TARGETED GENOME REGULATION AND EDITING IN PLANTS

Dissertation by

Agnieszka Piatek

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology,

Thuwal, Kingdom of Saudi Arabia

EXAMINATION COMMITTEE APPROVALS FORM

The dissertation of Agnieszka Piatek is approved by the examination committee.

Committee Chairperson: Magdy Mahfouz, Assistant Professor
Committee Members: Takashi Gojobori, Distinguished Professor, Peiying Hong, Assistant Professor
External Committee Member: Neal Stewart, Professor

ABSTRACT

The ability to precisely regulate gene expression patterns and to modify genome sequence in a site-specific manner holds much promise in determining gene function and linking genotype to phenotype. DNA-binding modules have been harnessed to generate customizable and programmable chimeric proteins capable of binding to site-specific DNA sequences and regulating the genome and epigenome. Modular DNA-binding domains from zinc fingers (ZFs) and transcriptional activator-like effectors (TALEs) are amenable to engineering to bind any DNA target sequence of interest. Deciphering the code of TALE repeat binding to DNA has helped to engineer customizable TALE proteins capable of binding to any sequence of interest. Therefore TALE repeats provide a rich resource for bioengineering applications. However, the TALE system is limited by the requirement to re-engineer one or two proteins for each new target sequence. Recently, the clustered regularly interspaced palindromic repeats (CRISPR)/ CRISPR associated 9 (Cas9) has been used as a versatile genome editing tool. This machinery has been also repurposed for targeted transcriptional regulation. Due to the facile engineering, simplicity and precision, the CRISPR/Cas9 system is poised to revolutionize the functional genomics studies across diverse eukaryotic species. In this dissertation I employed transcription activator-like effectors and CRISPR/Cas9 systems for targeted genome regulation and editing and my achievements include: 1) I deciphered and extended the DNA-binding code of *Ralstonia* TAL effectors providing new opportunities for bioengineering of customizable proteins; 2) I repurposed the CRISPR/Cas9 system for site-specific regulation of genes in plant genome; 3) I harnessed the power of

CRISPR/Cas9 gene editing tool to study the function of the serine/arginine-rich (SR) proteins.

ACKNOWLEDGMENTS

First and foremost thank you to the late King Abdullah Bin Abdulaziz Al Saud for having a great vision, believe in science, hope for peace and for establishing this amazing University. I am very grateful to have been a part of the developing and ever growing "House of Wisdom".

Thank you to my family and friends who always believe in me and wish me nothing but happiness and success in my endeavors. To my Mom, my brother and my grandparents for always being there for me. To Marek for standing strong by my side and making me a better person every day. To my baby Marcel for giving me so much strength and for putting a constant smile on my face.

Thank you to my advisor, Dr. Magdy Mahfouz and committee members for advice, insights, support and encouragement. Thanks to all lab members, previous and present, for teaching me new skills and guiding me in the lab. A big thank you to all CDA members, who became my friends and were always ready to lend a hand. To lab management and admin staff for handling the paperwork and being always very helpful. To Burgundy for being patient and answering millions of my questions.

Thank you to the publishers Elsevier, Springer, John Wiley and Sons, Taylor and Francis for giving me the permission to reuse my publications in my thesis. The copyright belongs to the publishers and it is indicated in each thesis chapter.

I would like to dedicate this thesis to my Dad.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A - adenine
ABA - abscisic acid
AD - activation domain
AS - alternative splicing
C - cytosine
CHS - chalcone synthase
CRISPR - clustered regularly interspaced short palindromic repeats
CS - constitutive splicing
DBD - DNA-binding domain
dCas9 - dead Cas9
DSB - double-strand breaks
dsRNA - double-stranded RNA
EBE - effector-binding element
FLASH - fast ligation-based automatable solid-phase high-throughput
G - guanine
GOI - gene of interest
GT - gene targeting
GUS - *β*-glucuronidase
HDR - homology directed repair
HGT - horizontal gene transfer
ICA - interactive capped assembly
Indel - insertion deletion
KRAB - Krüppel-associated box
LIC - ligation-independent cloning
NHEJ - non-homologous end joining
NLS - nuclear localization signal
PAM - protospacer adjacent motif
PCR - polymerase chain reaction
PDS - phytoene desaturase
PTGS - post-transcriptional gene silencing
RNAi - RNA interference
RNAP - RNA polymerase II
RTLs - *Ralstonia* TALE-like proteins
RVD - repeat variable diresidue
sgRNA - single-guide RNA
snRNP - small nuclear ribonucleoproteins
SSN - site specific nuclease
T - thymine
TALE - transcription activator like effector
TALEN - transcription activator like effector nuclease
TFs - transcription factors
TGS - transcriptional gene silencing
TSS - transcriptional start site

TYLCV - tomato yellow leaf curl virus
ZF - zinc finger
ZF-ATF - zinc finger based artificial transcription factors
ZFN - zinc finger nuclease

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Targeted Transcriptional Regulation and Editing *via* Engineered DNA-Binding Proteins and CRISPR/Cas9 System

Advances in high-throughput sequencing provides vast amounts of sequence information of variety of eukaryotic genomes including plants. Currently, the challenge is to convert the information about the gene sequence into knowledge about the gene function. Therefore, highly efficient technologies to study gene function are needed to link a genotype to phenotype.

Modern molecular tools are being developed to rewrite genetic sequence and alter transcript production in plants. To elucidate gene function techniques such as overexpression (Prelich, 2012), underexpression (Smyth, 1997) including RNAi (Lindbo, 2012) have been applied. However, such methods often lack reproducibility, specificity and stability over generations. Therefore, highly specific tools are needed for gene functional analysis.

A variety of genome editing tools provide user-defined sequence specificities and DNA-binding modules that can be engineered to bind any DNA sequence are highly important in order to generate programmable site-specific nucleases and regulators. Recently, programmable DNA-binding modules including zinc finger (ZF) proteins, transcription activator-like effectors (TALE) gained much attention due to their ability to bind DNA (Bibikova *et al.*, 2003; Boch and Bonas, 2010). The biggest advantage of these technologies is that it is possible to specifically direct engineered proteins to target virtually any DNA sequence and therefore directly control the DNA message with high efficiency and precision (Shan *et al.*, 2013a). These technologies have been heavily utilized as genome regulation tools by tethering the activator or repressor domains to the DNA-binding module of the protein, in order to precisely up-regulate or down-regulate genes in predictable constitutive and spatiotemporal manner (Cong *et al.*, 2012; Mahfouz

et al., 2012; Urnov and Rebar, 2002). Other effector domains such as nucleases or methylases have been fused to the DNA-binding proteins providing desired outcomes including but not limiting to gene editing and gene methylation (Bultmann *et al.*, 2012; Deng *et al.*, 2012b; Li *et al.*, 2012) (Figure 1).

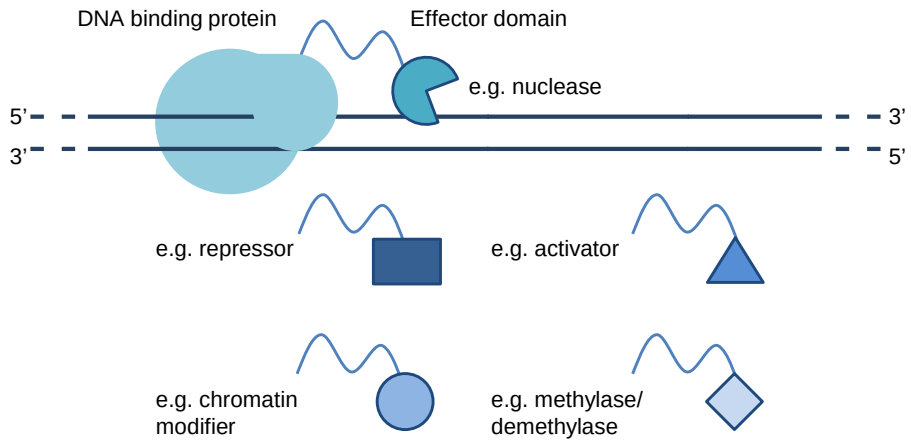


Figure 1. Customized synthetic regulators and modulators. Synthetic regulators or modulators are fusion proteins between DNA-binding proteins and various effector domains. DNA-binding domains are designed to target any user defined DNA sequence in promoter or exonic region of a gene of interest (GOI). Chimeric nucleases are used for targeted genome editing and to generate DNA DSBs. Chimeric activators and repressors are used for targeted genome regulation through recruitment of transcription factors and RNA polymerase. Chimeric modulators are used for targeted chromatin modifications and to induce epigenetic changes.

The first customizable DNA-binding modules were generated from the common Cys2-His2 zinc finger domain (Porteus and Carroll, 2005). Each zinc finger motif is composed of approximately 30 amino acids and binds to three nucleotides. Assembled zinc finger arrays customized to bind to a user-selected DNA sequence were fused to various effector domains in order to upregulate or downregulate gene expression profile. These assembly approaches for user-selected DNA target sequences have proven to be laborious, resource intensive, time consuming, expensive and they suffer from high

failure rates (Ramirez *et al.*, 2008). Because of these limitations, a major research goal has been to develop a DNA-binding module that can be easily engineered to bind to any sequence of interest with high efficiency and precision.

Nature provided us with an excellent tool for genome regulation studies, namely the transcription activator-like effectors (TALEs). The phytopathogenic *Xanthomonas* bacteria secrete a group of these proteins into plant cells to naturally reprogram the host transcriptional machinery to benefit the pathogen (Boch and Bonas, 2010). The simple code that governs the DNA recognition by the TALE protein has been harnessed to reprogram these proteins to bind any user-selected DNA sequence (Boch *et al.*, 2009; Moscou and Bogdanove, 2009). TALEs, just like ZF, have been tethered to various effector domains to better and precisely regulate gene sequence and expression respectively. The TAL-based genome engineering platform suffers from a number of limitations including laborious and time consuming engineering of proteins, the need to redesign DNA-binding domains for each target, potential off-target effects and cell toxicity.

Recently, an engineered version of the bacterial native defense system, namely the clustered regularly interspaced short palindromic repeats and Cas9 associated protein (CRISPR/Cas9) system, was developed as the first platform for targeting proteins to DNA target sites through RNA–DNA interactions rather than protein–DNA interactions (Jinek *et al.*, 2012). The CRISPR/Cas9 system was reengineered for transcriptional regulation purposes by retaining the DNA-binding capabilities of the Cas9 protein while inactivating the nuclease cleavage domains: this is termed dead Cas9 (dCsas9) (Qi *et al.*, 2013). Tethering dCas9 protein to various effector domains (e.g., activators, repressors,

or chromatin modifiers) has the potential to facilitate genomic screens through manipulating the expression time, abundance, and location of a particular mRNA.

All three platforms have been extremely useful in research to elucidate gene function. They have been successfully applied not only to genome regulations studies but also in genome editing studies.

In this dissertation I present results on the use of TALE and CRISPR platforms as genome regulation and genome editing tools. My work contributes to the field by the following achievements: 1) I discovered novel repeat variable residues (RVDs) from *Ralstonia* sp, that constitute the RVD-DNA code and govern the DNA recognition. Extending the code of RVDs provides new opportunities for bioengineering of new customizable proteins; 2) I contributed to the rich resource of DNA-binding modules from TAL like effectors from *Xanthomonas* sp. and proposed a novel, fast and efficient method for the RVD assembly method; 3) I developed dCas9 based programmable transcriptional regulators, both activators and repressors, and reported the first successful use of this gene modulation technology in plants; 4) I generated a rich resource of single and multiple SR gene knockouts collection and applied the CRISPR/Cas9 genome editing technology to study the interplay between stress responses and alternative splicing through the study of SR proteins in rice.

1.1 Targeted Genome Regulation

Regulation of gene transcription controls cellular functions, directs spatiotemporal development, and coordinates responses to physiological and environmental cues. Underexpression of a gene below certain threshold might result in a mutant phenotype.

Thus knockdown mutants are useful to determine gene function. Similarly, increased expression of a wild-type gene might also result in a phenotype, therefore gene overexpression has been exploited as a parallel approach in gene function studies. With the development of transformation technique in yeast (Beggs, 1978) and construction of genomic libraries (Nasmyth and Reed, 1980), overexpression has been widely applied as a genetic screening tool. Overexpression screens were soon applied in various organisms, including expression of a cDNA library in *Xenopus* (Smith and Harland, 1991), transposon based overexpression in *Drosophila melanogaster* (Staudt *et al.*, 2005) and full length human cDNA expression from *Homo sapiens* (Liu *et al.*, 2007).

The first method to achieve overexpression of genes in plants coincided with the discovery of *Agrobacterium tumefaciens* in the 1980s. *Agrobacterium* is a soil bacterium that transfers a part of its genome to the plant nucleus. The part of bacteria's genome called T-DNA, is integrated into the plant's DNA, causing the crown gall disease, described as cancerous growth mass. Jeff Schell, Mary-Dell Chilton and Marc van Montagu discovered the gene transfer mechanism between *Agrobacterium* and plants. Later it became possible to remove the disease-causing genes from bacteria's plasmid and replace them with a gene of interest (GOI). Using this approach bacteria would deliver GOI into the plant, adding a trait without causing disease. Hence, *Agrobacterium* became an established vector for plant transformation and gene overexpression, and marked the beginning of plant biotechnology. In the past few decades several studies reported different uses of overexpression such as: overexpression of wild type protein in mutant background (Prelich, 2012) or overexpression of dominant negative mutant genes (Barren and Artemyev, 2007; Voellmy, 2005).

Several other approaches have been generated to delineate gene function based on the screening of gain-of-function mutants. Approach known as activation-tagging system was developed in *Arabidopsis* and relied on the introduction of enhancer elements driven by the strong promoter, such as the cauliflower mosaic virus 35S promoter, through *Agrobacterium*-mediated DNA transfer (Hayashi *et al.*, 1992). Enhancers would activate the genes nearby the T-DNA insertion, resulting in elevated levels of genes' products. This technology was first applied in plant cell cultures (Kakimoto, 1996) and later in *Arabidopsis thaliana* to generate a population of transgenic plants (Weigel *et al.*, 2000). Later FOX (Full-length cDNA Over-eXpression) - hunting system, involving cDNA expression under 35S promoter was applied for systematic analysis of gain-of-function mutations in *Arabidopsis* (Ichikawa *et al.*, 2006).

Attempts to overexpress a chalcone synthase (CHS) gene in *Petunia hybrida*, homologous to petunia's native gene, resulted in partial or full inactivation of the gene. This led to the notion that abnormal copy number of a gene can lead to gene silencing. An extra gene copy, such as introduced transgene, may on occasion become inactivated, and also so does the corresponding endogenous gene. The term cosuppression describes such interactions. Cosuppression was observed in some of the transgenic petunia plants that had white flowers or flowers with mixed pigmentation (Napoli *et al.*, 1990). Disturbance in CHS enzyme activity resulted in a phenotype of flower petals lacking the usual purple pigmentation. mRNA analysis revealed that the white flowers had 50-fold less CHS transcripts when compared to wild type pigmented petals. The mechanism of cosuppression has been further studied and suggested to involve transcriptional gene silencing (TGS), through epigenetic changes either by methylation of chromatin

modification (Matzke and Matzke, 1991) or post-transcriptional gene silencing (PTGS), where the transgene and endogenous gene is transcribed, but the gene's transcripts are prevented from accumulating as a result of rapid RNA degradation (Cameron and Jennings, 1991). Cosuppression may also be initiated by the expression of the antisense transgenes. Antisense model proposes that gene silencing is governed by the formation of RNA duplex molecules between the homologous gene and the sense or antisense molecule delivered through transformation (Grierson *et al.*, 1991). The application of antisense technology has indicated the validity of this method to selectively inhibit the expression of particular genes. Gene silencing method by co-suppression through PTGS in plants is reminiscent of RNA interference (RNAi) in animals and quelling in fungi (Romano and Macino, 1992). All three methods involve inactivation of an endogenous gene by annihilating the RNA transcripts through formation of short interfering RNA that are produced by degradation of long double-stranded RNA (dsRNA) molecules.

The phenomenon of RNAi was first discovered in *C. elegans*, following the studies which reported that sense RNA was as effective as antisense RNA for suppressing gene expression in worms (Guo and Kemphues, 1995). Subsequently, it was shown that gene silencing can be achieved by introduction of purified dsRNA only partially homologous to target gene (Fire *et al.*, 1998). The main features of the RNAi degradation pathway include the formation of small RNAs by a molecule called Dicer and formation of RNA-induced silencing complex (RISC) (Bernstein *et al.*, 2001). Hamilton and Baulcome first discovered short double-stranded RNA molecules, of about 25 nucleotides (nt) in length, with target gene homology and associated them with the PTGS phenotype (Hamilton and Baulcombe, 1999). RNAi technology has become an important tool in basic and applied

research, due to its effectiveness in silencing particular genes of interest. For instance RNAi has been widely used in functional genomics studies in organisms like nematodes (Kamath *et al.*, 2003), insects (Boutros *et al.*, 2004) and mammalian cells (Lassus *et al.*, 2002). RNAi has been successfully used for functional genomics studies in hexaploid bread wheat (Travella *et al.*, 2006), maize (McGinnis *et al.*, 2007) as well as more model systems such as *Arabidopsis thaliana* and *Nicotiana benthamiana*. Examples of the use of this technology in plants include the alteration of plant architecture (Xu *et al.*, 2005), tolerance to abiotic stress (Jagtap *et al.*, 2011) and production of male sterile plants (Moritoh *et al.*, 2005).

