
2	GUS-OT2	23	1	4.3	27	0	0	0
3	GUS-OT3	23	7	30.4	27	0	0	0
4	GUS-OT11	23	6	26	27	0	0	0

[a] From Table S5-S6
[b] Characteristic insertions-deletions at the predicted DSB site.
[c] Percent lines showing off-target mutations regardless of the heat-shock treatment.
[d] Indels detected in room temperature samples.
[e] Indels detected in heat-shocked samples.

Table S1: Heat-shock induced CRISPR/Cas9 targeting of *PDS* gene in rice callus cultures

	Treatment	Sg1	Sg2	Subject to off-target analysis	Off-target mutation
1	RT	-	WT	Yes	-
2	RT	WT	WT	Yes	-
3	RT	Monoallelic[1]	Monoallelic[1]	Yes	-
4	RT	WT	WT	Yes	PDS-OT2
5	RT	-	WT	Yes	PDS-OT2
6	RT	-	WT	Yes	-
7	RT	WT	WT	Yes	-
8	RT	WT	WT	-	-
9	RT	WT	WT	Yes	-
10	RT	WT	Biallelic het.[1]	Yes	-
11	RT	WT	WT	Yes	-
12	RT	WT	WT	Yes	PDS-OT2
13	HS	-	Mosaic[2]	Yes	-
14	HS	Mosaic[2]	Mosaic[2]	Yes	-
15	HS	WT	WT	Yes	PDS-OT2
16	HS	WT	WT	Yes	PDS-OT2
17	HS	-	Mosaic[2]	Yes	-
18	HS	-	Mosaic[2]	Yes	-
19	HS	Mosaic[2]	Mosaic[2]	Yes	-
20	HS	WT	Mosaic[2]	Yes	-
21	HS	Mosaic[2]	Mosaic[2]	Yes	-
22	HS	WT	-	Yes	-
23	HS	WT	WT	Yes	-

[1]Mutations identified by CRISP-ID tool
[2]Multiple overlapping sequencing traces downstream of the predicted DSB sites

Table S2: Heat-shock induced CRISPR/Cas9 targeting of *GUS* gene in rice callus cultures

Samples	Treatment	Sg1	Sg2
1	RT	Monoallelic[1]	WT
2	RT	WT	WT
3	RT	WT	WT
4	RT	WT	WT
5	RT	WT	WT
6	RT	WT	WT
7	HS	WT	WT
8	HS	WT	Mosaic[2]
9	HS	WT	WT
10	HS	WT	WT
11	HS	Mosaic[2]	Mosaic[2]
12	HS	WT	Mosaic[2]

[1]Mutations identified by CRISP-ID tool
[2]Multiple overlapping sequencing traces downstream of the predicted DSB sites

Table S3: Analysis of T1 progeny of T0#1

T1 plant	Cas9 PCR	GUS staining[1]	Sg1 Site	Sg2 Site	Subject to off-target analysis
1	+	-	Monoallelic ±1	Monoallelic +1	Yes
2	+	-	WT¶	WT¶	-
3	+	-	WT¶	WT	-
4	+	-	WT¶	WT¶	-
5	+	-	WT¶	Monoallelic +1	-
6	-	-	Monoallelic -1	Monoallelic +1	Yes
7	+	-	Monoallelic -2	WT¶	-
8	-	-	Monoallelic -1	Monoallelic +1	Yes
9	+	-	Biallelic (-1/-7)	Monoallelic ±1	-
10	-	-	Monoallelic -1	Monoalellic +1	Yes
11	+	-	-	WT	-
12	+	-	-	Monoalellic +1	-
13	+	+	Monoallelic ±1	-	-
14	+	-	-	Monoallelic +1	-
15	+	-	Monoallelic ±1	-	Yes
16	+	-	-	Monoallelic +1	-
17	+	-	Monoallelic -1	-	-
18	+	+	-	-	-
19	+	+	-	-	-
20	-	-	Monoallelic -1	Biallelic (+1/±3)	-
21	-	-	-	WT	-
22	+	+	Monoallelic -1	-	Yes
23	-	-	Monoallelic -1	Monoallelic +1	-
24	+	-	-	-	-

[1]Strong (+) or diminished (-) GUS activity.

¶Baseline secondary sequence trace in the sequencing spectra, indicating rare mutations.

Table S4: Analysis of T1 progeny of T0#3

T1 plant	Cas9 PCR	GUS staining	Sg1 Site	Sg2 Site	Subject to off-target analysis
1	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	Yes
2	-	-	Monoallelic -1 (T)	Monoallelic +1 (G)	Yes
3	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
4	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
5	+	-	Biallelic homozygous -1 (T)	Monoallelic +1 (G)	Yes
6	-	-	WT	Monoallelic +1 (G)	-
7	+	-	Monoallelic -1 (T)	WT	-
8	+	-	Monoallelic -1 (T)	-	-
9	+	-	Monoallelic -1 (T)	-	-
10	-	-	Monoallelic -1 (T)	Monoalellic +1 (G)	-
11	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
12	+	-	Monoallelic -1 (T)	Monoalellic +1 (G)	-
13	-	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
14	+	-	WT	Monoallelic +1 (G)	-
15	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
16	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
17	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
18	-	-	Biallelic homozygous -1 (T)	-	-
19	+	-	Monoallelic -1 (T)	-	-
20	+	-	-	Monoallelic +1 (G)	-
21	+	-	-	Monoallelic +1 (G)	-
22	+	-	-	-	-
23	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
24	+	-	-	Monoallelic +1 (G)	-
25	+	-	-	Monoallelic +1 (G)	-
26	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
27	+	-	Monoallelic -1 (T)	Monoallelic +1 (G)	-
28	+	-	WT	Monoallelic +1 (G)	-
29	+	-	WT	Monoallelic +1 (G)	-
30	+	-	WT	Monoallelic +1 (G)	-

Table S5: Potential off-target sites of *GUS* sgRNAs

Off target site	Search criteria	Match to	Sequence[±]	Chromosome: Location	Mismatches	Off targeting
1	20 nt seed	sgRNA 1	CCACCAACGCTGATCtATTaCTa	1:33226683-33226705	3	None
2	20 nt seed	sgRNA 1	CCACCAACaCTGAcCAtTTCaAa	7:17045436-17045457	5	Yes[$]
3	20 nt seed	sgRNA 1	CCACCAACGCTGAcCAtTTCaAa	1:18693750-18693769	4	Yes[$]
4	20 nt seed	sgRNA 1	GTGGtAacGATtAGCGTTGGGGG	5:23611011-23611033	4	NA*
5	12 nt PAM-proximal	sgRNA 1	GATCAGCGTTGGaGG	12:21835725-21835739	1	None
6	20 nt seed	sgRNA 1	GTGGcATTGATCAGCGgTtGTGG	10:1526410-1526432	3	NA*
7	20 nt seed	sgRNA 1	GTaGAAagGATCAGaGTTGGAGG	8: 23532054-23532076	4	None
8	20 nt seed	sgRNA 1	GTGGcAATgTGATCgGCGTTGGTGG	2:15364552-15364576	3	None
9	20 nt seed	sgRNA 2	gaGCGgCgGCAAACCGAAGTGGG	4:28810154-28810176	4	NA*
10	20 nt seed	sgRNA 2	TCaCAAaCGCAAaCCGAAGGGGG	2:28847219-28847241	4	NA*
11	20 nt seed	sgRNA 2	cCaCGACCGCAAACCaAAGcAGG	3:34663865-34663887	4	Yes[$]
12	20 nt seed	sgRNA 2	TCGC-ACCGCAAAtCGtAG-CGG	2:31797105-31797135	4	NA*
13	20 nt seed	sgRNA 2	CCGcCTTCGGcTTGCGGcCGC-A	6:2588632-2588653	4	NA*
14	12 nt PAM-proximal	sgRNA 2	GCAAACCGAATGG	7:19473625-19473638	1	None

[±] Small red fonts are mismatches and red (-) dashes are gaps (deletions). PAM are underlined.
[*]Not analyzed (NA) due to no/non-specific amplification in negative controls.
[$]Shown in Table 4.

Table S6: Potential off-target sites of *PDS* sgRNAs

Off target site	Search criteria	Match to	Sequence[±]	Chromosome: Location	Mismatches	Off - targeting
1	20 nt seed	sgRNA1	AgAAGCacGaAGAATTCAGCTGG	8:15278101-15278123	4	None
2	20 nt seed	sgRNA1	AtAgcCCAGGAaAATTCAGCAGG	5:27556570-275565792	4	Yes[$]
3	20 nt seed	sgRNA2	aAaTGCATttATAACTCATCTGG	6:14957598-14957620	4	None
4	20 nt seed	sgRNA1	gCAAGCtAGGAtAATTaAGCAGG	3:7857294-7857316	4	None
5	20 nt seed	sgRNA1	ACAtaCgAGGAGAATTCAGtAGG	4:31131597-31131619	4	None
6	20 nt seed	sgRNA1	CCTGCTG-ATTCTtCTGGCTTcT	10:3090619-3090640	3	None
7	20 nt seed	sgRNA1	CCGGCTttATTCTCtTtGCTTGT	1:34782800-34782822	4	None
8	20 nt seed	sgRNA1	ACAAGCCAaGATAtATTCAGCAGG	4:29528959-29528979	2	None
9	12nt PAM-proximal	sgRNA1	GGAGAATTCAGCCGG	3:34889349-34889363	0	None
10	20 nt seed	sgRNA2	CCTGAT--GTTAcCCATtCAGTG	2:15451132-15451152	3	None
11	20 nt seed	sgRNA2	CCTGcTGAcTTtTCCAATGCAGTG	3:18156090-18156113	3	None
12	20 nt seed	sgRNA2	CCCGATGAG-T-TCCATGCtGTG	4:29456082-29456102	3	None
13	20 nt seed	sgRNA2	CCTGATGA-TTATAC-TG-AGTG	11:4868927-4868946	3	None
14	20 nt seed	sgRNA2	CCCGATGAGTTAcCCA-GtAGTG	12:5686362-5686383	3	None
15	12nt PAM-proximal	sgRNA2	CCTGATtAGTTATCC	1:39212196-39212210	1	None

[±] Small red fonts are mismatches and red (-) dashes are deletions. PAM is underlined.
[$]Shown in Table 4.

Table S7: **RUBI-CRISPR/Cas9 lines used in off-target analysis**

Gene	Line #	Tissue Type	On-Target mutation[1]		Off-target mutation
			Sg1	Sg2	
GUS	1	Leaf	No	No	OT2
	2	Leaf	Yes	Yes	OT3 and 11
	3	Leaf	Yes	Yes	OT3
	4	Leaf	Yes	Yes	OT3
	5	Leaf	Yes	Yes	OT3 and 11
	6	Leaf	Yes	Yes	OT3
	7	Leaf	Yes	Yes	OT3
	8	Leaf	No	Yes	OT3
	9	Callus	Yes	Yes	OT11
	10	Leaf	Yes	Yes	OT11
	11	Leaf	Yes	Yes	OT11
	12	Leaf	Yes	Yes	OT11
	13	Callus	ND	ND	none
	14	Leaf	No	No	none
	15	Leaf	No	No	none
	16	Callus	ND	ND	none
	17	Leaf	Yes	Yes	none
	18	Leaf	No	Yes	none
	19	Leaf	No	No	none
	20	Leaf	Yes	Yes	none
	21	Leaf	Yes	Yes	none
	22	Leaf	Yes	Yes	none
	23	Leaf	Yes	Yes	none
OsPDS	1	Callus	Yes	Yes	none
	2	Callus	Yes	ND	OT2
	3	Callus	Yes	Yes	none
	4	Callus	Yes	Yes	OT2
	5	Callus	Yes	ND	OT2
	6	Callus	Yes	ND	OT2
	7	Callus	Yes	ND	none
	8	Callus	Yes	ND	none

[1]Detected either by fragment deletion in PCR indicating dual-simultaneous activity of sg1 and sg2 or by sequencing of individual targets.
ND: not determined

Table S8: Primers used for the vectors construction and on-site targeted mutagenesis

Primer	Sequence (5' – 3')	Application
gGus1-F	TAGGTCTCCTGATCAGCGTTGGgttttagagctagaa	Construction of *GUS* sgRNA1: *5'*-GTGGAATTGATCAGCGTTGG-*3*'
gGus1-R	CGGGTCTCAATCAATTCCACtgcaccagccggg	
gGus2-F	TAGGTCTCCCCGCAAACCGAAGTgttttagagctagaa	Construction of *GUS* sgRNA2: *5'*-ACGCGACCGCAAACCGAAGT-*3*'
gGus2-R	CGGGTCTCAGCGGTCGCGTtgcaccagccggg	
gPDS1-F	TAGGTCTCCCAGGAGAATTCAGCgttttagagctagaa	Construction of *OsPDS* sgRNA1: *5*'-ACAAGCCAGGAGAATTCAGC-*3*'
gPDS1-R	CGGGTCTCACCTGGCTTGTtgcaccagccggg	
gPDS2-F	TAGGTCTCCATGGATAACTCATCgttttagagctagaa	Construction of *OsPDS* sgRNA2*: 5*'-CACTGCATGGATAACTCATC-*3*'
gPDS2-R	CGGGTCTCACCATGCAGTGtgcaccagccggg	
Ubi1812	TCTAACCTTGAGTACCTATCTATTA	Genotyping B1 (Ubi:GUS) locus
NosR2	GCGGGACTCTAATCATAAAAACCC	
PDS-F	GGTAGAAATGCCATGCGGGA	Genotyping *OsPDS* locus
PDS-R	GTGGTGAGGTTCGGCTGAAT	
Cas9F	AAAGACCGAGGTGCAGACAG	Ca9 genotyping & real-time PCR
Cas9R	ACCAGCACAGAATAGGCCAC	
BamH1-Cas9F	CGCGGATCCATGGACTATAAGGACCACGACGG	Construction of HS-Cas9
EcoR1-Kpn1-nosR	GGAATTCGGTACCGATCTAGTAACATAGATGACACCGCCCG	
U3-F	CGGGATCCGACCATGATTACGCCAAGCTTAAG	Construction of sgRNA vector
Ptg-R	CGGGATCCAAGCTTTCTAGACCGCCTTGACCCGAATTTGTG	
PDS sg1F	GGC ACAAGCCAGGAGAATTCAGC	Real-time quantitative PCR
PDS sg2F	GCACACTGCATGGATAACTCATC	
sgRNA-R	CGA CTC GGT GCC ACT TTT TCA AGT TG	
Ubi-F	CGCAAGTACAACCAGGACAA	
Ubi-R	GCTGTGACCACACTTCTTCTT	