Abovementioned techniques for gene overexpression and underexpression have had big impact on gene function studies, however they also carry many disadvantages. The lack of reproducibility and ability to simultaneously activate or repress genes in their native context are just a few shortcomings. Therefore, precise and efficient molecular tools are needed to characterize the function of single and multiple genes in linear and interacting pathways in a native context.

Natural DNA-binding domains of various proteins, such as TetR, LacI and LexA, were previously utilized to recruit effector domains for activation or repression of eukaryotic genes (Cronin *et al.*, 2001; Gossen and Bujard, 1992). Similarly, synthetic transcription factors were engineered to specifically control gene expression in a wide range of applications, from mammalian systems to diverse plant species. Examples of the use of synthetic transcription factors include stimulation of tissue regeneration (Rebar *et al.*, 2002), control over stem-cell differentiation (Bartsevich *et al.*, 2003) and genetic screens (Peng *et al.*, 2015). Engineered transcription factors can be directed to specific promoter

regions of endogenous genes in order to either activate or repress their expression through facilitating binding of RNA polymerase II (RNAP) to the promoter region or blockage of transcriptional initiation and interruption of transcriptional elongation, respectively.

The most frequently used technologies for targeting user-defined DNA sequences in the promoter for transcriptional regulation are DNA-binding domains of ZF proteins, TALE proteins and CRISPR system. Tethering DNA-binding proteins to various effector domains (e.g., activators, repressors, or chromatin modifiers) has the potential to facilitate genomic screens through manipulating the expression time, abundance, and location of a particular mRNA. Moreover, such DNA-binding based chimeras provide good platforms for recruitment of particular cellular transcription factors (TFs) to specific genomic loci in order to enhance regulation of gene expression. Engineered transcription factors can be tethered to various effector domains such as repressors (e.g., KRAB and SRDX domains) and activators (e.g., VP64, EDLL, and TAD domains) (Figure 2) to elicit transcriptional changes. Moreover, potential fusions with chromatin modifiers are also desired. For example it has been shown that TALE and CRISPR can bind methylated DNA (Deng *et al.*, 2012b; Hsu *et al.*, 2013). This suggested that TALE and CRISPR system could be used to manipulate epigenetic marks and thereby facilitate comprehensive studies of the association between the epigenome and transcriptional control.

ZFs are small structural protein motifs abundant in eukaryotic genomes. They play numerous roles in the cell from DNA recognition and transcriptional activation to apoptosis regulation and RNA packaging. They display various patterns of domains linked through a zinc knuckle. The first engineered ZF proteins were based on Cys2His2

zinc finger domains. Each of the finger domain is composed of approximately 30 amino acids and recognizes a specific consecutive tri-nucleotide DNA sequence (nucleotide triplet) in a modular fashion. The principle of DNA recognition is that the trinucleotide on one DNA strand is recognized by one amino acid residue occurring before the ZF alpha helix and two amino acids within the ZF alpha helix, therefore to achieve different DNA-binding preferences, the composition of the amino acids of the alpha helix needs to be altered.

Chimeric ZF proteins were subsequently used as DNA-binding modules to modulate gene expression profiles. Engineered ZF proteins were fused to numerous activation (Figure 2) and repression domains (Beerli *et al.*, 1998), to generate ZF based artificial transcription factors (ZF-ATF). For instance ZF repressors were targeted to HIV promoter in order to reduce the replication of the virus. A successful result of 100-fold decrease in the virus replication was observed in studied cells (Segal *et al.*, 2004). In addition, ZF activators were shown to upregulate the expression of a γ-globin gene when targeted to its promoter region. This experiment demonstrated the therapeutic potential of ZF-ATFs in sickle cell disease and β-thalassemia, in which the γ-globin gene is silent (Graslund *et al.*, 2005). To date ZF-ATF were applied to various organisms to regulate gene expression (Bartsevich *et al.*, 2003; Gaj *et al.*, 2013; Gersbach *et al.*, 2014; Sanchez *et al.*, 2002; Sanchez *et al.*, 2006).

ZF DNA-binding modules exhibit numerous limitations including the lack of information about affinity of some of the zinc fingers to nucleotide triplets and difficulty of construction of engineered zinc finger arrays. It has been challenging to generate new specific ZF DNA-binding modules which limited its broad applicability in research.

Moreover this particular toolkit needs further investigation focused on the potential toxicity to the cells and potential of off-targeting occurrence (Gersbach *et al.*, 2014). Therefore, there was a need to develop a more reliable and efficient genome engineering platform.

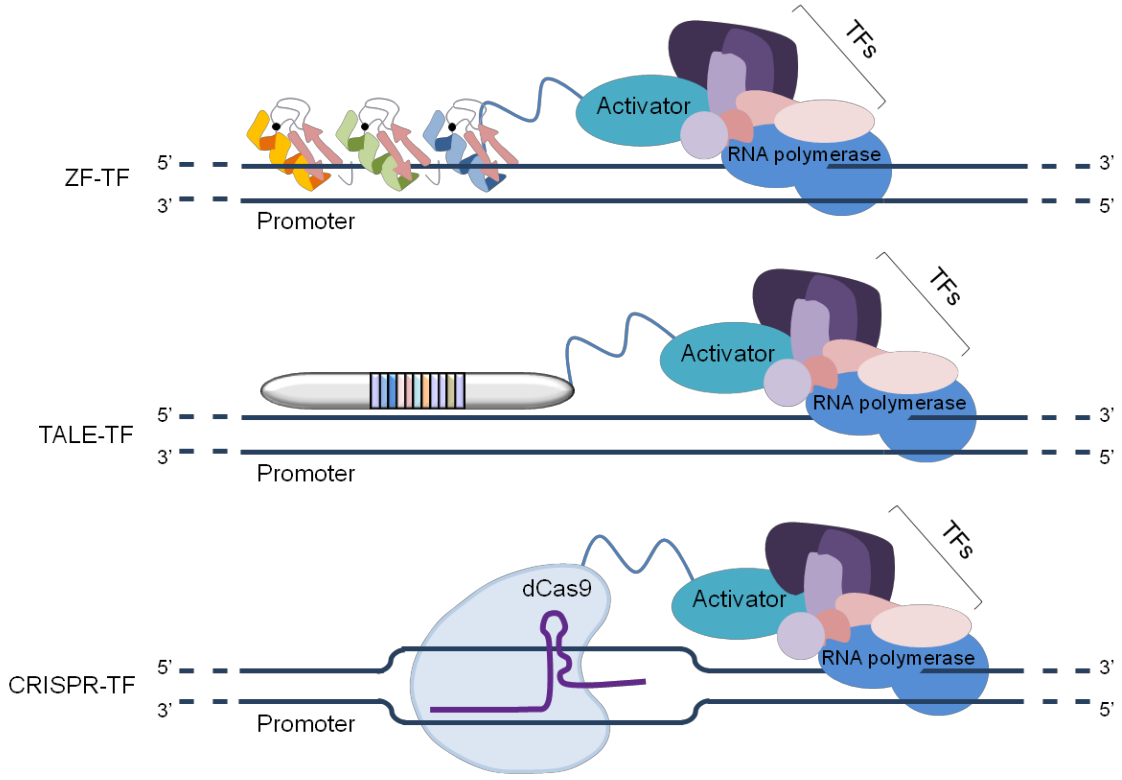


Figure 2. Targeted transcriptional activation with synthetic activators. Synthetic ZF proteins with activator. ZF recognize DNA through zinc finger arrays. Activation domain is fused to the protein through a linker sequence. Zinc finger transcription activators have been used to selectively activate genes. TALE based synthetic transcriptional activators. The specificity of DNA-binding is mediated by the central DNA-binding repeat domain. TALE protein is a natural transcription activator and has its own activation domain, however fusions with synthetic domains have also been reported for targeted genome regulation. CRISPR/Cas9 based synthetic activators. Catalytically inactive Cas9 (dCas9) is guided to the 20 nucleotide DNA target by a synthetic sgRNA molecule. Various strong natural activators may be fused to dCas9, such as VP64, p65 or EDLL domains. Activation domains help harvest different transcription factors from nucleus environment to improve transcriptional activation of the target gene.

1.2 Synthetic TALE Transcription Factors

TALEs are a class of DNA-binding proteins that were first discovered in *Xanthomonas* bacteria. It was documented that upon infection *Xanthomonas* injects TALEs to the host plant via the type III secretion system (Boch and Bonas, 2010). TALEs are localized to the host's nucleus, where host gene transcription is reprogrammed in favor of the pathogen, facilitating pathogenic spread and the development of the disease state.

TAL effector proteins have distinct structural features, which can be divided into 3 major segments: the N-terminus of the protein, ranging from 283 to 290 amino acids, the central repeat domain, and the C-terminus of the protein, ranging from 274 to 297 amino acids (Figure 3). The N-terminal and C-terminal regions are highly conserved in TALs, however at times slight variations are seen between different species. The N-terminus includes the secretion and translocation signal for the type three secretion system (T3SS), which facilitates the movement of the proteins from bacteria to plant cells. The C-terminus of the protein carries nuclear localization signals (NLSs), which mediate the protein localization to the nucleus of the plant cell. Activation domain (AD), placed downstream of the NLSs, is involved in recruiting the host’s transcriptional machinery and activating transcription, by mimicking the eukaryotic transcription factors. The most distinct feature of TALEs is their central repeat domain that consists of tandem nearly identical repeats each composed of 34 to 35 amino acids. The residues at positions 12 and 13 constitute the repeat variable diresidue (RVD). TALE binding to DNA is mediated by the repeat regions through their RVDs (Boch and Bonas, 2010). It was reported that a minimum of 6.5 repeats is necessary to induce transcription and 10.5 repeats are sufficient for full induction of the target genes (Romer *et al.*, 2007). The repeat region

ends with a truncated repeat, containing 20 amino acids (including the RVD), which is commonly referred to as a 'half repeat' in TAL effectors. The RVD to DNA recognition code has been determined by experimental and computational approaches (Boch *et al.*, 2009; Moscou and Bogdanove, 2009).

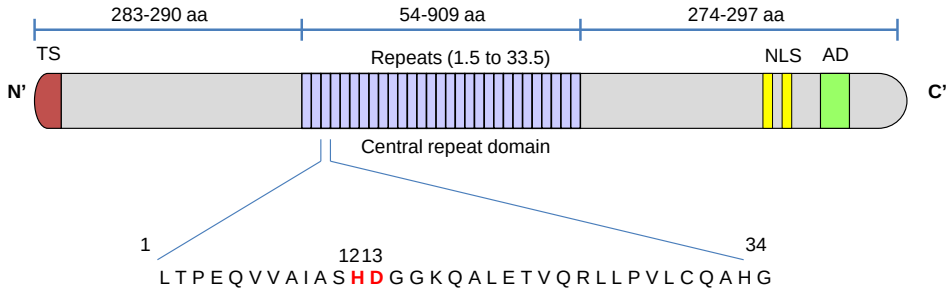


Figure 3. Structure of TAL effector. TAL effector consists of N-terminal segment including translocation signal (TS marked in red), central repeat domain composed of tandem, virtually identical repeats; nuclear localization signal (NLS marked in yellow); transcriptional activation domain (AD marked in green).

According to these findings, the code of RVD sequence to nucleotide is as follows: HD (histidine and aspartic acid) binds to the cytosine (C) nucleotide; NI (asparagine and isoleucine) binds to the adenine (A) nucleotide; NK (asparagine and lysine) binds to the guanine (G) nucleotide; NG (asparagine and glycine) binds to the thymine (T) nucleotide; NN (asparagine and asparagine) binds to A or G nucleotide and NS (asparagine and serine) could bind to the A, C, G or T nucleotides with equal affinity (Figure 4).

A better understanding of the DNA recognition code has been provided by the structural studies of TAL effectors (Deng *et al.*, 2012a; Mak *et al.*, 2012). Both groups have demonstrated that TAL repeats form a left-handed, two helix structure connected by a short loop where the two RVDs at positions 12 and 13 are positioned. All the repeats self-associate forming an overall right-handed superhelix which makes contact with major

groove of the target DNA site. Another significant finding communicates that the first residue (12th position in each repeat) has a stabilizing and structural role in the complex through formation of hydrogen bonds between repeats, and the second residue (13th position in each repeat) determines the sequence specificity.

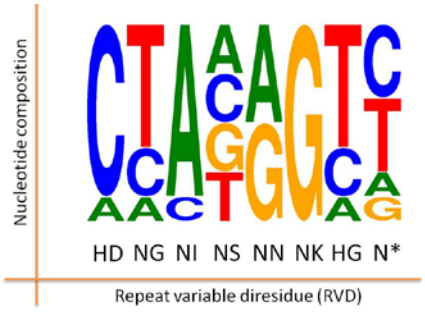


Figure 4. TAL effector recognition code. Relation of RVD to nucleotide. An asterisk indicates missing residue at position 13, resulting half repeat.

Studies of naturally-occurring TALEs in plant cells indicate that the DNA strand and DNA-binding site are key factors in gene activation. For instance, natural TALEs work in a directional manner by binding to a target box in the promoter of targeted genes. This shifts the transcription start site downstream of the binding site and activates transcription through recruitment of the transcriptional initiation complex (Romer *et al.*, 2009). Several TALE-based transcriptional regulators were developed in a variety of organisms. TALE repressors successfully suppressed endogenous gene expression in mammalian cell lines (Cong *et al.*, 2012) and in plants (Mahfouz *et al.*, 2012). To engineer an artificial transcriptional repressor, the DNA-binding domain of the TALE protein is fused to a repression domain that recruits heterochromatin-forming complexes. The Krüppel-associated box (KRAB) repressor domain occurs naturally in many eukaryotic TFs and is the preferred fusion domain for use in mammalian studies (Garg *et al.*, 2012). In plants, targeted repression of transgenic and endogenous RD29A expression in *A. thaliana* exposed to abiotic stress, such as abscisic acid and low temperatures, was achieved using

a TALE protein coupled to the ERF-associated repression domain SRDX (Mahfouz *et al.*, 2012). Similar synthetic TAL-based transcriptional activators were able to effectively induce the expression of several endogenous genes (Bultmann *et al.*, 2012; Cong *et al.*, 2012; Maeder *et al.*, 2013b; Perez-Pinera *et al.*, 2013b). However, the level of expression induction varied widely. For instance, activation of the human Oct4 locus with synthetic TALE yielded modest changes in gene expression level, suggesting that a single TALE might not always be sufficient to induce the desired enhancement in expression. In addition, as a result of epigenetic effects, different genetic loci are unlikely to be equally accessible to TALE proteins (Zhang *et al.*, 2011a). When TALE activators were co-delivered with chemical inhibitors of epigenetic modifiers to bypass the effects of chromatin state, targeted transcriptional activation of epigenetically-silenced target genes was observed (Bultmann *et al.*, 2012). Using multiple TALE proteins to target different regions of a single endogenous promoter (at variable vicinity from transcription start site) also enhanced gene activation (Maeder *et al.*, 2013b; Perez-Pinera *et al.*, 2013b). Use of several TALE activators is now preferred for the strong induction of gene expression; however, simultaneous delivery and expression of multiple proteins is challenging.

TALE proteins customization seems to have obvious advantage over ZFs due to the precise TALE DNA-binding specificity. Not only TALE proteins have predictable sequence specificity, but also can be easily designed to match long DNA target sequences. However, using TALE-ATFs requires complete redesign of the protein itself for each new user-defined DNA site. The must to redesign the binding domain of the TAL protein is challenging because TALE arrays are very similar in composition. In

addition, other limitations include unpredicted off-target binding due to repeat-context effects and the requirement to express large amounts of proteins for every experiment.

1.3 Targeted Genome Editing

Historically, genetic material has been altered by conventional cross-breeding of various plants or selecting the best animals (in terms of desired traits) for reproduction. Later genome editing technology came to play a role in modifying species' genes in order to introduce novel traits. Earlier techniques for genome modification included introducing mutations to the DNA by: irradiation (Russell *et al.*, 1958), chemical based mutagenesis (Neet and Koshland, 1966), or delivery of a gene into a living cell by transgenesis (Gordon *et al.*, 1980). These techniques had numerous limitations including the lack of reproducibility, random nature of insertion, or disturbance of other functional genes. In 1980's scientists Mario R. Capecchi, Oliver Smithies and Martin Evans had been working on developing a technique called gene targeting, which enabled modification at a specific DNA sequence. A process of homologous recombination was utilized as the delivery tool for gene targeting (GT). This technology, based on the introduction of a homologous sequence located on another strand of DNA, was used as a model for targeted gene inactivation or modification, for which the aforementioned researchers were awarded with the Nobel Prize in Physiology or Medicine in 2007. Since then this technology gained increasing attention and importance for the production of transgenic organisms and for human gene therapy. Although gene targeting was very successful in various organisms, including: mouse embryonic stem cells (Capecchi, 1989), *Physcomitrella patens* (Schaefer and Zryd, 1997), *Drososphila melanogaster* (Rong and Golic, 2000), sheep (McCreath *et al.*, 2000) and human somatic cells (Hanson and

Sedivy, 1995), it remained problematic in other organisms, including plants. Nevertheless, many attempts were taken to succeed with gene targeting in flowering plants, with first reports coming from Paszkowski (Paszkowski *et al.*, 1988). However, this technology was still inefficient and offered limited reproducibility. Over the years, numerous trials of GT mediated modification of endogenous genes have surfaced (Lee *et al.*, 1990; Miao and Lam, 1995) including one example of an *Arabidopsis thaliana* plant with an anticipated gene disruption (Kempin *et al.*, 1997).