Table S9: Primers used in the off target analysis

Gene	Off target site	Primers	Sequence (*5'-3'*)	Application
PDS	1	PDSoff1F PDSoff1R	TGTGGTTTTGGTTGAGGGCA GCCCTAAAAGAGGCCGTTCA	Genotyping PDS off target site 1
	2	PDSoff2F PDSoff2R	CACTCTTAGCAGTAGGCTATGG GTAGGAGTTGTACTCACGGATG	Genotyping PDS off target site 2
	3	PDSoff3F PDSoff3R	GAGAGGGAGAAACCACACAATC GCCTCCTGAACTTCTGCTATATTC	Genotyping PDS off target site 3
	4	PDSoff4F PDSoff4R	TCAGAGCGATCTCCCAGAAT TCTTCTCGGGCTCGACCATA	Genotyping PDS off target site 4
	5	PDSoff5F PDSoff5R	CAGAGAAGACCACTTACAGA CACTGACTTACTTCCATCAAGG	Genotyping PDS off target site 5
	6	PDSoff6F PDSoff6R	ATACACGCGCCACAGACAAT ATCGCAGGCGATCTCGAA	Genotyping PDS off target site 6
	7	PDSoff7F PDSoff7R	CACTCCATCTCTACACAGCT C TGTACTGTGACACCGGGTAG	Genotyping PDS off target site 7
	8	PDSoff8F PDSoff8R	ACATGGCGTGGTGCATAA GATTCAGGGATCAGGATGACAC	Genotyping PDS off target site 8
	9	PDSoff9F PDSoff9R	CGTCCAATGTATCTCCTCTTC GCTTGTTGTGGGCTTAGTTG	Genotyping PDS off target site 9
	10	PDSoff10F PDSoff10R	CAAAGGACTTACAGGACGTG TATAGAGAGGGAAGGACCCA	Genotyping PDS off target site 10
	11	PDSoff11F PDSoff11R	CCAGTCGAACCATTCAGTGAC TACAGCCAGAGGTGGTATG	Genotyping PDS off target site 11
	12	PDSoff12F PDSoff12R	ACTCCCGACCTCTAGTTTC CTTGTTGTACGCCTGCAT	Genotyping PDS off target site 12
	13	PDSoff13F PDSoff13R	CCAAGTATGCCAAAGGTGTG GTACGGAGCAAAGTGTTTCC	Genotyping PDS off target site 13
	14	PDSoff14F PDSoff14R	GTTTCCGTGCAAATCTGATG TCTTCGAGCATCCTATCCA	Genotyping PDS off target site 14
	15	PDSoff15F PDSoff15R	GCT GAC TAG TGT TAC GTG CA CAGCACTCACAGCAACATAGC	Genotyping PDS off target site 15
GUS	1	Gusoff1F Gusoff1R	ATGCGCTCGCCATAGAATAG TCAGCGTGGAAGATGAAGTG	Genotyping GUS off target site 1
	2 and 3	Gusoff2F Gusoff2R	CGAAGATTCCTCCGCGATTAC CATGGATGGAACCAACCTAGAC	Genotyping GUS off target site 2 and 3
	5	Gusoff5F Gusoff5R	CCGAACCCATCTTGATTCTCTT AGAAGAAGCTCCCACCATTTAC	Genotyping GUS off target site 4
	7	Gusoff7F Gusoff7R	TCGTACCCGTTCAGTATACGG GATGACATGCGTCCACAAACAC	Genotyping GUS off target site 6
	8	Gusoff8F Gusoff8R	CCTTGTCGTCGTTGGTTCTG CAAGCGGCACGAGATTTG	Genotyping GUS off target site 7
	11	Gusoff10F Gusoff10R	TCGCTGCTCCAAGCTCTC CAA CAG GTT GCT AGA GCG	Genotyping GUS off target site 10
	14	Gusoff14F Gusoff14R	CCCTTCAACACCGGATCGAAG GAAGAGGCCGGACAGGTTCTT	Genotyping GUS off target site 13

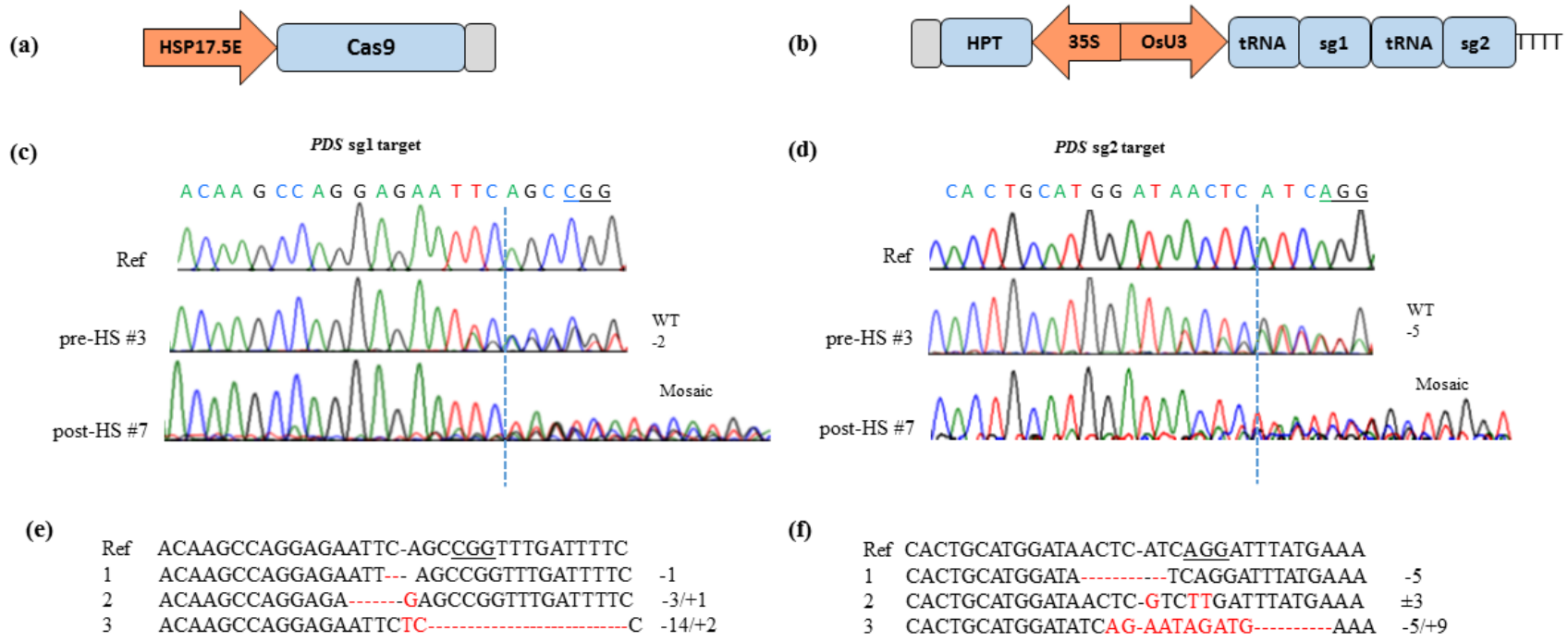


Figure 1: Efficacy of heat-shock (HS) -inducible CRISPR/Cas9 on the rice Phytoene Desaturase (PDS) gene. (a) HS-Cas9 expression construct consisting of the soybean heat-shock protein 17.5E (HSP17.5E) gene promoter and the *Streptococcus pyogenes* Cas9 coding sequence; (b) standard sgRNA construct consisting of the rice sno U3 promoter expressing a pair of sgRNAs via the tRNA processing mechanism. For the plant selection, hygromycin resistance gene consisting of the 35S promoter and the hygromycin phosphotransferase (HPT) gene was included in the construct. Pol III terminator is shown as TTT, and gray bars represent nos 3′ terminators; (c–d) Sequencing spectra of the PDS target sites (PAM underlined) in the wild type reference, and the representative HS CRISPR/Cas9-transformed callus lines, without heat-shock (pre-HS) or after a few days of HS (post-HS). Targeted mutations are indicated by two or multiple overlapping sequence traces (mosaic) near the predicted double-stranded break (DSB) site (dotted line) in the spectra; (e-f) Alignments of the reference sequence with the mutant reads as identified by the CRISP-ID tool or TA cloning.

Figure 1 continued…

Insertion-deletions (indels) are indicated by the red fonts and the dashed lines. Number of insertions or deletions is also indicated. PAM site (underlined) and predicted DSB sites (-) are indicated in the reference sequences.

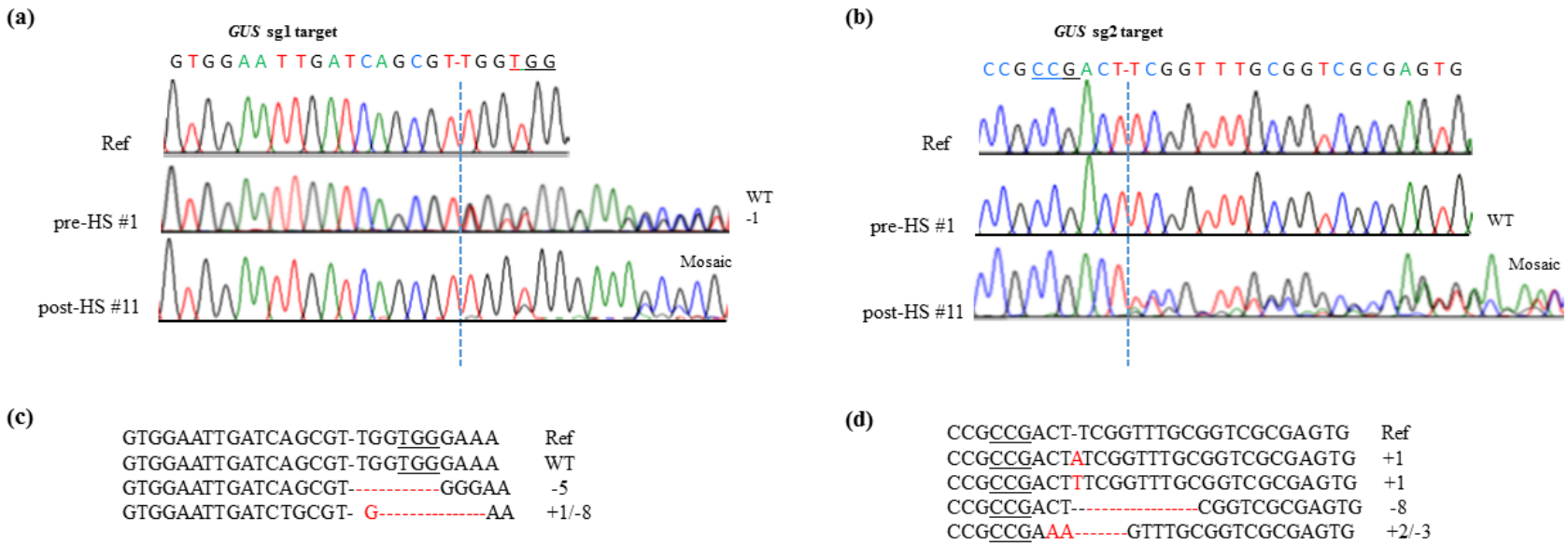


Figure 2: Efficacy of HS-CRISPR/Cas9 on the GUS transgene located in the rice genome. (a, b) Sequencing spectra of the GUS target sequences from the parental B1 line (ref., PAM underlined), and the targeted callus lines, without heat-shock (pre-HS) or with HS treatment (post-HS). Dotted vertical lines represent the predicted DSB sites. Overlapping sequence traces in the spectra indicate the mosaic mutation pattern; (c, d) Mutations in the spectra as identified by the CRISP-ID tool (c) or TA cloning (d). Dashes indicate deletions, and the red letters indicate insertions. Number of insertions-deletions in each sequence is indicated. PAM site (underlined) and the predicted DSB sites (-) are also indicated.

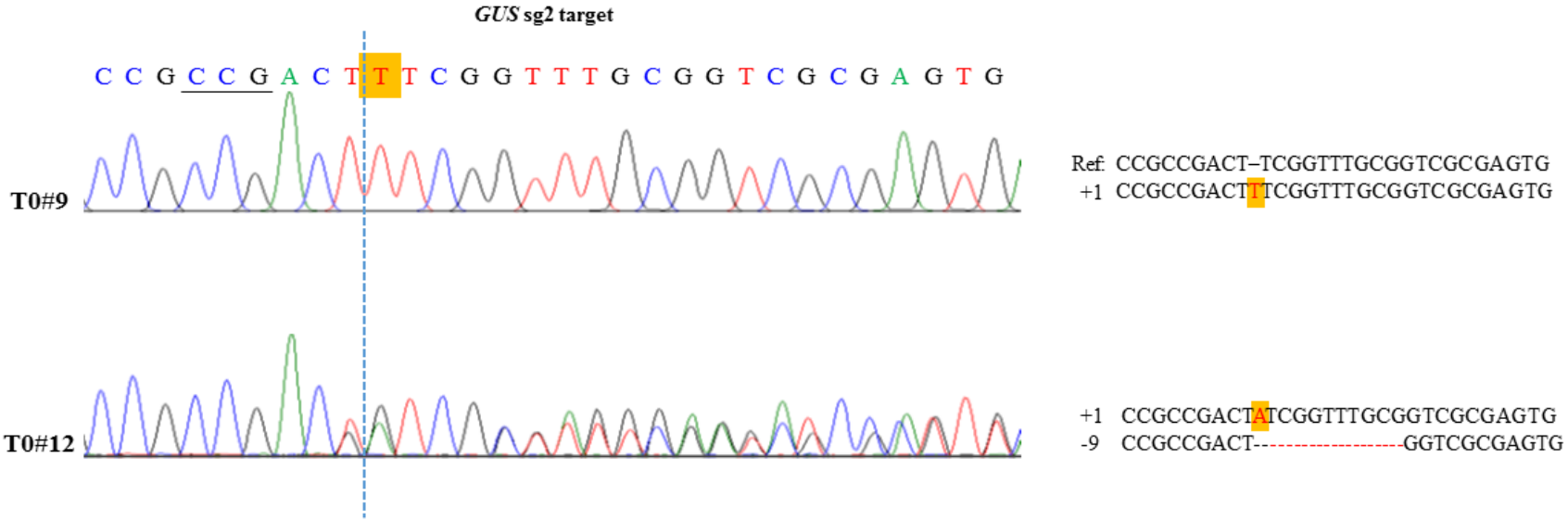


Figure 3: Sequencing of the *GUS* sg2 target site in T0 plants #9 and #12 harboring HS-CRISPR/Cas9 constructs. Mutation types are shown adjacent to each spectrum along with the reference sequence. Dashed vertical line indicates the predicted DSB site. PAM site is underlined. Shaded red letter indicates insertions, and dashes indicate deletions. The two sequences in T0#12 were separated using the CRISP ID tool.