Nowadays modern genome editing technologies (ZF, TALE and CRISPR) offer user-defined DNA sequence modification, nearly with no restriction to the chosen DNA site. In order to alter the genome at defined positions, certain conditions need to be met, i.e. introduction of double-strand breaks at the DNA level (McMahon *et al.*, 2012). When double-strand breaks occur on the DNA, the cell mobilizes multiple repair pathways, including non-homologous end joining (NHEJ) and homology directed repair (HDR) (Figure 5).

These highly conserved pathways can be harnessed to generate defined genetic outcomes across broad range of cell types and species. In case of NHEJ repair pathway, broken DNA ends are re-joined, however random insertions or deletions (InDels) of nucleotides may occur thereby shifting the open reading frame of a gene (Lieber, 2010). Such indels may result in gene knockout. In the process of HDR where the broken DNA is substituted with a homologous template, sequence can be modified through insertion of desired encoded features to the original DNA template. The exchange of the material occurs between a foreign (donor) DNA molecule and a partially homologous sequence on the chromosomal (target) DNA of the target cell. HDR is a powerful process which has been

successfully used for gene targeting in mice (Hall *et al.*, 2009), fruit flies (Venken and Bellen, 2005) and human cell lines (Tenzen *et al.*, 2010). Whereas HDR is relatively easy to promote in other organisms, it is very difficult to trigger in plant species, because DNA repair is naturally achieved by NHEJ and not by HDR. Therefore, attempts have been made to develop methods for gene targeting in plants, for instance by enhancing the HR rate or by the induction of genomic double-strand breaks (DSBs) in the target cell genome, through rare cutting enzymes with exceptional features (Puchta, 2005; Puchta *et al.*, 1993). Technologies for manipulating genetic material at the DNA level depended primarily on the development of synthetic site specific nucleases (SSNs). Such nucleases are composed of an adaptable DNA-binding module fused to a catalytic nuclease domain.

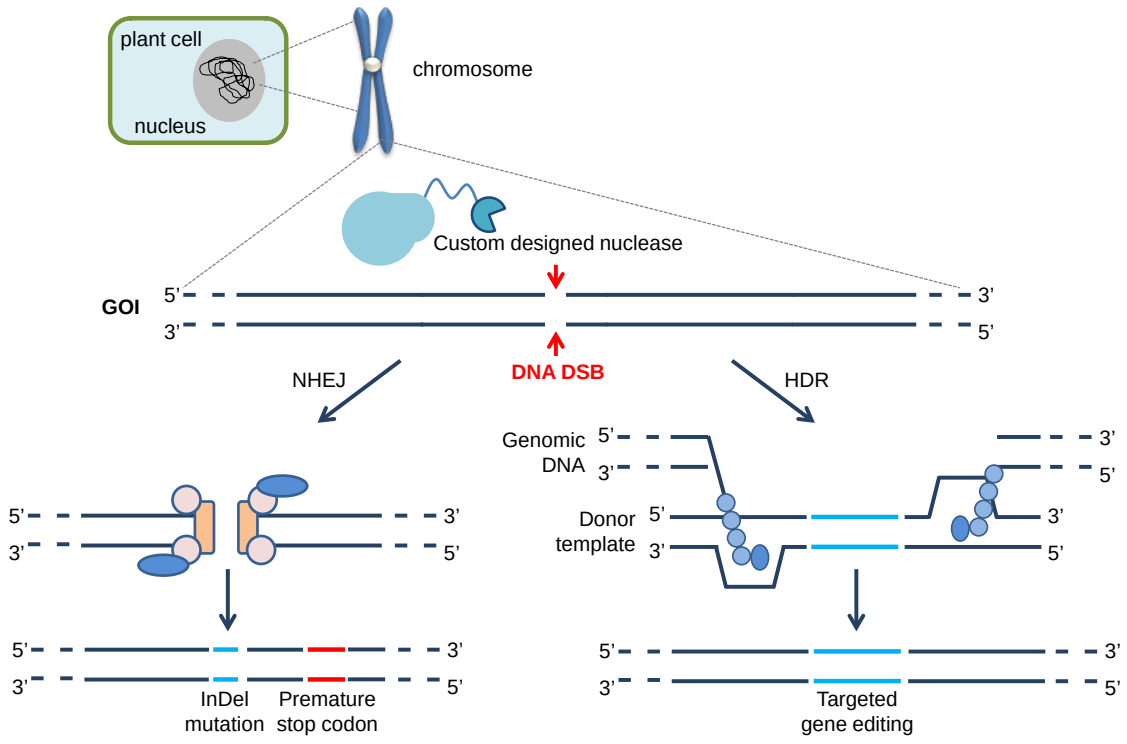


Figure 5. DNA repair mechanisms promote genome editing. DNA DSB is generated through synthetic, custom designed DNA-binding chimeric nuclease. NHEJ repair mechanism re-joins DNA broken ends introducing random InDels at the place of broken sites. Random InDels might result in open reading frame shift generating a premature

stop codon. HDR can be harnessed to edit a sequence of any given gene using a donor DNA template, resulting in gene editing. Figure adapted from (Ran *et al.*, 2013).

Several platforms of SSNs have been developed in the past years, including homing meganucleases, recombinases, and SSNs based on ZF proteins, termed zinc finger nucleases (ZFNs) and TALE proteins, termed transcription activator-like effector nucleases (TALENs). The efficiency of these platforms largely depends on the customizability of the DNA-binding module and its binding specificity. ZFNs and TALENs use the principles of protein to DNA recognition. In order to generate ZFN and TALEN a catalytic cleavage domain of a nuclease must be tethered to the ZF or TALE protein. The cleavage domain of the *Fok*I endonuclease has been preferably and widely used as the tethering agent. Summarizing, ZF and TALE provide the user-defined binding to the DNA through the DNA-binding domain and *Fok*I provides the catalytic activity resulting in DNA DSBs at the defined position. *Fok*I is a monomeric protein with one active site that is fused to the ZF or TALE protein through a linker sequence. To cut both strands of the DNA, the monomer at the target site needs to recruit a second monomer from the environment. Thus, the nature of the *Fok*I enzyme requires self-dimerization in order to create DNA DSBs (Bitinaite *et al.*, 1998). Dimerization involves two engineered nucleases to be present at the target site in a tail to tail orientation with a proper spacing between the two binding sites (Li *et al.*, 2011b). One nuclease will bind to one DNA strand and the other to an adjacent site on the opposite DNA strand. *Fok*I domain self-dimerization is a drawback. Not only does it require the optimization of the spacer length between two cleavage domains, but also the expression of two ZFNs or TALENs proteins. To bypass this, search for other alternative nuclease domains that facilitate the generation of monomeric protein based nucleases was proposed. Such monomeric

protein-based nucleases comprise a single polypeptide that is capable of generating a single- or double-strand break at the user-defined target DNA sequence. Such catalytic derivatives include the *I-Tev*I and *I-Sce*I nuclease domains that have been successfully used generate mutations at the DNA target site (Kleinstiver *et al.*, 2014; Kleinstiver *et al.*, 2012; Lin *et al.*, 2015). Yet such mutations often result in InDels and do not promote gene replacement or correction through delivery of DNA donor template. It has been proposed that separate nicks of single DNA templates might stimulate the HDR pathway. Therefore attempts have been made to develop DNA-binding proteins with nickases (Liu *et al.*, 2013; Ramirez *et al.*, 2012; Wu *et al.*, 2014). Taken together, the ability to perform targeted genome modifications can facilitate efficient and robust addition, deletion, activation and inactivation of genes. Such precise and efficient modification of the genome can for example: expand the plant species amenable to genetic modification, accelerate beneficial trait development and expand the range of traits.

In the mid-1990s the DNA-binding domain of ZF protein was coupled to the endonuclease domain of the *Fok*I restriction enzyme, generating ZFNs that were originally termed chimeric restriction enzymes (Porteus and Carroll, 2005). Two landmark publications demonstrated the utility of ZFNs for inducing targeted modifications to the fruit fly (Bibikova *et al.*, 2003) and human (Porteus and Baltimore, 2003) genomes. To date, ZFNs mediated gene disruption and correction have been shown to work effectively in other animal models, including rats (Geurts *et al.*, 2009), mice (Carbery *et al.*, 2010), and zebrafish (Meng *et al.*, 2008). The initial successes in plant genome engineering with ZFNs were reported in *Arabidopsis thaliana* (Zhang *et al.*, 2010a) and tobacco (Wright *et al.*, 2005). For example, *Arabidopsis thaliana* plants

insensitive to abscisic acid and glucose were generated by targeting the ABI4 gene with ZFN (Osakabe *et al.*, 2010). Tobacco herbicide-resistant plants were produced by targeting the acetolactate genes (Townsend *et al.*, 2009). ZFN were also applied to mutagenise the genome of crop plants including maize and soybean, resulting in heritable mutations of target genes (Curtin *et al.*, 2011; Shukla *et al.*, 2009). In addition, human clinical trials have been initiated by Sangamo BioSciences (a company that develops zinc-finger-protein-based therapeutics) who uses ZFNs to target the cell surface receptor CCR5 to prevent HIV-1 infection (CCR5 is an HIV-1 co-receptor) (Tebas *et al.*, 2014).

1.4 TALENs

The decoding of the TALE DNA recognition mechanism coughed the attention of genome engineers who recognized its potential for biotechnological applications, leading to the generation of TALENs. TALENs are fusion proteins that contain a nuclease cleavage domain of the *Fok*I restriction enzyme in addition to the TALE DNA-binding module. As mentioned above, *Fok*I needs to dimerize in order to generate a DNA break. Dimerization involves two engineered nucleases to be present at the target site (Figure 6). TALENs have been used to generate targeted modifications in a variety of organisms including yeast, fruit fly, *C. elegans*, zebrafish, rats, mice and human cell lines. Target-specific mutations have also been introduced in plants using TALENs.

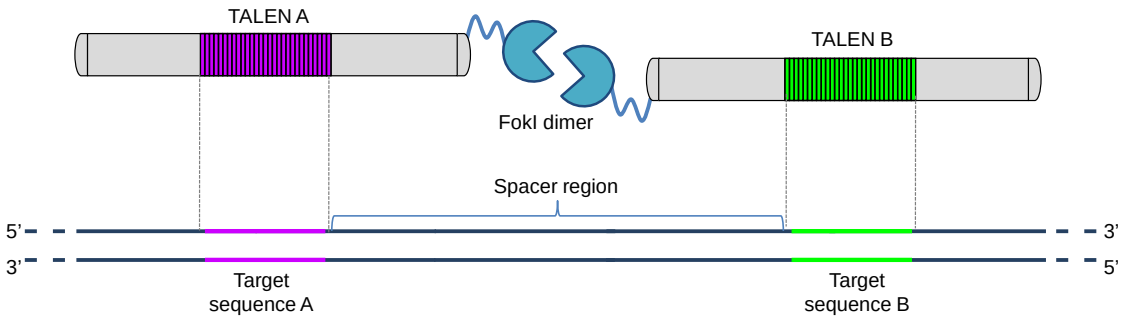


Figure 6. TALE effector nuclease. Chimeric TALEN composed of TALE DNA-binding domain (marked in purple and green) and *Fok*I nuclease domain needs to dimerize with its counterpart in order to generate DNA DSB. TALENs are custom designed to target any used defined DNA sequence. Two DNA target sites are separated by a spacer.

As a proof of principle *Arabidopsis thaliana* protoplasts were targeted with specific TALEN pair to mutate the ADH1 gene (Cermak *et al.*, 2011). Next, *Nicotiana benthamiana* genes were modified in transient assays (Mahfouz *et al.*, 2011) and *Arabidopsis thaliana* stable transgenics were recovered indicating that the mutation can be carried to the next generation of plants (Christian *et al.*, 2013). TALENs have also been used for application in crops. A disease resistant rice was generated through targeted mutagenesis in the OsSWEET14, that is a sucrose efflux transporter, a disease susceptibility gene. Moreover, in this experiment scientists were able to breed out the selection marker gene along with the TALEN integrated gene, essentially generating plants with targeted mutation only, making it a candidate for non GMO plant (Li *et al.*, 2012). Large genomic deletions and big scale targeted mutagenesis were also reported in rice plants using different sets of TALENs (Shan *et al.*, 2013a). In addition TALEN technology has been applied to other crops including barley (Gurushidze *et al.*, 2014; Wendt *et al.*, 2013) and maize (Liang *et al.*, 2014).

1.5 CRISPR/Cas9 System For Targeted Genome Editing

In 2012, an engineered version of the bacterial native defense system, clustered regularly interspaced short palindromic repeats and Cas associated proteins (CRISPR/Cas9) system was developed as the first platform for targeting proteins to DNA target sites through RNA-DNA interactions, rather than protein-DNA interactions. CRISPR system was first observed in *E. coli* in the late 1980s (Ishino *et al.*, 1987). The CRISPR/Cas9 system provides bacteria and archaea with an adaptive molecular immunity against invading phages and plasmids (Barrangou *et al.*, 2007; Horvath and Barrangou, 2010). The CRISPR/Cas9 system is encoded by as many as 40% of bacteria and 90% of archaea. The main components of the natural CRISPR/Cas locus are: an operon of Cas genes, the CRISPR array, and non-coding RNA molecules (Figure 7).

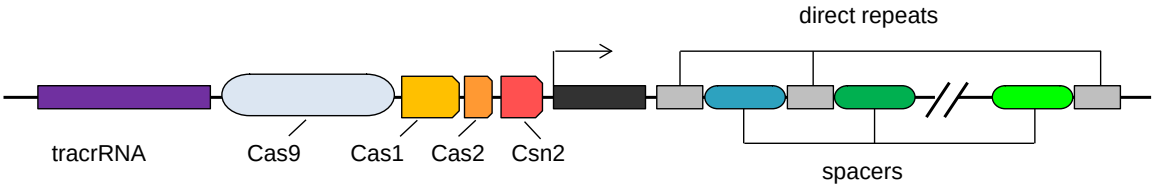


Figure 7. Schematic representation of Type II CRISPR locus. CRISPR locus is composed of four major segments, the tracrRNA, Cas associated genes, leader sequence and the CRISPR array.

The Cas operon represents CRISPR associated genes that encode a family of endonuclease proteins. This physical connection between repeat arrays and Cas genes led to the suggestion that they are co-functional (Jansen *et al.*, 2002). To date, over 50 different CRISPR-associated gene families have been identified which are generally located close to CRISPR arrays. Developed classification of CRISPR-Cas systems divides them into three types - I, II and III. Although, recent new classification distinguishes CRISPR Class I that comprises type I and III and Class 2 that includes type

II (Makarova *et al.*, 2015). Each of the types substantially differs in their sets of component genes, and each is characterized by a distinctive signature gene. The signature gene for the type I CRISPR-Cas array is Cas3 (Sinkunas *et al.*, 2011). Another signature protein Cas9 belongs to type II. It's a large protein containing RuvC-like and HNH nuclease domains and originated from the *Streptococcus pyogenes*. Finally, Cas10 belongs to the type III and is considered as a protein containing a domain homologous to the palm domain of nucleic acid polymerases and nucleotide cyclases (Makarova *et al.*, 2011). The CRISPR array is a stretch of identical repetitive sequences (short directed repeats) intertwined with unique DNA fragments called spacers. Directed repeats vary greatly in length and sequence composition between different organisms and exhibit a partial dyad symmetry, implying the formation of a secondary structure such as a hairpin in RNA transcripts, however they are not truly palindromic (Kunin *et al.*, 2007). The spacers are comparable in length but have highly variable sequences (Karginov and Hannon, 2010). With the development of sequencing technology and computational analysis, the CRISPR spacers were reported to match the sequences of various phages and plasmids. Such finding suggests that exogenous DNA is incorporated into bacteria's own genome in the form of spacers, implying that new spacers can be added fast in response to phage infection. Spacers, encode small RNA molecules that act as specific guides to target complementary DNA sequences (Gasiunas *et al.*, 2014).

The proposed mechanism of CRISPR-Cas system describes that exogenous DNA (foreign phage's or plasmid's DNA) is processed by bacteria's proteins (encoded by CRISPR-associated genes) into small elements, which are subsequently inserted into the CRISPR locus near the leader sequence. Various CRISPR-Cas types recruit different

proteins to perform this step. Next, a combination of small interfering RNA is produced after the transcription of the CRISPR array. Transcription of a CRISPR array results in a precursor RNA (pre-crRNA), followed by a maturation step in which the pre-crRNA is diced into smaller CRISPR RNAs (crRNAs). Such small interfering RNA molecules are then congregated with present Cas-encoded machinery (comprising proteins of helicase and nuclease activities) to specifically target and destroy nucleic acid exhibiting sequence complementarity to the spacers (Figure 8).