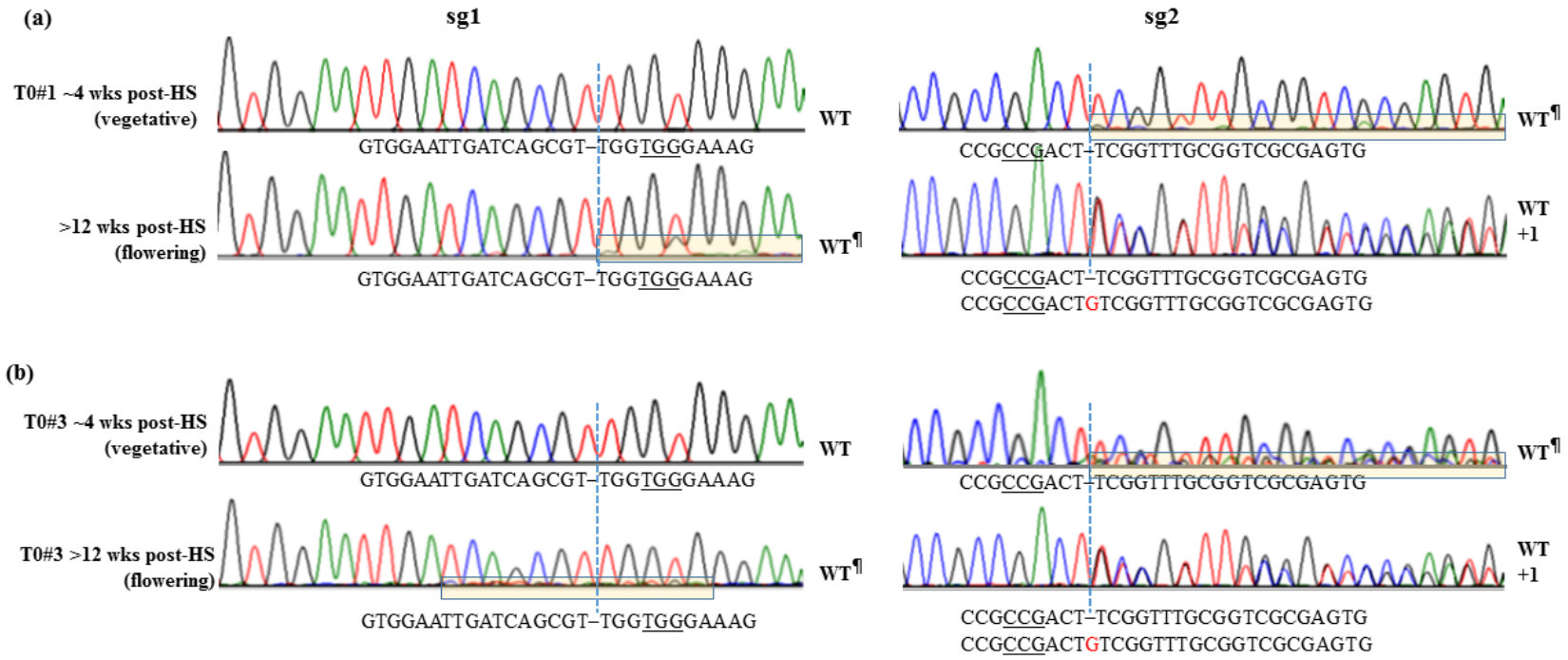


Figure 4: Genotyping of T0 plants #1 (a) and #3 (b) at *GUS* sg1 and sg2 sites by PCR-sequencing at two growth stages, ~4 weeks after heat-shock (HS) or the vegetative stage and ~12 weeks after HS or the flowering stage. Mutation types are shown below each sequencing spectra with the PAM sequence underlined. The predicted DSB sites are indicated by the vertical lines. The baseline secondary sequence traces in the spectra are boxed, indicating a low rate of mutations in largely wild type samples (WT¶; see Table 2). The spectra containing two overlapping sequences were analyzed by the CRISP-ID tool to identify monoallelic +1 mutations in the two plants. Major sequences in the remaining are shown below each spectrum

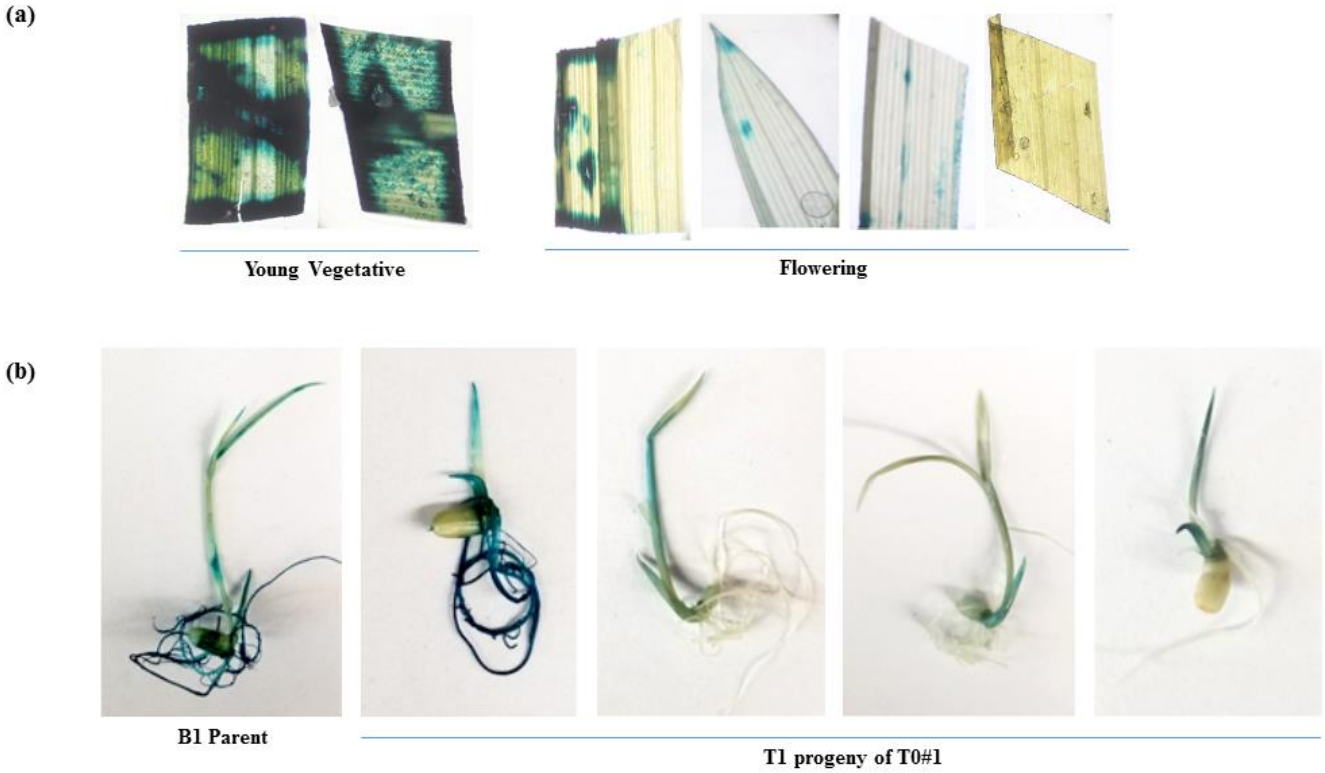


Figure 5: Histochemical GUS staining in the HS-CRISPR/Cas9 line. (a) Leaf cuttings from the post-HS T0#1 plant at the young vegetative stage and from the flowering plant. Note the staining in the cut end and poked points, and diminished staining in the leaves of flowering plant; (b) Seedlings of the control B1 line harboring the GUS gene and the progeny of the HS-CRISPR/Cas9 line #1.

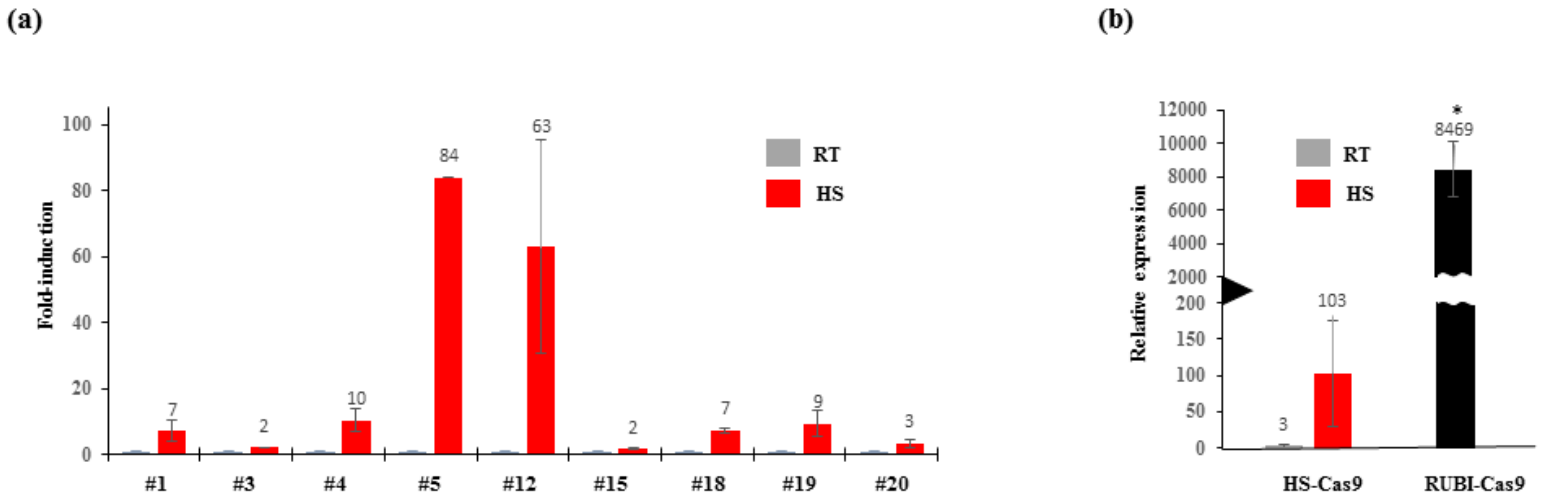


Figure 6: Cas9 expression analysis. (a) Fold-induction of Cas9 in T0 plants by the heat-shock (HS) treatment (3 h exposure to 42°C) as compared to the background room-temperature (RT) values; (b) Relative expression of Cas9 in HS-Cas9 lines with respect to the constitutive RUBI-Cas9 lines. The expression in HS-Cas9 lines was calculated at RT and upon HS. The average of 8 HS-Cas9 lines and 3 RUBI-Cas9 lines is shown with standard errors (*p-value < 0.001).

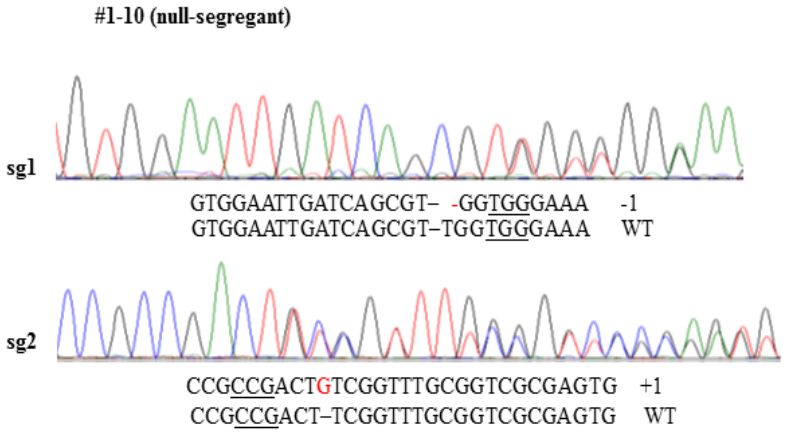
(a)

Types of sg1 mutations		No. of progeny T0#1	T0#3
WT	GTGGAATTGATCAGCGT–TGGTGGGAAA	4	5
-1	GTGGAATTGATCAGCGT–--GGTGGGAAA	8	20
±1	GTGGAATTGATCAGCGT–--GGGGGAAA	3	
-2	GTGGAATTGATCAGCGT–----GTGGGAAA	1	
-7	GTGGAATTGATCAGCGT–----------AAA	1	

(b)

Types of sg2 mutations		No. of progeny T0#1	T0#3
WT	CCGCCGACT–TCGGTTTGCGGTCGCGAGTG	6	1
+1	CCGCCGACTGTCGGTTTGCGGTCGCGAGTG	8	24
+1	CCGCCGACTATCGGTTTGCGGTCGCGAGTG	1	
+1	CCGCCGACTTTCGGTTTGCGGTCGCGAGTG	1	
±1	CCGCCGACA- TCGGTTTGCGGTCGCGAGTG	1	
±3	CCGCCGACTGCGGGTTTGCGGTCGCGAGTG	1	

(c)



(d)

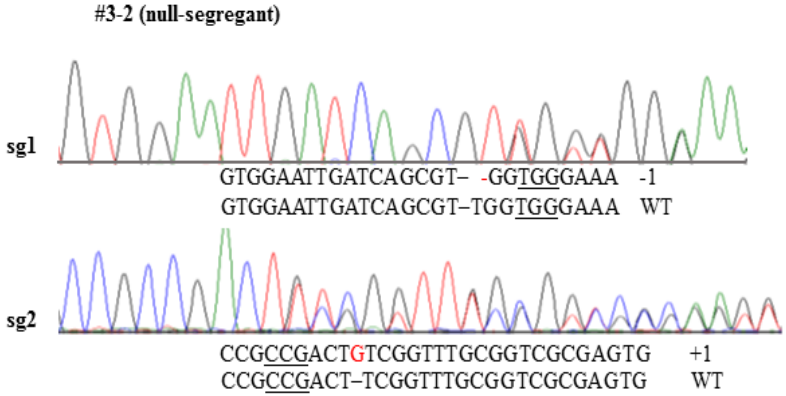


Figure 7: Inheritance of HS-CRISPR/Cas9-induced mutations by the progeny of T0#1 and #3. (a, b) Number of T1 plants harboring monoallelic or biallelic indels at the GUS sg1 and sg2 target sites. Indels are shown as dashes and the red letters; (c, d) Inheritance of mutations in the two Cas9 null-segregants harboring monoallelic mutations at the sg1 and sg2 sites. The sequence reads as identified by separating overlapping reads by the CRISP-ID tool and their alignments are shown below each spectrum. Insertion and deletion are shown by red letter or dashes. PAM is underlined

Figure 8: Off-target site analysis. Sequencing alignments of the predicted *PDS* and *GUS* off-target (OT) sites in the constitutive (RUBI) and the inducible (HS) CRISPR/Cas9 lines. (a) Sequence alignments of the off-target sites in the reference (WT or B1 parent) and the RUBI-CRISPR/Cas9 lines indicating insertion-deletions (indels) at the predicted DSB sites; (b) alignment of PDS OT2 in HS-CRISPR/Cas9 pre-HS and post-HS lines. Predicted DSB site (^) and PAM (underlined) are indicated. Blue fonts indicate mismatches between the reference sequence and the sgRNA, purple fonts indicate single-nucleotide polymorphisms between mutant reads and the reference sequence, red dashes are deletion, and red small fonts are insertions. Types of mutations in each line and the Cas9 presence are also shown. The line numbers are given in Table S1 (HS-CRISPR/Cas9) and Table S7 (RUBI-CRISPR/Cas9).

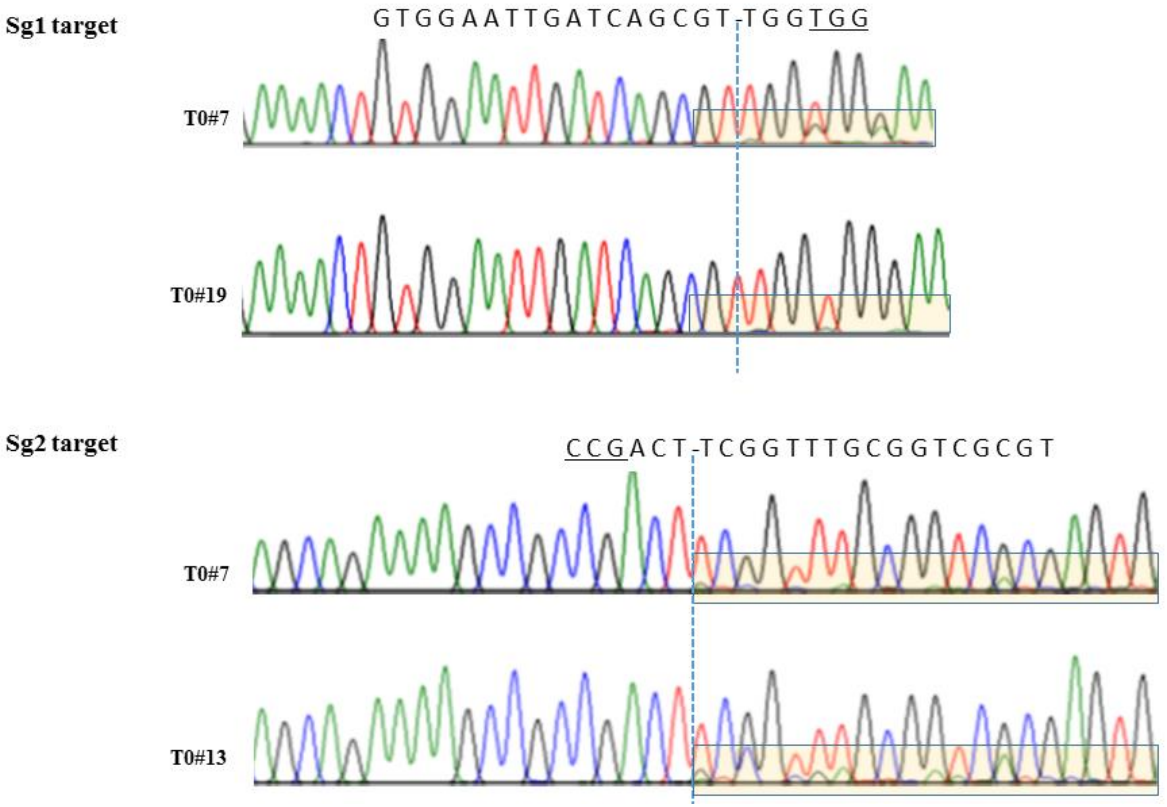


Figure S1: Representative sequence spectra with baseline secondary sequence trace (boxed area) indicating a low rate of mutagenesis induced by HS-CRISPR/Cas9 activity. The target sites with PAM (underlined) are shown above each spectra.