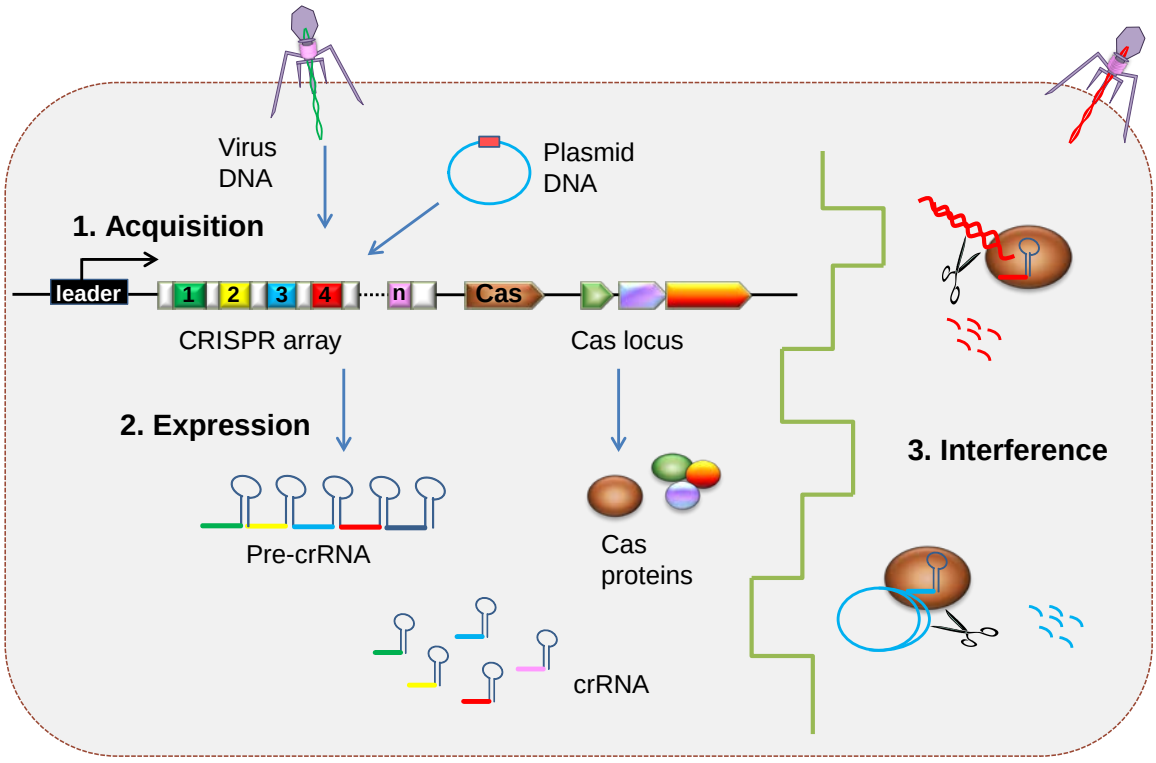


Figure 8. Schematic representation of CRISPR-Cas mediated interference. Foreign DNA might be injected into bacteria's cell through phages or enter the cell in a form of plasmid. Such extrachromosomal DNA is processed by internal CAS proteins and incorporated into the genomic CRISPR locus as a spacer (indicated with green, yellow, blue and red colors in CRISPR array). Spacers are intertwined by repeats (marked in white). The repeat/spacer region is flanked by a 'leader' sequence and by Cas locus. Next, in the expression phase, the CRISPR array is transcribed as a long pre-crRNA and further processed to mature, short crRNA. In any subsequent event of foreign DNA invasion, the

role of crRNA, in the step of interference, is to guide the Cas9 nuclease to invading foreign nucleic acid, resulting DNA sheering and infection halt.

It has been shown that Cas9 is directed to the specific DNA sequence by a short guide crRNA, and a non-targeting RNA called tracrRNA (Jinek *et al.*, 2012). These two small RNA molecules, crRNA and tracrRNA, were synthetically combined to generate a single-guide RNA (sgRNA) capable of directing the Cas9 protein to its genomic target in a site-specific manner (Mali *et al.*, 2013b) (Figure 9).

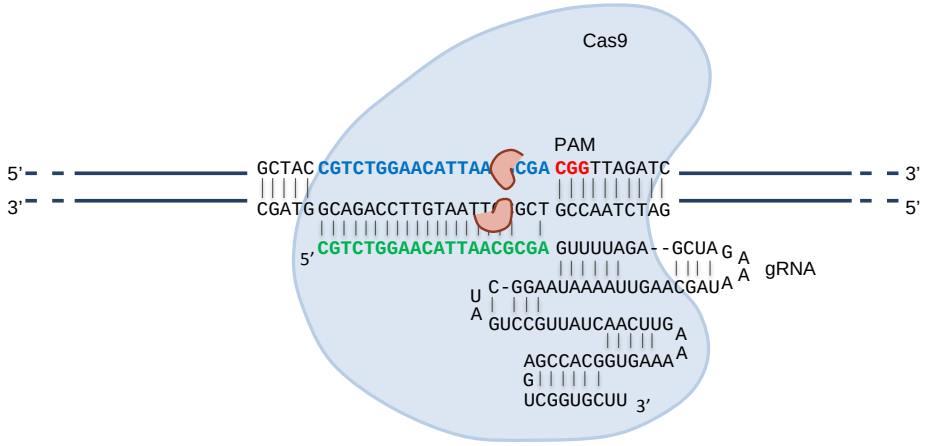


Figure 9. Targeted genome editing with RNA-guided Cas9. Cas9 protein comprises two active sites RuvC and HNH (marked in dark pink), that each cleave one strand of a double-stranded DNA molecule. Cas9 is guided to the target DNA (sequence in blue) by a gRNA molecule that contains a complementary spacer sequence (sequence in green). Three nucleotides, NGG (N represents any of the four nucleotides and G represents guanine), PAM sequence (sequence in red) located downstream the target site is a prerequisite for DNA cleavage.

In order for the DNA cut to be made through the CRISPR/Cas9 system certain conditions need to be met. First, the ribonucleoprotein complex (Cas9 protein and the sgRNA) moves along DNA and scans for a protospacer adjacent motif (PAM), and for the specific DNA target sequence. The PAM sequence is located upstream of the target site and is essential for proper functioning of the CRISPR complex. PAM usually encodes a

trinucleotide NGG sequence, where N represents any of the four nucleotides and G indicates guanine. After recognizing the PAM sequence, the Cas9/sgRNA complex binds to the DNA target via Watson and Crick complementary base pairing. If the complex finds a PAM sequence but the DNA/RNA complementarity requirement is not met, the Cas9/sgRNA complex continues to scan. Likewise, the complex continues to scan if complementarity is found but there is no PAM sequence preceding the target. When the two conditions are fulfilled, the Cas9/sgRNA complex stops scanning and generates a DNA DSB at a gene of interest (Figure 9).

CRISPR/Cas system has proven to be highly effective for genome editing. Exchanging the 20 nucleotides in the sgRNA can target the system to virtually any place on the DNA. CRISPR/Cas9 system has been harnessed to induce mutations across prokaryotic (Jiang *et al.*, 2013a) and eukaryotic species, including yeast (DiCarlo *et al.*, 2013), zabrafish (Chang *et al.*, 2013; Hwang *et al.*, 2013), mice (Shen *et al.*, 2013; Wang *et al.*, 2013) and human cell lines (Mali *et al.*, 2013b). A powerful advantage of this system is that one can modify multiple target sites simultaneously in one experiment approach. The method of using multiple sgRNA with different target sequences concurrently, referred to as multiplexing, has proven to be successful (Cong *et al.*, 2013; Mali *et al.*, 2013b). Genomes of plant species have also been modified using CRISPR toolkit. Model plant species such as *Arabidopsis thaliana* and *Nicotiana benthamiana* were the first choices to demonstrate the potential of the CRISPR system to introduce DNA DSBs (Feng *et al.*, 2013; Jiang *et al.*, 2013b; Li *et al.*, 2013a; Mao *et al.*, 2013; Nekrasov *et al.*, 2013). First *Arabidopsis* protoplasts (Li *et al.*, 2013a), tobacco (Nekrasov *et al.*, 2013) and rice plants (Shan *et al.*, 2013b) were modified using transient assays. Next, germinal transmission of

target mutation, through multiple generations, was shown in *Arabidopsis* (Feng *et al.*, 2014; Jiang *et al.*, 2014). Finally, CRISPR system was successfully applied to crop species including rice (Feng *et al.*, 2013; Jiang *et al.*, 2013b; Miao *et al.*, 2013; Shan *et al.*, 2013b), maize (Liang *et al.*, 2014), sorghum (Jiang *et al.*, 2013b), soybean (Jacobs *et al.*, 2015), wheat (Upadhyay *et al.*, 2013; Wang *et al.*, 2014b) suggesting the potential of this technology in production of enhanced staple foods. Collectively the data provided insight into efficacy of the CRISPR system in plant species, its cleavage specificity and efficiency, rate of mutations, and possibility of large chromosomal deletions. Hints on insertions of donor sequences into plant genomes through HDR with help of CRISPR toolkit were also reported (Schiml *et al.*, 2014).

Many transiently expressed targets and endogenous genes were chosen to be modified with the CRISPR/Cas9 technology as a proof of principle that the system is viable in plant species. Now the CRISPR/Cas9 toolkit is applied to help answer important biological questions in basic research and used for application in many fields including agronomy. For instance, CRISPR/Cas9 and TALEN technologies have been successfully used to generate a certain wheat strain resistant to a fungal pathogen, namely the powdery mildew. As wheat has a hexaploid genome, all three copies of genes responsible for the susceptibility to the disease had to be inactivated in one plant in order for the wheat plant to resist the disease (Wang *et al.*, 2014a). In another example, our research group reported that CRISPR/Cas9 can be used in plants to confer resistance to infectious DNA viruses, block further infection and suppress disease progression. Ali *et al.* engineered *Nicotiana benthamiana* plants resistant to tomato yellow leaf curl virus (TYLCV), that causes major destruction of tomato crops worldwide (Ali et al., 2015a; Ali et al., 2015b).

1.6 Introduction to the Chapters, rationale and motivation

In this dissertation I use two most recently discovered, state of the art genome engineering technologies, TALE proteins and CRISPR/Cas9 system, to induce targeted genome modification through introduction of mutations at specific DNA target sites; and targeted genome regulation through control of gene expression patterns in plants. In subsequent chapters I studied novel DNA-binding domains from *Ralstonia* sp., developed novel methods of TALE proteins assembly, repurposed CRISPR/Cas9 technology for targeted genome regulation and applied CRISPR/Cas9 to generate comprehensive SR protein knockout mutant collection in rice. Below I summarize my research in a nutshell and explain the rationale and motivation behind each project in individual chapter.

Chapter 2: "Characterization and DNA-Binding Specificities of Ralstonia TAL-Like Effectors."

The TALE code from *Xanthomonas* has been well recognized by the time the study on RVDs originating from *Ralstonia* sp. was undertaken. The RVD to DNA relation has been deciphered and *Xanthomonas* RVDs began to be applied in custom design of DNA-binding modules of TALE proteins and used in first proof of principle genome editing studies. Soon it was reported that certain RVDs express weak DNA-binding and certain RVD compositions, (within DNA-binding domain) may not be functional. Also, it was not clear whether RVDs from other species can be transferable and functional in other organisms. Therefore, this prompted us to look for new RVD specificities in other species than *Xanthomonas*.

We chose to characterize novel RVDs from *Ralstonia solanacearum* sp. because homologs of TAL effectors have been identified in this bacteria and because *Ralstonia* is a phytopathogenic bacterium with host range of over 200 species, causing major economical loss.

Our findings were the first to advocate that *Ralstonia* TAL like proteins (RTL proteins) localize in the plant cell nucleus, mediate DNA-binding, and most importantly function as transcriptional activators, suggesting they play a major role on the pathogenicity of this bacteria. We have identified new RVDs and deciphered their DNA-binding specificities. Also, we proved that the new RVDs are functional when transferred to TAL proteins architectures originating from *Xanthomonas*, providing greater specificity and flexibility for biotechnological applications in agriculture. This research was the first report on *Ralstonia* originating RVDs and first to suggest commercially efficient delivery platform for novel and newly synthesized DNA-binding units. The importance of this research has been recognized and further studied by other groups (de Lange *et al.*, 2013).

Chapter 3: "Rapid and Highly Efficient Construction of TALE-Based Transcriptional Regulators and Nucleases for Genome Modification."

At the time this study was undertaken, although TALE proteins have been widely accepted and commonly used as genome engineering tools across different species, there was no quick, efficient and affordable TALE assembly strategies. The problem was that TALE proteins were/are very difficult to engineer due to repetitive nature and nearly identical sequence of TAL repeats in the DNA-binding domain. The synthesis of the designer TALEs has been shown previously with PCR-based methods coupled to Golden

Gate cloning or Golden Gate cloning only. However, this strategies were laborious, resource-intensive, often error prone and required several restriction and ligation steps using various backbone vectors (Morbitzer *et al.*, 2011; Weber *et al.*, 2011; Zhang *et al.*, 2011a).

We sought to establish a robust and reliable method of TAL effectors assembly that would facilitate the application of TALE-based genome modifications in various organisms including plants. Our approach was the first in the field to assemble repeat domains of various TAL effectors for any 14-nucleotide DNA target sequence in one sequential restriction-ligation cloning step and in only 24 h. At the time of publishing the results our design was the fastest and the easiest method of assembly. Our approach was based on a library of 100 plasmids carrying multiple combinations of RVDs designed and synthesized to facilitate cloning of the repeats to any user-selected recognition sequence. The final product of the reaction contained DNA-binding domain comprising fragments ligated to a backbone of choice. We have successfully tested various custom designed DNA-binding arrays through our method and fused them in backbones generated for constructing transcriptional activators (dTALE-TFs) and chimeric nucleases (dTALENs).

Chapter 4: "RNA-Guided Transcriptional Regulation in Plants via Synthetic dCas9 Based Transcription Factors."

In the last three years CRISPR/Cas9 technology has become an attractive alternative to DNA-binding modules of TALE and ZF proteins due to its simplicity of engineering and broader range of DNA targeting. In the same way as CRISPR/Cas9 technology has been eliminating TALENs, CRISPR/dCas9 system started to become a substitute for targeted

genome regulation using TAL effectors. Reports in the literature prompted the use of a catalytically inactive Cas9 nuclease (dCas9), that rendered the DNA-binding capabilities but lacked its catalytic activity. This meant that dCas9 was no longer able to cut DNA but was still able to bind it when guided by the short gRNA molecule. Studies reported comparable or better performance of dCas9 system to TALE based TF in terms of control of gene expression levels.

This encouraged us to develop dCas9 based targeted transcription regulation system that can be applied in plants and was not reported previously. Further reason for advancing and implementing CRISPR/dCas9 technology in plants was that we recognized the problem that custom TALE-TF face, namely the requirement to redesign whole TALE protein for each new DNA target and the delivery method into plant species for targeting multiple DNA sequences.

We sought to generate artificial dCas9 based activators and repressors that could be used to modulate gene expression *in planta* in a proof-of-principle transient assays and also for endogenous genes. We demonstrated that dCas9 based effectors can robustly induce gene transcription or reduce gene expression with significant rates. We also showed that targeting dCas9 effectors with multiple gRNAs resulted in synergistic increase or decrease in target gene expression. This was the first evidence of using this targeted genome regulation technology in plants. Following our work others started using this approach in their research on model plants (Lowder *et al.*, 2015).

Chapter 5: "Understanding the Functional Roles of Rice SR Proteins Under Stress Conditions Using CRISPR/Cas9 System."

In the past years CRISPR/Cas9 tool has emerged as the genome engineering technology that revolutionized modern molecular biology. Indeed, the increasing list of applications of this technology spans many fields including human health, biotechnology, agriculture and in addressing basic research questions.

One of the challenges in plant genome engineering is to efficiently generate multiple knockout or knockdown mutants for functional genomics. In this study we harness the power of CRISPR/Cas9 technology to better understand the biological and functional role of one gene family with diverse molecular functions in rice, namely the Serine-Arginine (SR) rich protein family.

We undertook an arduous task of generating a large rice mutant resource of SR knockouts. We took advantage of the multiplexing feature of CRISPR/Cas9 technology to generate single and multiple SR knockouts.

We also harness the power of CRISPR/Cas9 technology to address basic biological questions regarding SR proteins. This protein family has diverse functions in different stages of RNA metabolism starting from expression to translation. SR proteins were shown to help in exon definition, enhance splicing, facilitate spliceosome assembly and counteract the effects of hnRNPs. In addition, SR proteins are post-translationaly modified by protein kinases and their phosphorylation status determines their functional roles. SR proteins were shown to be alternatively spliced under different stress conditions. Our understanding of molecular underpinnings of SR proteins may help us to engineer plants for better and improved stress tolerance. To study their functional roles, we attempt to generate a suite of SR protein knockouts of the entire SR family in rice.

This would help us understand the plant specific proteins because SR protein family is extended in plants.