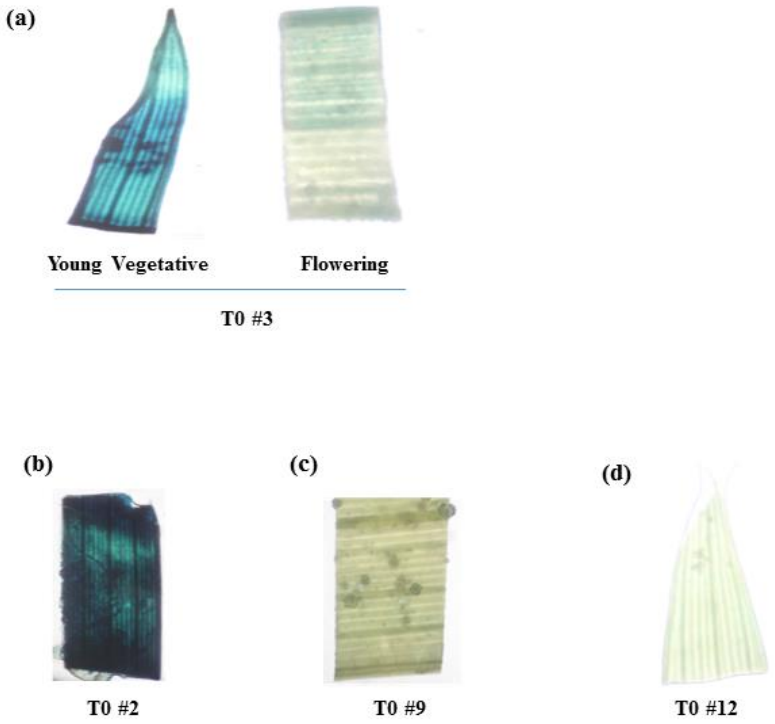


Figure S2: Histochemical GUS staining in the HS-CRISPR/Cas9 lines. **(a)** Leaf cuttings from the post-HS T0#3 plant at the young vegetative stage and the flowering plant. **(b-d)** Leaf cuttings from the post-HS plants of T0#2, #9, and #12 at the flowering stage.

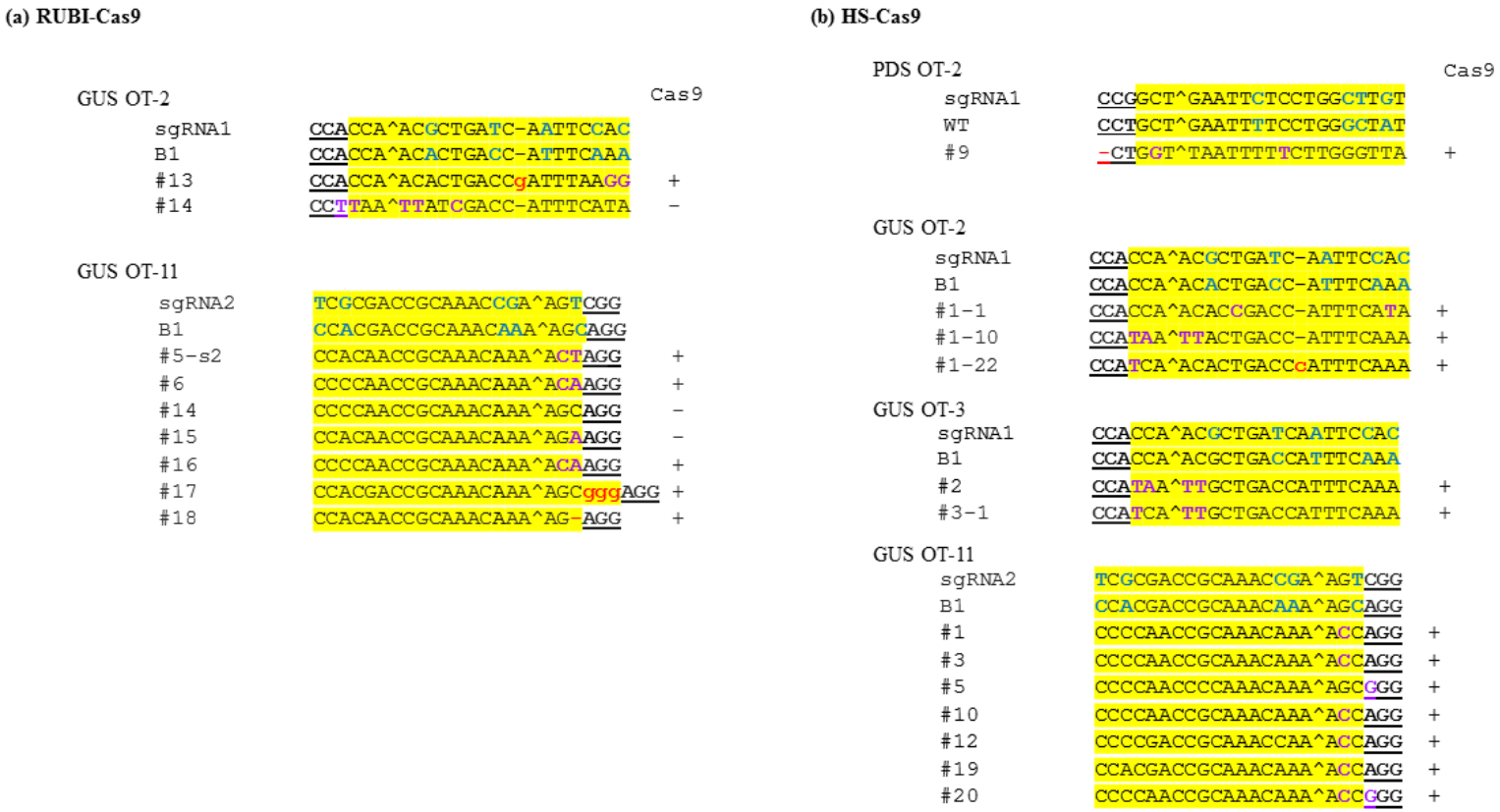


Figure S3: Other (tissue culture) effects in the off-target sites. Sequence alignments of off-target sites with significant matches to PDS or GUS sgRNAs between wild-type reference and the mutant reads obtained from constitutive (RUBI) or heat-inducible (HS) CRISPR/Cas9 lines. Predicted DSB site (^) and PAM (underlined) are indicated. Blue fonts indicate mismatches between the reference sequence and the sgRNA, purple fonts indicate single-nucleotide polymorphism, red dashes are deletion, and red small fonts are insertions. Presence of Cas9 in each line is indicated.

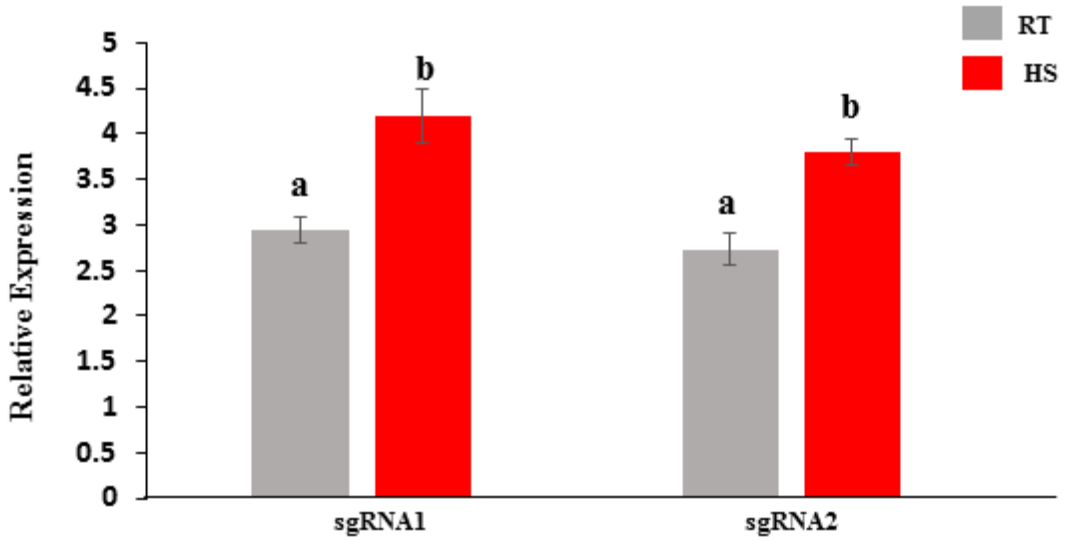


Figure S4: sgRNA expression analysis by real-time quantitative PCR in PDS HS-CRISPR/Cas9 lines. Relative expression at room temperature (RT) and upon heat-shock (HS) at 42oC for 3 h. Average of 5 independent HS-CRISPR/Cas9 lines is shown as log10 transformed values relative to WT. Statistical differences (a, b) were determined by Student *t* test.

Conclusions:

This study evaluated approaches for gene stacking, marker gene excision, and targeted mutagenesis for crop genome engineering and fast track breeding. Cre-*lox* recombination was used to catalyze site-specific integration of a DNA fragment encoding 5 genes. The resulting site-specific integration (SSI) locus harbored 3 constitutively expressed genes and 2 inducible genes that were induced by heat or cold treatment. It was observed that >50% transgenic lines recovered contained full-length site-specific integration (SSI) of the 5 genes, and each of genes expressed according to their promoter-specificity. Gene expression analysis using the leaf tissues showed that the genes expressed by constitutive promoters strongly expressed at normal growth conditions, while genes expressed by the inducible promoters mostly stayed silent at ambient temperatures but were strongly induced upon hot or cold treatment. In the progeny analysis, the expression of the constitutively expressing *NPT, GFP* and *GUS*, as determined by protein or enzyme activity, showed similar expression levels among independent SSI lines, with a correlation with the allelic number of the locus. Expression of the inducible genes (*AtDREB1A* and *pporRFP*) by heat- or cold-inducible promoters was also found to be duly regulated by heat or cold treatment. These data indicate that Cre-*lox* site-specific recombination is an effective approach for stacking multiple genes in the rice genome that could then simplify breeding of multi-genic traits.

The *CCR5*-zinc finger nuclease (*CCR5*-ZFN) and I-*Sc*eI were tested for their efficiency in generating targeted excisions in rice and *Arabidopsis*. The constitutively expressing I-*Sc*eI and *CCR5*-ZFN overexpressing constructs were refractory to the transformation. The inducible I-*Sc*eI in *Arabidopsis* showed the somatic excision, but failed to transmit to the progeny. The inducible expression of *CCR5*-ZFN in rice showed a stable inheritance in the progeny, but was ineffective

in creating targeted excisions. Therefore, these nucleases did not appear to be effective in generating the heritable targeted deletions. Future studies on the off target effects of these nucleases, and their activities in other plant species, are needed to determine their utility in plant biotechnology.

The constitutively expressing CRISPR/Cas9 was tested for its efficiency in generating targeted mutagenesis in rice on three loci including genic and intergenic regions, representing 6 target sequences. The targeted and heritable genomic excisions were observed at a lower frequency in the calli. The genomic deletions by dual targeting of the paired guide RNAs were mainly observed in callus and plants regenerated from these calli. The randomly regenerated plants showed mixed genomic effects but did not harbor heritable genomic deletions. The point-mutagenesis was studied in 114 plants, including primary transgenic lines and their progeny. Point-mutagenesis was observed in 78% of the lines. Thus, point mutations were highly favored over genomic deletions.

Due, to the concern of the off-target effects of the CRISPR/Cas9, a heat-inducible (HS) approach of CRISPR/Cas9 expression was tested by analyzing targeted mutations and off-target effects, and compared with the lines expressing strong constitutive CRISPR/Cas9. Only a low rate of pre-HS mutagenesis was detected in the lines harboring HS-CRISPR/Cas9, but an increased rate of mutagenesis was observed after HS treatment. The HS-induced mutations were transmitted to the progeny at a higher rate, generating monoallelic, and biallelic mutations that segregated independently from *Cas9* gene. However, the genomic deletions through dual targeting were undetected in the HS-CRISPR/Cas9 lines. Further, the off-target effects were either undetectable or detected at a lower rate in HS-CRISPR/Cas9 lines when compared with the constitutive-overexpression CRISPR/Cas9 lines.

In conclusion, this study has tested the feasibility and utility of multiple approaches and their components towards developing biotechnology approaches for crop improvement. The Cre-*lox* recombination and CRISPR/Cas9 were found to be highly efficient in gene stacking and targeted mutagenesis, respectively. Use of nucleases such as I-*Sce*I and *CCR5*-ZFN in rice and *Arabidopsis*, on the other hand, was problematic due to cytotoxicity to the cells.

UNIVERSITY OF ARKANSAS

Office of Research Compliance

November 26, 2018

MEMORANDUM

TO:	Dr. Vibha Srivastava
FROM:	Ines Pinto, Biosafety Committee Chair
RE:	Protocol Renewal
PROTOCOL #:	07012
PROTOCOL TITLE:	Precise Transgene Integration and Gene Editing Systems for Plants
APPROVED PROJECT PERIOD:	**Start Date** November 19, 2006 **Expiration Date** November 18, 2021

The Institutional Biosafety Committee (IBC) has approved your request, dated October 3, 2018, to renew IBC # 07012, "Precise Transgene Integration and Gene Editing Systems for Plants".

The IBC appreciates your assistance and cooperation in complying with University and Federal guidelines for research involving hazardous biological materials.

1424 W. Martin Luther King, Jr. • Fayetteville, AR 72701
Voice (479) 575-4572 • Fax (479) 575-6527

Department of
Crop, Soil, and Environmental Sciences

October 1, 2019

To: The Graduate School, University of Arkansas

From: Vibha Srivastava, Professor

Re: Bhuvan Pathak, Dissertation- Recombinant DNA

I am writing this letter for Bhuvan Pathak whose dissertation study entitled "Analyzing multigene stacking and genome editing strategies in rice" required the use of recombinant DNA.

The protocol # 01072, in which Ms. Pathak is named, has been approved by the Institutional Biosafety Committee. Ms. Pathak has over eight years of experience in the recombinant DNA and plant transformation field. She has been properly trained in handling the recombinant DNA material, and has fulfilled biosafety training requirements on timely basis as required by the University of Arkansas.

Please let me know if you need additional information or have any questions.

Thank you,

Vibha Srivastava

Department of
Crop, Soil, and Environmental Sciences

November 12, 2019

To Whom It May Concern:

I certify that Bhuvan Pathak is the first/equal author in the following manuscripts. She contributed 55 to 70% of the work described in these manuscripts.

Pathak BP, Pruett E, Guan H, Srivastava V (2019) Utility of I-*Sce* I and *CCR5*-ZFN nucleases in excising selectable marker genes from transgenic plants. BMC Research Notes. 12(1):272. https://doi.org/10.1186/s13104-019-4304-2

Pathak B, Zhao S, Manoharan M, Srivastava V (2019) Dual-targeting by CRISPR/Cas9 leads to efficient point mutagenesis but only rare targeted deletions in the rice genome. 3 Biotech. 9(4):158. https://doi.org/10.1007/s13205-019-1690-z

Nandy S, Pathak B, Zhao S, Srivastava V (2019) Heat-shock-inducible CRISPR/Cas9 system generates heritable mutations in rice. Plant Direct. 3 (5):e00145. https://doi.org/10.1002/pld3.145

Sincerely,

Vibha Srivastava