We reasoned that studying SR proteins through the mutant collection will shed more light on the splicing patterns of SR proteins themselves and their downstream targets. Moreover, we hypothesize that SR proteins that affect the splicing events of other genes might influence the response of the plants under different environmental pressures.

At the time of writing this thesis we transformed rice calli with CRISPR/Cas9 reagents, targeting single or multiple SR genes, and regenerated viable rice plants. In this chapter I present the results of Cas9 induced mutations at the SR target sites, indicating the success, efficiency and robustness of CRISPR/Cas9 system in rice plants.

CHAPTER 2

Characterization and DNA-Binding Specificities of *Ralstonia* TAL-Like Effectors

Reprinted with permission by Elsevier from:

Li, L., Atef, A., Piatek, A., Ali, Z., Piatek, M., Aouida, M., Sharakuu, A., Mahjoub, A., Wang, G., Khan, S., Fedoroff, N.V., Zhu, J.K. and Mahfouz, M.M. (2013) Characterization and DNA-binding specificities of *Ralstonia* TAL-like effectors. *Molecular plant* 6, 1318-1330. doi: 10.1093/mp/sst006.

2.1 Abstract

Transcription activator-like effectors from *Xanthomonas* sp. have been used as customizable DNA-binding modules for genome-engineering applications. *Ralstonia solanacearum* TALE-like proteins (RTLs) exhibit similar structural features to TALEs, including a central DNA-binding domain composed of 35 amino acid-long repeats. Here, we characterize the RTLs and show that they localize in the plant cell nucleus, mediate DNA-binding, and might function as transcriptional activators. RTLs have a unique DNA-binding architecture and are enriched in repeat variable di-residues, which determine repeat DNA-binding specificities. We determined the DNA-binding specificities for the RVD sequences ND, HN, NP, and NT. The RVD ND mediates highly specific interactions with C nucleotide, HN interacts specifically with A and G nucleotides, and NP binds to C, A, and G nucleotides. Moreover, we developed a highly efficient repeat assembly approach for engineering RTL effectors. Taken together, our data demonstrate that RTLs are unique DNA-targeting modules that are excellent alternatives to be tailored to bind to user-selected DNA sequences for targeted genomic and epigenomic modifications. These findings will facilitate research concerning RTL molecular biology and RTL roles in the pathogenicity of *Ralstonia* spp.

2.2 Introduction

Molecular tools that can be used to achieve precise, site-specific modifications of genomes are valuable for many basic research and biotechnological applications in agriculture and medicine. Developing genome-editing tools for targeted modifications of the genome requires the development of customizable DNA-binding modules that can be easily engineered to bind to any user-defined genomic sequence. These DNA-binding modules can be fused to functional domains to generate chimeric proteins for the desired type of genomic modification (Bogdanove and Voytas, 2011; Carroll, 2011). A group of TALE proteins have been previously shown to bind to DNA *via* a central DNA-binding domain of tandem repeats (Romer *et al.*, 2007).

Several other bacterial species contain TALE-like proteins or proteins with TALE-like structural features, particularly the repeat domain region (Bogdanove *et al.*, 2010). One of such bacteria include *Ralstonia solanacearum*. *Ralstonia solanacearum* is a soil-borne, gram-negative, pathogenic bacterium that colonizes the xylem of more than 200 species of plants (Genin and Denny, 2012). The vascular dysfunction provoked by this widespread colonization causes wilting symptoms and subsequent plant death (Genin and Denny, 2012). One of many *Ralstonia solanacearum* targets includes the plant of potato, which distinctly constraints the production rate in tropical and subtropical regions worldwide, causing the brown rot disease (Poueymiro and Genin, 2009). *Ralstonia solanacearum* type III effectors are required for its pathogenicity on *Arabidopsis* and tomato plants (Remigi *et al.*, 2011; Zhang *et al.*, 2011b). DNA sequences homologous to the *Xanthomonas* Avr*Bs3* TAL effectors were detected in strains of the *R. solanacearum* species complex (Heuer *et al.*, 2007; Salanoubat *et al.*, 2002). The repeats are composed

of 35 amino acids and are similar to the Hax2 TAL effector of *Xanthomonas campestris* pv. *armoraciae*. Not all *Ralstonia* species contain TAL effectors, which suggests that these effectors were acquired through horizontal gene transfer (HGT) (Heuer *et al.*, 2007). The DNA sequences of the repeats are prone to recombination, which might enhance the adaptive responses of the pathogen to different hosts through interactions with disease or resistance host regulators.

Here, we functionally characterize three *Ralstonia* TAL-like proteins as transcriptional activators and we determine the DNA-binding specificities for the RVD sequences ND, HN, NP, and NT. The RVD ND interacts with C nucleotides, HN interacts with A and G nucleotides, and NP interacts with C, A, and G nucleotides. Characterization of novel DNA-binding architectures and RVD DNA-binding specificities will provide us with great flexibility in the generation of specific targeting modules. Moreover, we demonstrate that the RTL proteins can be tailored to bind to user-selected DNA target sequences and can thereby be used to generate invaluable genome-engineering tools with a variety of potential biotechnological applications in agriculture and genetic medicine. Furthermore, our data represent a major step towards further molecular characterization of the DNA-binding activities of RTLs and their biological roles in the pathogenicity of *Ralstonia* spp.

2.3 Research Aim and Objectives

The aim of this study was to identify novel RVDs from other species than *Xanthomonas*, in order to facilitate the production of highly efficient TALE architectures for genome-engineering applications. Objectives of this study included: identification of TALE homologous proteins in *Ralstonia* sp.; determining the localization of RTL proteins in the

plant; establishing whether RTLs mediate DNA-binding; deciphering the DNA counterparts of newly described RVDs; establishing whether RTLs function as transcriptional activators and whether *Ralstonia* originating RVDs are transferable and functional in other species.

2.4 Results

2.4.1 Identification of *Ralstonia* RTLs

The *Ralstonia solanacearum* species complex possesses a high degree of genetic diversity and is divided into four phylotypes that reflect geographical distribution. The full genome sequence of seven strains, including the reference strain GMI1000, has been published (Remenant *et al.*, 2010; Salanoubat *et al.*, 2002) and genomic sequences of additional strains are accessible online. Because TALE proteins exhibit distinct features, including an N-terminal secretion signal, a C-terminal transcriptional activation domain, and an NLS and central DNA-binding repeat domain, the available *Ralstonia* proteomes were searched for proteins with features similar to those of TALE proteins. Because the presence of the central DNA-binding repeat domain is the most distinctive feature of the TALEs, a k-means-based algorithm was used to identify tandem repeats in *Ralstonia* proteomic sequences available in databases (Jorda and Kajava, 2009). Three proteins (CAQ18687.1, NP_519936.1, and YP_003750492.1) that appeared to be derived from full-length clones and to contain TALE-like structural features, including the N-terminus, the central DNA-binding repeats, and the C-terminus domain, were selected for further analysis. The general structural features of the RTL proteins and the alignment of the repeats, including the consensus among the three RTL proteins, are shown in Figure 10.

Repeats with a high similarity to the canonical central repeats were identified in the upstream and downstream regions flanking the modular repeat domain (Figure 10). These repeats, however, do not display the well-characterized RVDs, either because the 13th amino acid is missing or because the 13th position contains an amino acid that has not been previously shown to bind to DNA or to mediate protein–DNA interactions.

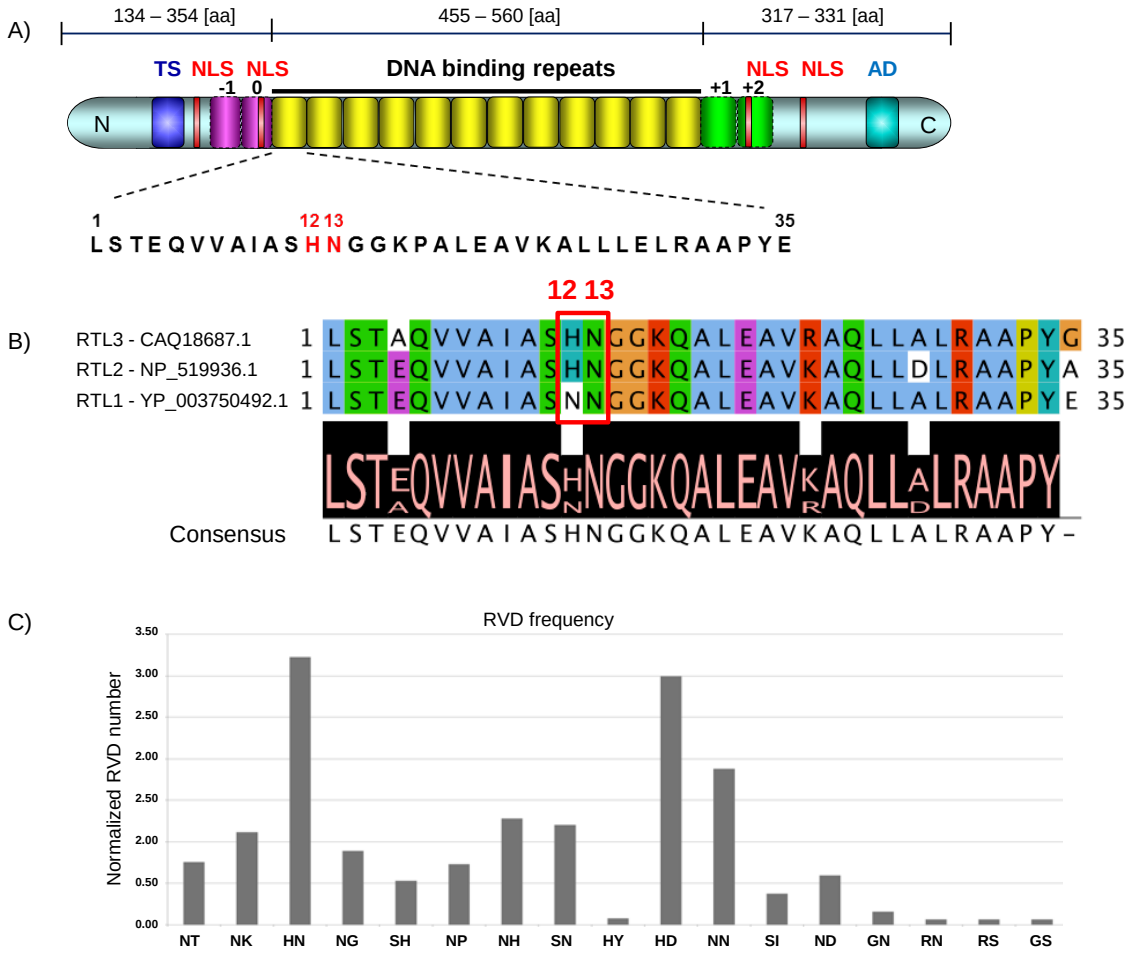


Figure 10. Structural Representation of RTL Effectors. (A) Structural features present in studied RTL effectors. A typical RTL effector contains a central DNA-binding domain consisting of 35 amino acid-long repeats (yellow boxes), putative –1 and 0 repeats upstream and +1 and +2 half-repeats downstream the DNA-binding domain (pink and green boxes), an acidic activation domain (AD, turquoise box), and a translocation signal (TS, blue box). The conserved RTL repeat sequence that was used for RTL repeats assembly is also shown. The repeat variable di-residues (RVD) are shown in red. (B) Alignment of RTL1 (YP_003750492.1), RTL2 (NP_519936.1), and RTL3 (CAQ18687.1) consensus sequence of

repeats. The consensus sequence of repeats among the three proteins is shown at the bottom of the panel below the alignment. The RVD is highlighted in the red box. (C) The frequency of the naturally occurring RVDs in RTL effectors. The 12th and 13th amino acids have been extracted from all identified RTLs containing canonical tandem repeats and their total number have been normalized per protein.

To determine the conservation level of the tandem repeats in the DBD region among all the RTLs, we performed a multiple sequence alignment using Jalview (Waterhouse *et al.*, 2009). The [LSTEQVVAIASXXGGKPALEAVKALLLELRAAPYE] consensus repeat sequence is highly conserved among all the reported RTLs. Putative functional domains also were identified, including a nuclear localization signal composed of short stretches of positively charged amino acids surrounding the modular repeat domain. Positively charged amino acids constitute a significant fraction of the sequence flanking the tandem repeat domain.

2.4.2 Design, synthesis and construction of RTL proteins

To characterize the RTL effectors and analyze their molecular function, we designed three artificial RTLs based on the amino acid sequences of the YP_003750492.1, NP_519936.1, and CAQ18687.1 clones. Codon-optimized versions of the clones, herein referred to as RTL1 (YP_003750492.1), RTL2 (NP_519936.1), and RTL3 (CAQ18687.1), were generated (Supplementary Sequence 1). Each RTL protein is composed of three major parts: the N-terminus, the C-terminus, and the central DNA-binding repeats. The backbone of each RTL effector was custom-synthesized as two DNA fragments that correspond to the N- and C-termini of the protein (Supplementary Sequence 2). The first fragment corresponds to the sequence of the N-terminus of the clone and was engineered to contain *Apa*I, *Nde*I, and *Spe*I (for RTL1) or *Sal*I, *Nde*I, and *Spe*I (for RTL2 and RTL3) restriction-enzyme recognition sequences at the 3’ end. The

second fragment sequence corresponds to the C-terminus of the clone and was flanked by *Nde*I and *Spe*I restriction-enzyme recognition sequences. For generation of the backbone clone (full-length minus the central repeat domain), the second fragment was digested with *Nde*I and *Spe*I and ligated into the first fragment, which had been digested with *Nde*I and SpeI; this produced the backbone clones RTL1ΔR, RTL2ΔR, and RTL3ΔR (Figure 11 A).

Two assembly approaches were used to generate RTL central repeat fragments. First, a protocol for an ordered assembly of RTL di-repeats was devised. Six fragments of di-repeats (210 bp) were designed based on the sequence that is conserved among the RTL effector repeats (LSTEQVVAIASXXGGKPALEAVKALLLELRAAPYE). All the di-repeat DNA fragments were synthesized in the pUC19 minus *MCS* vector with the proper flanking sequences. Fragment 1 was digested with *Apa*I or *Sal*I and *BsmA*I, fragments 2-5 were digested with *BsmA*I, and fragment 6 was digested with *BsmA*I and *Nde*I. The restricted fragments were gel-purified and ligated into an RTLΔR backbone clone digested with *Apa*I or *Sal*I and *Nde*I. A library containing fragments 1-6 with each of the desired RVD combinations was generated to allow the RVDs to be assembled in the desired sequence and order (Supplementary Sequence 3). The details of the ordered assembly of the di-repeats are provided in Figure 11 and Supplementary Information for Chapter 2.

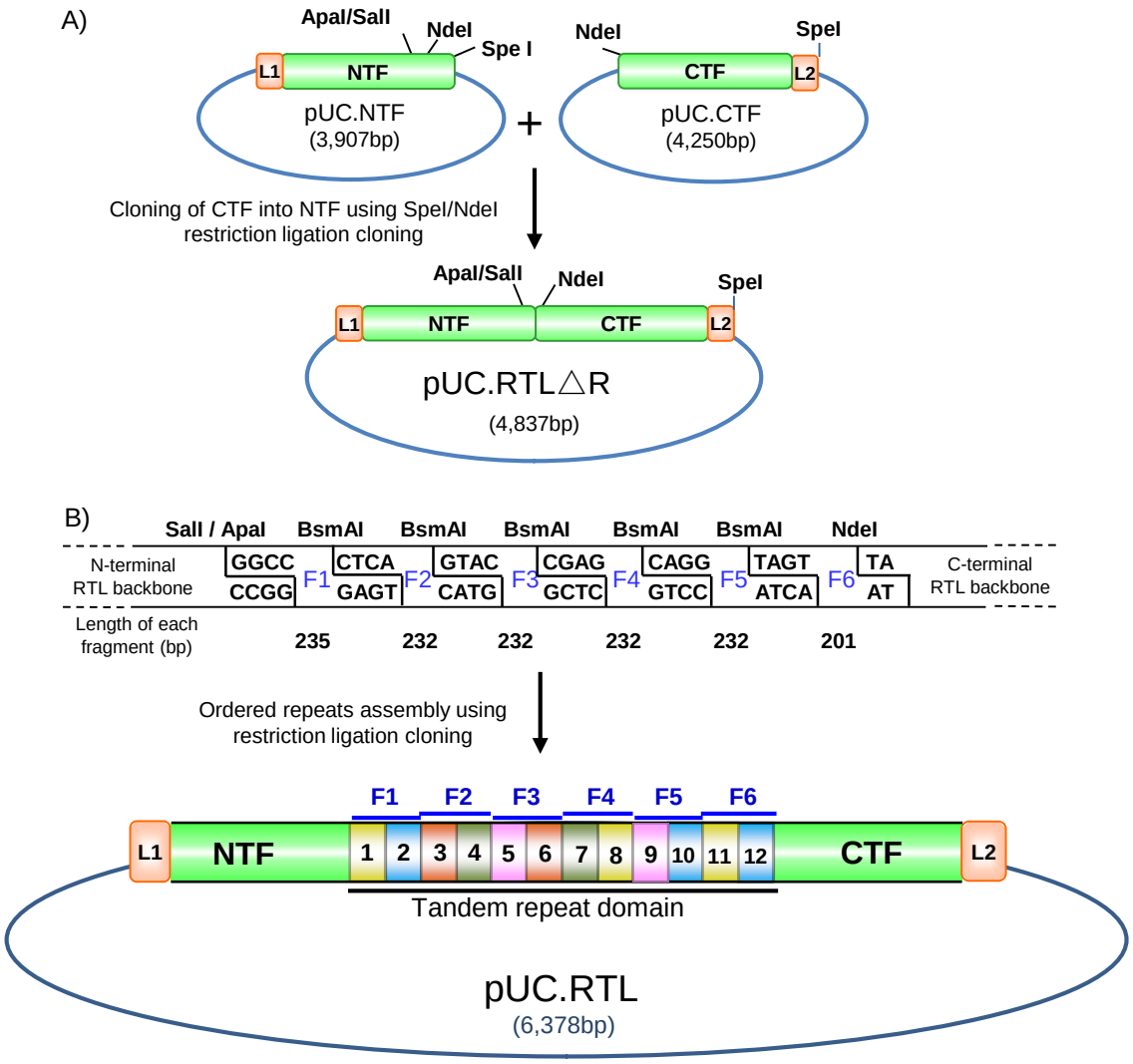


Figure 11. Design and Construction of RTL Effectors and Assembly of Their Repeats. (A) Each RTL backbone clone (RTL minus-repeats, RTLΔR) was custom-synthesized as an N-terminus fragment (NTF) and C-terminus fragment (CTF). The CTF was restricted with *Nde*I/*Spe*I and subcloned into the NTF clone to generate the RTLΔR clone. RTLΔR clones are flanked with L1 and L2 gateway recombination sites to allow cloning into desired destination vectors. (B) Schematic representation of the ordered assembly of RTLs repeats showing the six restriction fragments and the overhang sequence of each fragment produced by a *BsmA*I restriction digestion.

For the second assembly approach, a PCR-based procedure was developed to clone RTL pre-assembled repeats into the *Xanthomonas* backbone, as described below. The central repeat fragments were PCR-amplified using forward and reverse primers that contain

*Apa*I or *Sal*I and *Nde*I restriction enzyme recognition sequences. The fragments were then digested and ligated into the RTLΔR backbone clones to generate the full-length RTL DNA clones.

Determining whether RTL repeats (RRs) would recognize and bind to DNA in the *Xanthomonas* TAL backbone was essential. To study the DNA-binding specificities of the RRs in the *Xanthomonas dHax3* backbone, we designed three types of constructs (Supplementary sequence 4). The first construct contains part of the N-terminus of the *dHax3* backbone and the first repeat, the second construct type represents a single repeat, and the third construct contains the last half repeat and the C-terminus of the *dHax3* backbone. The repeats with identical or mixed RVD types in the *dHax3* backbone can be assembled by sequential restriction digestion, concatamerization of the second single-repeat type, and ligation of these three types of fragments to generate the desired repeat number and RVD order. Clones of *dHax3* containing repeats of variable length (ranging from 2 to 15 repeats) were selected by restriction and Sanger sequence analysis. These repeat concatemer clones were generated for all desired RVDs of the RRs. Using the two repeat assembly approaches described above, we generated clones with the desired repeat number and RVD order in both the *Xanthomonas* and *Ralstonia* TAL backbones. It should be noted that all the backbone clones are flanked by gateway L1 and L2 sequences to allow LR gateway cloning in desired destination vectors.

2.4.3 RTLs are nuclear proteins

As indicated in Figure 10, the general structural features of the RTL effectors are very similar to those of the *Xanthomonas* TALEs. In addition to a central DNA-binding region consisting of a variable number of nearly identical 35 amino acid repeats, RTL effector

structural features also include multiple putative NLSs and AD in the C-terminus. All of these features were previously shown to be essential for the functionality of TALEs, indicating that, if the RTLs are truly transcriptional activators, these features should be essential for the activity of the RTLs as well (Boch and Bonas, 2010). Because the functionality of the RTL NLS motif has not been experimentally verified, we determined whether RTL effectors possess a functional NLS and thereby localize to the cell nucleus. The RTL1, RTL2, and RTL3 clones fused to GFP were transiently expressed in *Nicotiana benthamiana* tobacco leaves. Two days post agroinfiltration, leaf discs were collected and the samples were analyzed by confocal laser scanning microscopy. The three RTL:GFP proteins localized to the nucleus, indicating the functionality of their NLS sequences (Figure 12). The nuclear localization of the RTLs suggests that these proteins may be transcriptional activators that modulate host gene expression for their own benefit.

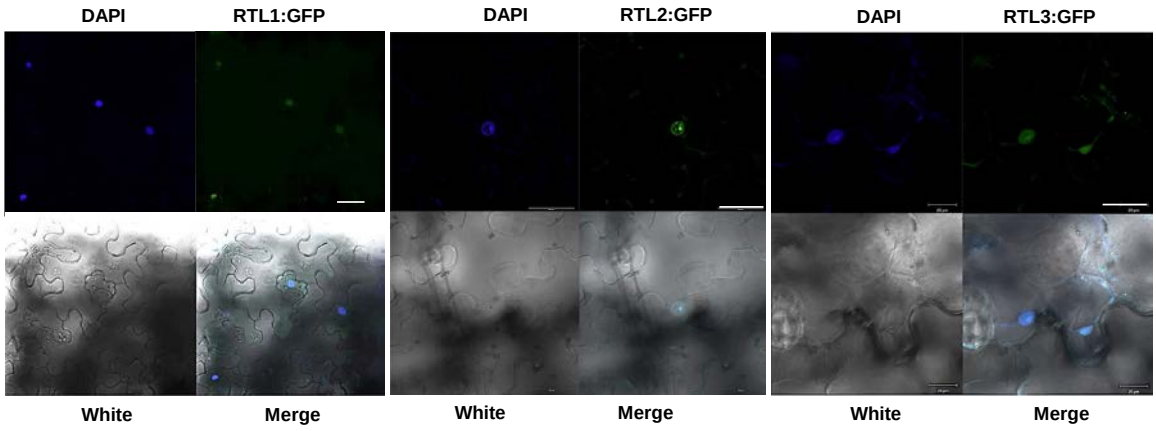


Figure 12. RTL Effector Proteins Localize in the Plant Cell Nucleus. RTL1, RTL2, and RTL3 were fused in frame to green fluorescent protein, respectively, to generate RTL:GFP and were transiently expressed in tobacco leaves via *Agrobacterium*-mediated delivery. Two days post agroinfiltration, confocal laser scanning microscopy was used to analyze samples. 4′,6′-Diamidino-2-phenylindole (DAPI) staining indicates nuclei. Scale bars = 30 µm. RTL:GFP fusion proteins localize in the plant cell nucleus.

2.4.4 RRs mediate DNA-binding in the *Xanthomonas* TALE backbone

To determine whether the repeats of the RTL DBDs mediate DNA-binding in the *Xanthomonas* TAL backbone, we engineered the RRs to include *Xanthomonas* RVD sequences with known DNA-binding specificities. Eleven identical RRs were cloned with the RVD sequence HD in the *Xanthomonas dHax3* backbone to generate the *dHax3*.RR.HD12 clone. To determine whether the RRs containing the RVD HD mediate DNA-binding and recognize the C nucleotide, we cloned all possible effector-binding elements (EBEs) (polyA, polyT, polyG, or polyC) into the *Bs3* minimal promoter, which drives *uidA* expression (Supplementary Sequence 5). The clones of the *dHax3*.RR.HD12 effector were co-agroinfiltrated with each of the promoter clones of the EBEs (polyA, polyG, polyC, or polyT:*Bs3*:*uidA*) in *Nicotiana benthamiana* leaves. Two days after agroinfiltration, the leaf discs were collected for qualitative and quantitative analysis of the transcriptional activation of the GUS reporter. As indicated in Figure 13, *dHax3*.RR. HD12 bound to the polyC EBE and strongly activated the expression of the *uidA* reporter *in vivo*. These data demonstrate that the *Ralstonia* repeats mediate DNA-binding and that the *Xanthomonas* RVD DNA-binding specificity code can be applied to the *Ralstonia* repeats. The ability of RRs to mediate DNA-binding strongly suggests that RTL effector proteins may function as transcriptional activators that bind to their genomic targets to reprogram host gene expression to the benefit of the phytopathogen.

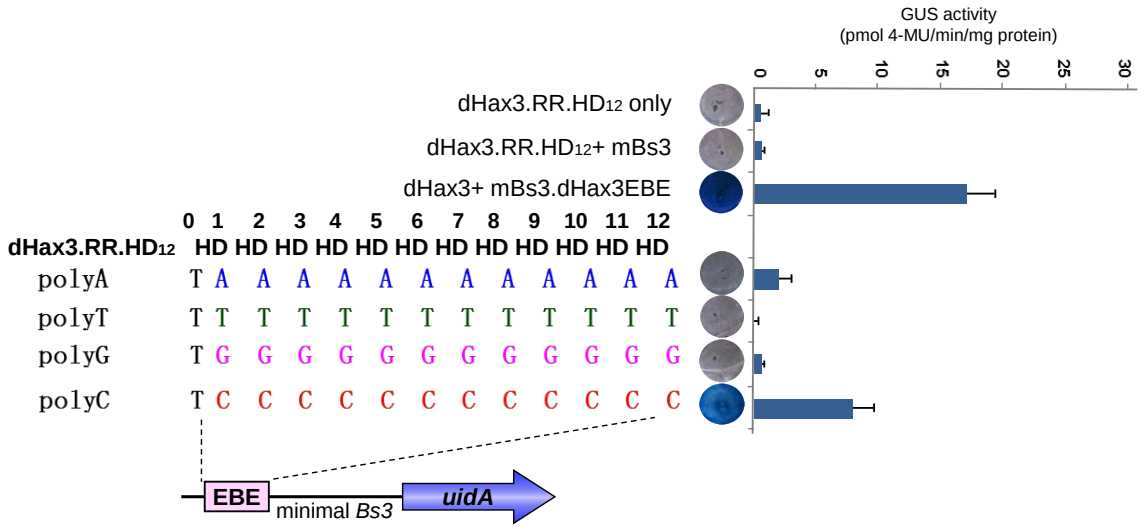


Figure 13. *dHax3*.RR.HD12 DNA-Binding Activity and Efficiency. Sequence of effector *dHax3*.RR.HD12 and T-preceded EBE boxes, polyA, G, T, and C, in minimal *Bs3* promoter (m*Bs3*) that drives *uidA* reporter gene expression. The constitutive 35S-driven *dHax3*.RR.HD12 clone was co-delivered, separately, with m*Bs3* containing polyA, G, T, or C EBE clone via A. *tumefaciens* into N. *benthamiana* leaves. *dHax3*.RR.HD12 alone and *dHax3*.RR.HD12 with m*Bs3* served as negative controls and *dHax3* with m*Bs3*.*dHax3*.EBE served as positive control. Leaf discs were collected 2 d post infiltration; half were stained with X-Gluc and half were used for quantitative measurement of GUS activities. The results show that the *dHax3*.RR.HD12 effector strongly activated the expression of m*Bs3*polyC driven *uidA* reporter. These experiments were repeated at least three times, with similar results. RR, *Ralstonia* repeat; EBE, DNA-binding element.

2.4.5 RTLs may function as transcriptional activators

Because the RRs mediated DNA-binding in the *Xanthomonas dHax3* backbone, we determined whether RTL effectors can function as genuine transcriptional activators. An RTL protein (RTL2_RR.HD12) was engineered that contained the previously reported *Xanthomonas* HD RVD in the RTL backbone. The RTL effectors were separately co-infiltrated into *Nicotiana benthamiana* leaves with m*Bs3* clones that contained polyA, polyT, polyG, or polyC EBEs. Two days post infiltration, leaf discs were collected and qualitative and quantitative GUS assays were performed. As shown in Figure 14, the

RTL2_RR.HD12 was able to bind to the promoter target sequence and activate the expression of the *uidA* gene. These data provide strong evidence that the RTLs may function as transcriptional activators and used by the *Ralstonia* phytopathogen to reprogram host transcription.

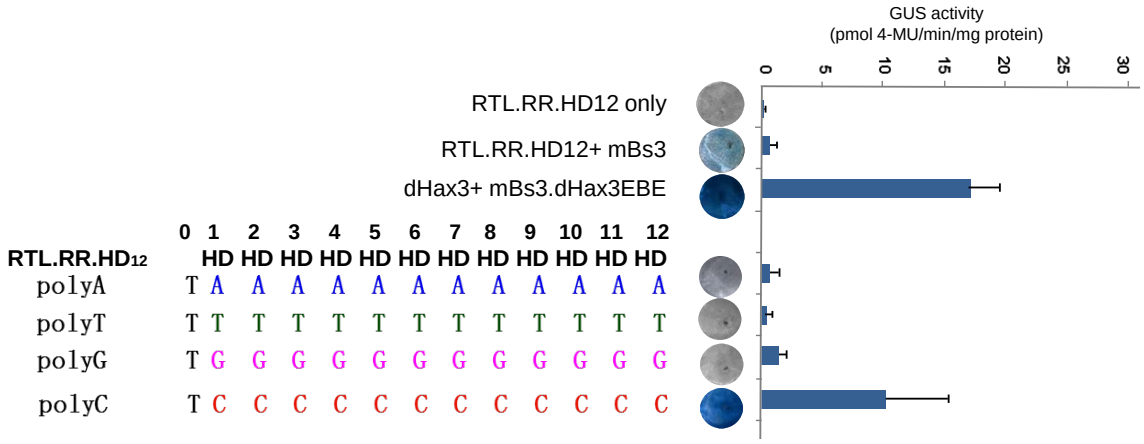


Figure 14. RTLs May Function as Transcriptional Activators. The effector clone of RTL2_RR.HD12 and clones of EBE boxes, T-preceded polyA, G, T, and C, in m*Bs3* promoter-driven *uidA* reporter were co-argoinfiltrated into N. *benthamiana* leaves, respectively. GUS activities were analyzed qualitatively and quantitatively 2 d post infiltration. RTL2_RR.HD12 alone and RTL2_RR.HD12 with m*Bs3* served as negative controls and *dHax3* with m*Bs3*.*dHax3*.EBE as positive control. RTL2_RR.HD12 strongly activated the expression of m*Bs3*polyC-driven *uidA* reporter gene.

2.4.6 The RVD ND interacts specifically with cytosine nucleotides

RTLs possess unique RVDs, and the DNA-binding specificities of the equivalent RVDs from TAL effectors of the *Xanthomonas* phytopathogen have not been determined. Moreover, the frequency of certain repeats is much higher in the RTLs than in the TAL effectors of *Xanthomonas* (Figure 10C). To determine the DNA-binding specificity of the ND RVD, we generated a DBD that possesses 12 RRs containing the RVD ND in the *dHax3* backbone (*dHax3*.RR.ND12). Two days after coagroinfiltration of the effector with each of the polyC-, G-, T-, or A- EBE promoter clones, the leaf discs were analyzed by GUS assays as described above. The *dHax3*.RR.ND12 selectively and strongly

activated the polyC EBE-containing promoter (Figure 15A), indicating that the RVD ND specifically recognizes and selectively binds to C nucleotides. To test whether the binding specificity of the other RVDs can be determined using the same approach, we co-expressed effectors containing poly(RVDs) (HN, NP, and NT) in the *dHax3* backbone along with promoter clones that contain DNA-binding boxes of polyC, G, A, and T nucleotides. *dHax3*.RR.HN12 showed a weak binding and activation to A and G nucleotides (Figure 15C). Interestingly, *dHax3*.RR.HN12 RVD repeats more strongly bind to and activate a target EBE-containing polyAG nucleotides stronger than to a target EBE with only polyA or polyG nucleotides (Figure 15C). Moreover, *dHax3*.RR.NP12 RVD showed moderate interaction and activation of C, A, and G (Figure 15D), and *dHax3*.RR.NT12 RVD showed weak binding and activation of T, A, G and C (Figure 15E).

To function, *Xanthomonas* TAL effectors require the presence of the T nucleotide preceding the DNA-binding box in the promoter region (Boch *et al.*, 2009). We therefore determined whether the T nucleotide that corresponds to the cryptic repeat 0 (T0) is important for DNA-binding and hence for transcriptional activation of RTL effectors. Several EBE DNA-binding boxes were generated, with or without (T0 is replaced by G) the T nucleotide, in *Bs4* minimal promoters fused to the *uidA* reporter and were cloned into binary vectors for plant expression. These EBE promoter constructs were co-delivered into tobacco leaves with their respective RTLs via agroinfiltration (Figure 15B). RTL effectors with EBEs lacking the T0 nucleotides showed minimal transcriptional activation background as negative controls (Figure 15B). Thus, the T

nucleotide that corresponds to repeat 0 is essential for RTL binding and transcriptional activation.

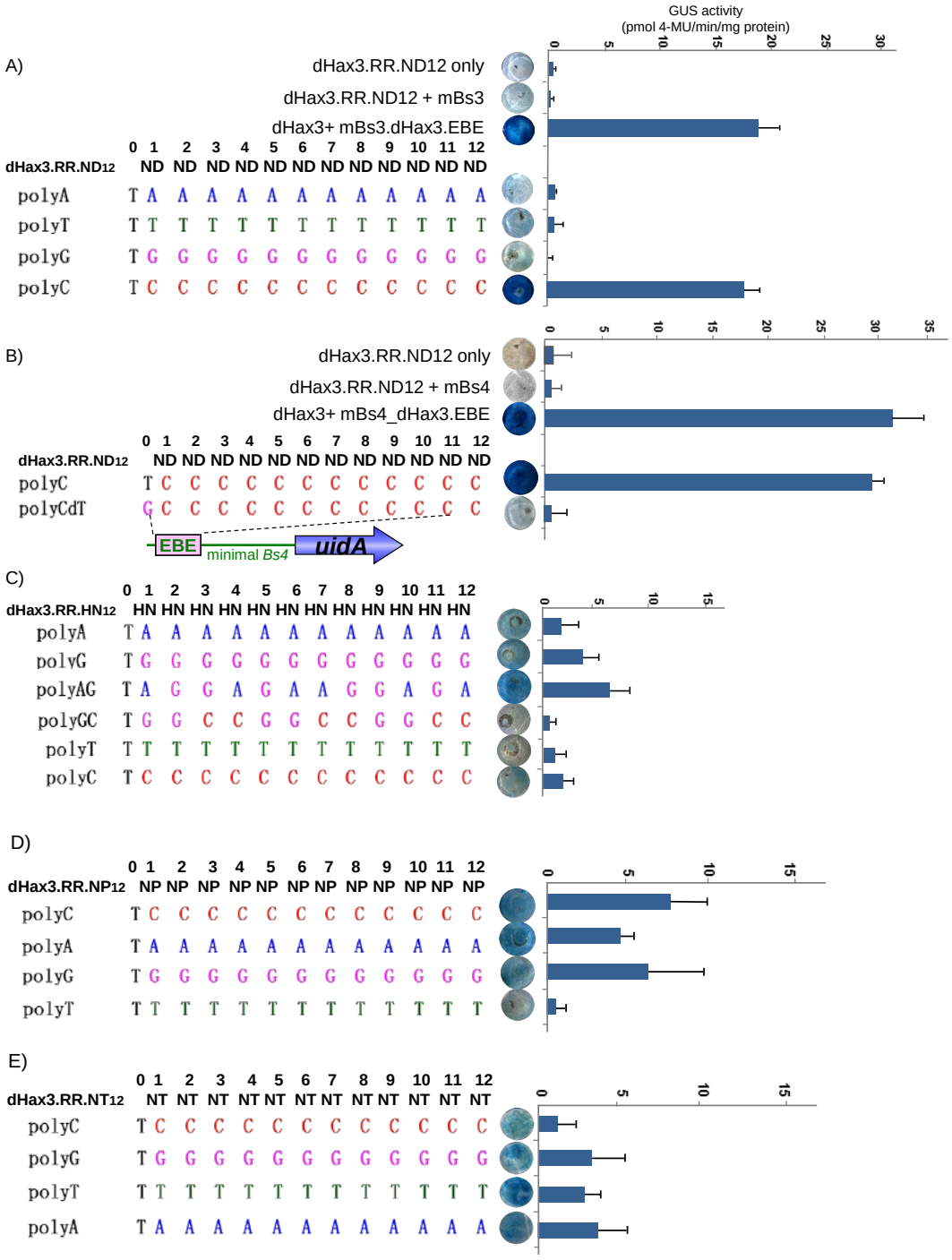


Figure 15. Determining the DNA-Binding Specificities of *Ralstonia* RVDs. The RVDs sequences of effectors, *dHax3*.RR.ND12, *dHax3*.RR.HN12, *dHax3*.RR.NP12, and *dHax3*.RR.NT12, and DNA sequences of EBE boxes in m*Bs3* or

m*Bs4* promoter that drive *uidA* reporter expression are shown. Clones of the RTL effectors and EBE-containing m*Bs3* were co-argoinfiltrated into N. *benthamiana* leaves, respectively, as shown. *dHax3*.RR.ND12 strongly and specifically activated the expression of polyC, preceded by T, -m*Bs3*-driven *uidA* reporter gene (A, B), but not polyCdT-*Bs3* (B). *dHax3*.RR.HN12 showed a weak activation to A and G (C); *dHax3*.RR.NP12 showed moderate interaction and activation of C, A, and G (D); and *dHax3*.RR.NT12 showed weak binding and activation of T, A, G, and C (E).

2.4.7 Determining the DNA-binding specificities of the RTL RVDs

The binding of the poly(RVD) appears to depend on the nature of the RVD. Thus, the use of RTLs containing poly(RVDs) may not be the best experimental approach to determine the DNA-binding specificities of RVDs. Consequently, RTL effectors were designed and assembled that contain mixed RVDs of known and unknown binding specificities. Following the same procedure that we used to determine the binding specificity of the RVD ND, we assembled an RTL2 effector with the RVD [HN-HN-ND-ND-HN-HN-ND-ND-HN-HN-ND-ND]. The possible EBEs for this assembly were fused to the *Bs3* minimal promoter that was used to drive *uidA* expression. The RTL2_HN-ND12 effector was agroinfiltrated with each of the possible *Bs3* promoters (Figure 16A). The RTL2_HN-ND12 effector induced GUS expression only in combination with the *Bs3* promoter containing the [TAACCAACCAACC] or [TGGCCGGCCGGCC] EBE. The HN-type RVD therefore mediates specific binding for A and G nucleotides. To determine the binding specificity of the NP-type RVD, we generated a *dHax3*_NP-ND-NT-NG11 effector containing the RVD sequence [NP-ND-ND-ND-NP-NP-NP-NT-NP-NP-NG] and fused the possible EBEs for this assembly to the *Bs3* minimal promoter. Our findings demonstrate that the NP-type RVD mediates DNA-binding to C, A, and G nucleotides, when: ND binds to C, NG binds to T, and NT binds to T nucleotides (Figure 16B). To further confirm the binding specificities, we assembled RTL2_RR.HN-NP12 (HN-HN-

NP-NP-HN-HN-NP-NPHN-HN-NP-NP), cloned possible EBEs in the m*Bs3* minimal promoter (Figure 16C), and separately co-expressed the effector with each possible EBE respectively via agroinfiltration in *N. benthamiana* tobacco leaves. The GUS assays showed that HN binds to A and G, NP binds more strongly to C than to G and A (Figure 16C).

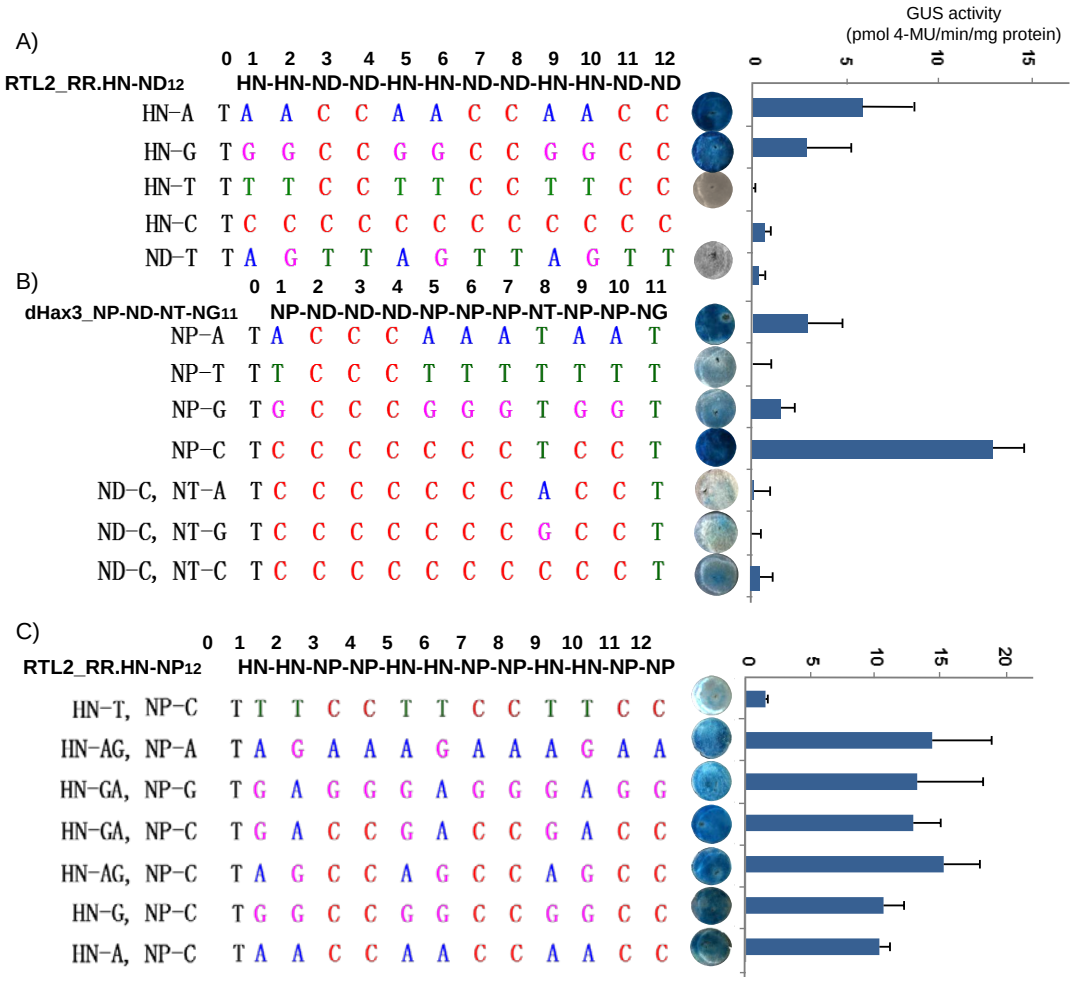


Figure 16. Determine the DNA-Binding Specificities of RTL RVDs. RTL2_RR.HN-ND12 (A), *dHax3*_NP-ND-NT-NG11 (B), and RTL2_RR.HN-NP12 (C) were constructed with sequences as shown. Possible target EBE boxes were cloned in m*Bs3* promoter with *uidA* reporter downstream. Clones of 35S promoter-driven RTLs were co-agroinfiltrated with possible EBEs, respectively. Leaf discs from 2 d post infiltration were collected for qualitative and quantitative analyses. RTL2_RR.HN-ND12 activated promoters containing EBEs that HN corresponds to A or G strongly (A); *dHax3*_NP-ND-NT-NG11 activated promoters containing EBE that NP corresponds to C stronger, but to A and G

weaker, when ND binds to C and NT binds to T, but, when NT was changed to A, G, or C, *dHax3*_NP-ND-NT-NG11 did not activate *uidA* reporter expression (B); RTL2_RR.HN-NP12 activated promoters containing EBEs that HN corresponds to A or G, NP corresponds to A, G, and C, strongly (C).

2.5 Discussion

Developing customizable DNA-binding modules that can be easily engineered to bind to any user-selected DNA sequence is of paramount importance in the development of tools and reagents for genome-engineering applications. TAL effectors from *Xanthomonas* phytopathogenic bacteria have been recently used as programmable DNA-binding scaffolds for genome-editing applications (Bogdanove *et al.*, 2010; Bogdanove and Voytas, 2011; Mahfouz and Li, 2011). Although the TAL DNA-binding code is surprisingly simple, many questions remain about the biology of TAL effectors and their molecular function. Identifying novel TAL-like DNA-binding architectures from other phytopathogens would advance our understanding of the function of TAL effectors in disease biology and facilitate their use in customizable platforms for DNA targeting. We report the identification and characterization of *R. solanacearum* TAL-like effectors and demonstrate the DNA-binding specificities of several of their naturally occurring RVDs, including ND, HN, NP, and NT.

Based on their overall structural similarity to *Xanthomonas* TAL effectors, three TAL-like *Ralstonia* proteins (RTL1, RTL2, and RTL3) were selected for molecular characterization and *in vivo* testing of their DNA-binding specificities. To determine whether RTLs function as transcriptional factors, we first assessed the ability of their DNA repeats to mediate DNA-binding. The data demonstrate that the RRs in the *Xanthomonas* TALE backbone do mediate DNA-binding as inferred by the transcriptional activation of their targets (Figure 13). These data prompted us to study

whether the RTLs are genuine DNA-binding proteins that mimic eukaryotic transcriptional factors. Cellular localization studies indicate that the RTLs are localized to the nucleus (Figure 12), which provides evidence that those RTLs may possess functional NLSs. Because our data showed that the RRs in the *Xanthomonas* backbone mediate DNA-binding and possess functional NLSs, we next determined whether they function as genuine transcription factors and activate the expression of target genes. RTLs with poly(HD)-containing repeats induced the transcriptional activation of promoters that contained the respective DNA-binding sequences (Figure 14). These data indicate that RTLs may function as transcriptional activators. The finding that RTLs may function as transcriptional activators prompted us to determine the DNA-binding specificities of RTL RVDs. The ND, HN, NP, and NT RVDs were selected for analysis of their DNA-binding specificities. Our data clearly demonstrate that the ND RVD binds to the C nucleotide with high specificity. Surprisingly, RTL1, RTL2, and RTL3 containing poly(NP), poly(HN), and poly(NT) did not noticeably induce the expression of the polyA/T/G/C:m*Bs3*::*uidA*. These data indicate that the RTL binding may depend on the composition of the RVDs. It should be noted that RVDs possess different binding affinities (Deng *et al.*, 2012a; Mak *et al.*, 2012). Because the RTLs containing poly(ND) bind specifically and strongly to the C nucleotide, the inability of RTLs containing other poly(RVDs) to bind to DNA might be the result of weak binding of those RVDs. In other words, RTLs composed of strong RVDs may bind strongly to DNA, and RTLs composed of weak RVDs may bind weakly to DNA, which could affect the overall function of the RTL. Because the HD and ND RVD types bind to DNA with high affinity, their

poly(RVDs) were able to bind to DNA and activate the expression of the GUS reporter with high efficiency.

To exclude the possibility that the inability to bind DNA is due to the nature of the RTL scaffold and that the binding might be scaffold-dependent, we generated poly(RVDs) in the *Xanthomonas dHax3* backbone and tested their binding activities. As observed with the RTL scaffold, only *dHax3*.poly(ND) showed strong binding to the C nucleotide, whereas the *dHax3*.poly(NT), (HN), and (NP) did not show strong binding with any nucleotide (Figure 15). These data clearly indicate that, when the repeats contain a single RVD type, the binding of the TAL or RTL effectors will be weak if the binding of that particular RVD is naturally weak.

When mixed repeats containing strong and weak RVDs were generated, the RTL effector strongly induced GUS expression, indicating strong DNA-binding. These data suggest that the presence of strong RVDs could stabilize RTLs and allow the weak RVDs to bind specifically (Figure 16). Thus, the binding strength of an engineered RTL could be increased by including strong RVDs spaced throughout the repeats. Similar conclusions regarding DNA-binding specificities and efficiencies were drawn recently from work on *Xanthomonas* TAL effectors and repeats (Streubel *et al.*, 2012). Further studies, however, are needed to determine the positional effects of the RVDs on the DNA-binding strength and on transcriptional functions of the RTLs or TALEs.

Recently, the structural basis of TAL binding to DNA has been reported. These studies indicated that the 12th residue of the RVD makes the conformation more rigid, whereas the 13th residue mediates base-specific contacts (Deng *et al.*, 2012a; Mak *et al.*, 2012). Our data indicate that *Ralstonia* TAL effectors may assume a similar structure, because

RVDs that end with D (e.g. ND) specifically recognize the C nucleotide, whereas the RVDs that end with N (e.g. HN) recognize the A and G nucleotides. At least two important questions remain to be answered with structural and molecular genetics approaches. First, what are the functions of the two repeats (repeat 0 and repeat –1) that precede the canonical repeats containing the DBD? These two repeats are quite similar to the canonical repeats but lack the RVD residues. Also, at the end of the C-terminus of the repeat domains, there are two half-repeats, +1 and +2, that lack known RVDs. It will be quite important to determine the molecular functions of these repeats and whether they contribute to the RTL binding to DNA. Second, what are the molecular mechanisms of the conformational plasticity of the TALEs and RTLs? Such plasticity might be important for their binding specificity to epigenetically modified DNA in its cellular chromatic environment (Deng *et al.*, 2012b). Answers to these questions would guide the design of RTLs and TALEs to ensure their conformational plasticity and thereby their specificity and molecular activity *in vivo*.

2.6 Significance

We are presenting novel structures that can be used as targeting modules and novel RVDs that can be used to engineer RTLs to bind to any user-selected DNA sequence. That knowledge will allow to generate targeting modules with greater specificity and flexibility for biotechnological applications e.g. in agriculture. These findings should facilitate the development of several immediate biotechnological applications, including the use of RTLs as customizable DNA-binding scaffolds to generate a variety of chimeric proteins for targeted genomic and epigenomic modifications. RTLs could also be tailored to generate sequence-specific transcriptional activators and could be fused to other

functional domains, such as transcriptional repressors, nucleases, methylases, demethylases, recombinases, or chromatin modifiers. In addition to supporting the development of new technologies, our findings may enhance the basic understanding of the molecular interactions between pathogenicity-related genes and RTLs.

2.7 Materials and Methods

2.7.1 Generation and cloning of the RTL repeat concatemers in the *dHax3* backbone

To assemble RRs in the *dHax3* backbone, a collection of plasmids (pUCminusMCS) that contains three types of fragments was generated. The first fragment type contains the first repeat plus an overlapping part of the N-terminus of the *dHax3* backbone to facilitate cloning using the *PpuM*I restriction-enzyme. The second fragment type contains a single-repeat clone that can be cleaved with and can concatamerize in the ligation reaction, thereby generating different fragment lengths. The third repeat type contains the last repeat plus the C-terminus of the *dHax3* backbone to facilitate cloning using the *Sac*I enzyme. All RVD possibilities were generated in these fragments to facilitate the assembly of the desired number and order of the *Ralstonia* RVDs in the *dHax3* backbone. Sequence information for all of these fragments is provided in the Supplementary Information for Chapter 2. To assemble the repeats, three 20-µl digestion reactions containing 1000–1500 ng of DNA were used. The first, second, and third fragments were released from the plasmid by digestion with *PpuM*I and *BsmA*I, *BsmA*I, and *Sac*I and *BsmA*I, respectively. All digested fragments were purified from the agarose gel. To allow concatemer formation of the second fragment, each type of RVD was ligated to the last fragment in a 1:3 ratio. After 4 h of incubation with T7 ligase, the same RVD and ligation buffer was added, and the ligation reaction

was repeated three times. After the completion of the ligation reaction, 1.2–1.5-kb products were separated and eluted from the E-gel Ex 2% Agarose (Life Technologies, Grand Island, NY, USA), ligated to the first fragment (which had been digested with *PpuM*I–*Bsm*AI), and finally ligated to the backbone (pENTR221.*dHax3* digested with *PpuM*I–*Sac*I) with T7 ligase. After 2 h of ligation, the product was used to transform ONE-shot Top 10 cells (Life Technologies), which were spread on Kan^+ LB plates. The repeat fragment size was confirmed either by colony PCR or by *PpuM*I–*Hind*III restriction digestion. Sanger sequencing confirmed the RVD number and order. NEB (New England Biolab) restriction enzymes were used for all restriction digestions.

2.7.2 Design and construction of the backbones of RTL proteins

To facilitate the synthesis and restriction ligation assembly of overlapping fragments, the codons of the *Ralstonia* clones (YP_003750492.1, NP_519936.1, and CAQ18687.1) were optimized. The sequences of the optimized clones are provided in the Supplementary Information for Chapter 2. Each of the RTL clones was divided into fragments containing the N-terminus, the central repeats, and the C-terminus, with overlapping ends containing unique restriction enzymes. For example, to assemble the YP clone, the N-terminus fragment (NTF) was synthesized and cloned into pUC19minusMCS flanked by the L1 gateway cloning sequence at the 5’ end, and by a sequence containing *Apa*I-*Nde*I-*Spe*I restriction recognition sequences at the 3’ end. The C-terminus fragment (CTF) was designed and synthesized to be flanked by an *Nde*I site at the 5’ end and by the L2 gateway cloning sequence followed by a *Spe*I site at the 3’ end. The pUCminusMCS.L1.NTF.*Apa*I-*Nde*I-*Spe*I clone was used as a backbone to subclone the CTF using *NdeI* and *Spe*I to generate the pUCminusMCS.L1.NTF.*Apa*I/*Sal*I-*Nde*I-

CTF.L2.SpeI backbone clone. This RTL- minus-repeats clone (RTLΔR) was used as a scaffold to assemble the repeats in a random or defined order. The other RTL clones were assembled using the same approach, and the details of their assembly are described in the Supplementary Information for Chapter 2.

2.7.3 Subcloning of the RRs into RTL backbone clones

The RRs in the *dHax3* backbone were amplified by primers that contain the *Apa*I/*Sal*I and *Nde*I enzyme sites at the 5' and 3' ends, respectively, and were cloned into a TOPO TA vector. The RTL repeat fragments were digested with *Apa*I or *Sal*I and *Nde*I and subcloned into the RTLΔR backbone. Sanger sequencing was used to confirm the identity of the full-length clones with the *et al*. assembled clone. The RTL clones were subcloned into the pK2GW7 and pK7WGF2 plant overexpression vectors or the pET32a-GW bacterial expression vector using the LR gateway cloning reaction.

2.7.4 Ordered assembly cloning of the RRs in the RTL backbone clones

To achieve the ordered assembly of the RRs in the RTL backbone clones, a library of di-repeat clones was assembled that contains the *Ralstonia* RVDs in all possible combinations. The di-repeat clones, flanked by *BsmA*I sites, produce distinct overhangs after digestion to ensure the ordered assembly of 12 repeats. The sequences of the di-repeats in the library are provided in the Supplementary Information for Chapter 2. Several RTL clones were generated with repeats in the desired order, and full-length clones were confirmed by Sanger sequencing. Generating an extended library of the di-repeats will make it possible to generate any RTL with any desired RVD sequence to target any DNA sequence of interest.

2.7.5 Cloning of minimal *Bs3* and *Bs4* promoters

Minimal *Bs3* (m*Bs3*) and *Bs4* promoters containing different EBEs, which were designed to test the transcriptional activation of different effectors, were PCR-amplified using specific primers listed in the Supplementary Table 1 and cloned into a D/TOPO vector. After the clones were confirmed by Sanger sequencing, they were transferred by LR reaction to the gateway destination vector pKGWFS7 to drive the expression of *uidA* reporter genes.

2.7.6 RTL cellular localization studies

To confirm whether RTLs function as transcription activators, pK7WGF2/RTL clones that express GFP–RTL fusion proteins *in planta* were generated. The clones were transformed into *Agrobacterium tumefaciens* GV3101 and used for transient expression in tobacco leaves. Two days post agroinfiltration, infiltrated discs were collected for confocal laser scanning microscopy analysis. The discs were prepared on slides, embedded with 50% glycerol, and covered with coverlids. The specimens were examined and imaged with an LSM 710 laser scanning confocal microscope (Zeiss, Germany) with ZEN 2009 software. Stained DAPI was excited by a 405-nm laser, while the detection range was set at 410–483 nm. The expressed GFP was activated by a 488-nm laser with an emission range of 493–598 nm. Images were recorded with a 2000 × 2000 pixel frame size and 8 bits of depth.

2.7.7 Measurement of RTL transcriptional activity using GUS reporter assays

To test the transcriptional activity of the RTLs *in planta*, pK2GW7/RTLs were generated. The clones were transformed into *A. tumefaciens* GV3101 and used for transient expression in tobacco leaves. *A. tumefaciens* harboring m*Bs3*/*Bs4::uidA* and its RTL

activator were co-infiltrated or separately infiltrated into tobacco leaves. Infiltrated discs were collected at 48 h post inoculation. The qualitative GUS activity assay was performed by immersing leaf discs in GUS staining solution (10 mM sodium phosphate, pH 7, 10 mM EDTA, 0.1% Triton X-100, 0.1% 5-bromo-4-chloro-3-indolyl-β-D-glucuronide, 1 mM potassium ferricyanide, 1 mM potassium ferrocyanide) at 37°C for 24 h. The discs were cleared in ethanol. For the quantitative GUS analysis, three leaf discs were collected, and the tissue was homogenized, diluted, and incubated with 4-methyl-umbelliferyl-β-D-glucuronide (MUG) as previously described (Kay *et al.*, 2007).

CHAPTER 3

Rapid and Highly Efficient Construction of TALE-Based Transcriptional Regulators and Nucleases for Genome Modification

Reprinted with permission by Springer from:

Li, L., Piatek, M.J., Atef, A., Piatek, A., Wibowo, A., Fang, X., Sabir, J.S., Zhu, J.K. and Mahfouz, M.M. (2012) Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. *Plant molecular biology* 78, 407-416. doi: 10.1007/s11103-012-9875-4.

3.1 Abstract

Transcription activator-like effectors (TALEs) can be used as DNA-targeting modules by engineering their repeat domains to dictate user-selected sequence specificity. TALEs have been shown to function as site-specific transcriptional activators in a variety of cell types and organisms. TALE nucleases (TALENs), generated by fusing the FokI cleavage domain to TALE, have been used to create genomic double-strand breaks. The identity of the TALE repeat variable di-residues, their number, and their order dictate the DNA sequence specificity. Because TALE repeats are nearly identical, their assembly by cloning or even by synthesis is challenging and time consuming. Here, we report the development and use of a rapid and straightforward approach for the construction of designer TALE (dTALE) activators and nucleases with user-selected DNA target specificity. Using our plasmid set of 100 repeat modules, researchers can assemble repeat domains for any 14-nucleotide target sequence in one sequential restriction-ligation cloning step and in only 24 h. We generated several custom dTALEs and dTALENs with new target sequence specificities and validated their function by transient expression in tobacco leaves and *in vitro* DNA cleavage assays, respectively.

3.2 Introduction

The existence of a large number of TALEs and information regarding their target genes have been essential for deciphering how these proteins bind to the promoter regions of their target genes (Boch *et al.*, 2009; Moscou and Bogdanove, 2009). By using the TALE DNA-binding code, a suite of precise and efficient molecular genome-modification tools can be developed that are applicable to many different organisms. These genome-engineering tools are based on the design of chimeric proteins composed of TALE, which confers very high DNA sequence specificity when coupled to a variety of functional domains including nucleases, activation and repression domains, methylases, and integrases. TALEs fused to the *Fok*I cleavage domain generate chimeric nucleases (TALENs) that bind to the DNA and create double-strand breaks (DSBs) in several systems and cell types (Cermak *et al.*, 2011; Christian *et al.*, 2010; Huang *et al.*, 2011; Morbitzer *et al.*, 2011; Weber *et al.*, 2011; Zhang *et al.*, 2011a). The development of an efficient and simple repeat assembly approach for the construction of designer TALEs (dTALEs) is needed to advance the use of TALE-based genome-engineering applications across different species and systems.

DNA fragments longer than 200 bp can be generated by assembly PCR (Smith *et al.*, 2003) or microarray (LeProust *et al.*, 2010). However, DNA constructs composed of repetitive modules, like TAL effectors, cannot be engineered by aforementioned methods due to high likelihood of errors occurrence by oligonucleotides randomly incorporating into constructs (Xiong *et al.*, 2008). A few strategies for constructing customized TALEs with specific arrangements of tandem repeat monomers have been reported. First method, the Golden Gate (vector based) cloning (Engler *et al.*, 2009) uses type IIS restriction

enzymes' properties to cleave outside of their recognition site. Directional assembly of multiple DNA fragments is possible due to the design of type IIS restriction sites to have different sequences and to create multiple sticky ends. When the recognition sites are placed in inverse orientation, on the DNA fragment, they are removed in the cleavage process, therefore allowing two DNA fragments with compatible sequence overhangs to ligate in order and faultlessly. The first successful assembly of TALE effectors using the Golden Gate cloning method was performed to target a promoter sequence driving GFP expression in a transgenic plant (Weber *et al.*, 2011). This technique requires various pre-made plasmids and takes about one week to finalize the experiments. Another method, quicker than the abovementioned one, bases on the modification of the Golden Gate cloning and uses PCR fragments instead of plasmid fragments (Sanjana *et al.*, 2012; Zhang *et al.*, 2011a). Nevertheless, none of these methods are adjusted for high-throughput production due to requirements for PCR, gel purification of fragments, and characterization of intermediate constructs.

Here, we present a rapid and efficient approach for the modular assembly of TALE repeats using the *dHax3* scaffold as backbone for TALE transcription factors (TALE-TFs) and TALENs. With this approach, a dTALE with 12.5 or more repeats can be assembled in just one sequential restriction-ligation step and in only 24 h from a set of 100 plasmids containing all possible combinations of repeat modules. In contrast to previously reported approaches, our approach involves just one sequential restriction-ligation step and results in a high efficiency and reproducibility of assembly, does not require highly skilled researchers, and can be performed in only 24 h. Because our

assembly strategy facilitates the rapid construction of TALE-TFs and TALENs, it should expedite TALE-based genome-engineering applications in plants and other eukaryotes.

3.3 Research Aim and Objectives

The aim of this study was to develop robust and efficient protocol for rapid TALE protein assembly that would facilitate the application of TALE-based genome modifications in various organisms including plants. Objectives of this study included: generation of a plasmid library containing possible combinations of repeat modules; assembly of multiple repeats in one sequential restriction-ligation step; *in vivo* testing of the custom engineered TALE transcription factors with novel user selected sequence specificities; *in vitro* functional testing of custom engineered TALENs in DNA cleavage assays.

3.4 Results

3.4.1 Assembly of dTALE repeats in a one-step sequential restriction-ligation reaction

We selected the *dHax3* TALE protein as a scaffold and backbone for the construction of dTALE-TFs and dTALENs. The *dHax3* backbone has been successfully used in several systems as a transcriptional activator and chimeric nuclease. Our goal was to establish a tool kit that facilitates a rapid and ordered assembly and cloning of the repeats to any user-selected recognition sequence in dTALE-TF or dTALEN backbones. To achieve this goal, we designed, synthesized, and cloned a plasmid library of fragments containing all possible combinations of RVDs in the pUCMCS minus vector. The *dHax3* repeat sequence was optimized to include a few restriction enzymes at specific positions so that their restriction produces several distinct fragments that can be correctly and orderly assembled by restriction-ligation cloning. These fragments can be assembled in a one-

step sequential restriction-ligation cloning to produce a custom order of repeats for any user-selected target specificity. Our assembly approach is based on a 100-plasmid library of repeat fragments and type IIS restriction enzymes (Figure 17). We used *Ppu*MI and *Sac*I to clone the assembled repeats into the TALE backbone (Figure 18 and Supplementary Sequence 1).

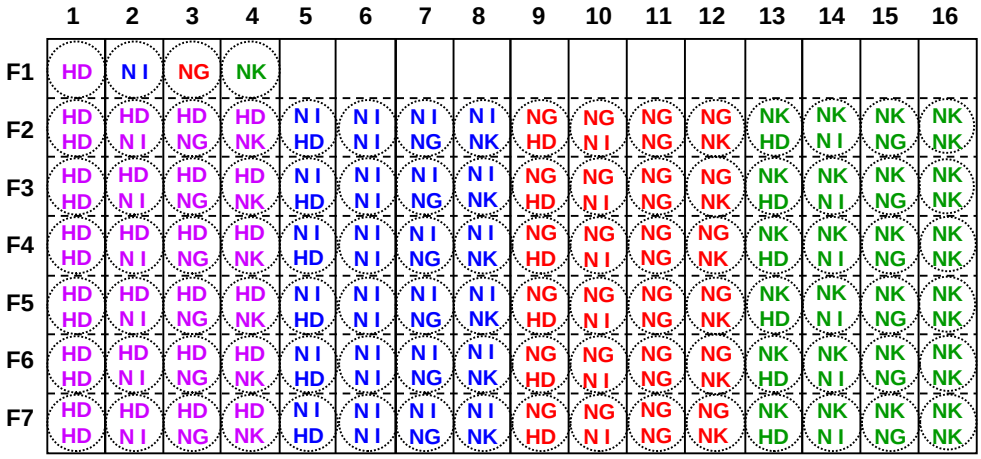


Figure 17. Diagram of library fragments. Diagram of library fragments to be used as a reference in assembly and stocking.

To assemble the repeats, we used *BsmA*I and *BsmB*I that cleave outside their recognition sequence and produce 4-bp overhangs. Repeat fragments were designed to contain *BsmA*I and *BsmB*I enzymes so that ordered assembly of the fragments can be achieved after restriction with *BsmA*I and *BsmB*I by sequential restriction ligation cloning method (Figure 18). We also used *Xho*I enzyme to assemble fragment 6–7. The introduction of this enzyme also serves another purpose when confirming the clones by restriction digestion, we used *PpuM*I and *Xho*I because *Xho*I is absent in the original backbone. As indicated in Figure 17, our plasmid library is based on seven fragments; the first fragment contains one RVD and hence has four plasmids. The second fragment, and also the third

through seventh fragment, contains two RVDs, and hence each has 16 plasmids for all possible combinations of RVDs (Figure 17). The restriction of these fragments by the designated enzymes produces fragments that can be assembled in a desired order. The overlaps of the fragments were designed so that only an ordered and correct assembly is possible. For each fragment, we generated all possible variants that specify the RVDs, including HD that binds to the C nucleotide, NI that binds to the A nucleotide, NK that binds to the G nucleotide, NN that binds to the A or G nucleotides, and NG that binds to the T nucleotide.

We performed the assembly of the repeats using a one-step sequential restriction-ligation protocol and confirmed the correct assembly of the clones. More than 95% of the clones confirmed by restriction digestion followed by Sanger sequencing had the correct order and number of repeats, indicating that this assembly approach is highly efficient. Using this protocol, we can produce TALENs and TALE-TFs with 12 and 13 repeats in just 24h with much less effort than required for other methods. The assembly of 13 repeats can be achieved if fragment 6 is restricted with *BsmA*I and *Xho*I enzymes (e.g. Supplementary Sequence 2). This number of repeats should be sufficient to produce dTALENs that are highly specific, with a single genomic target. Although a dTALE-TF with 13 or 14 repeats might have more than one target in the genome, it would still be sufficiently specific for targeting a group of genes.

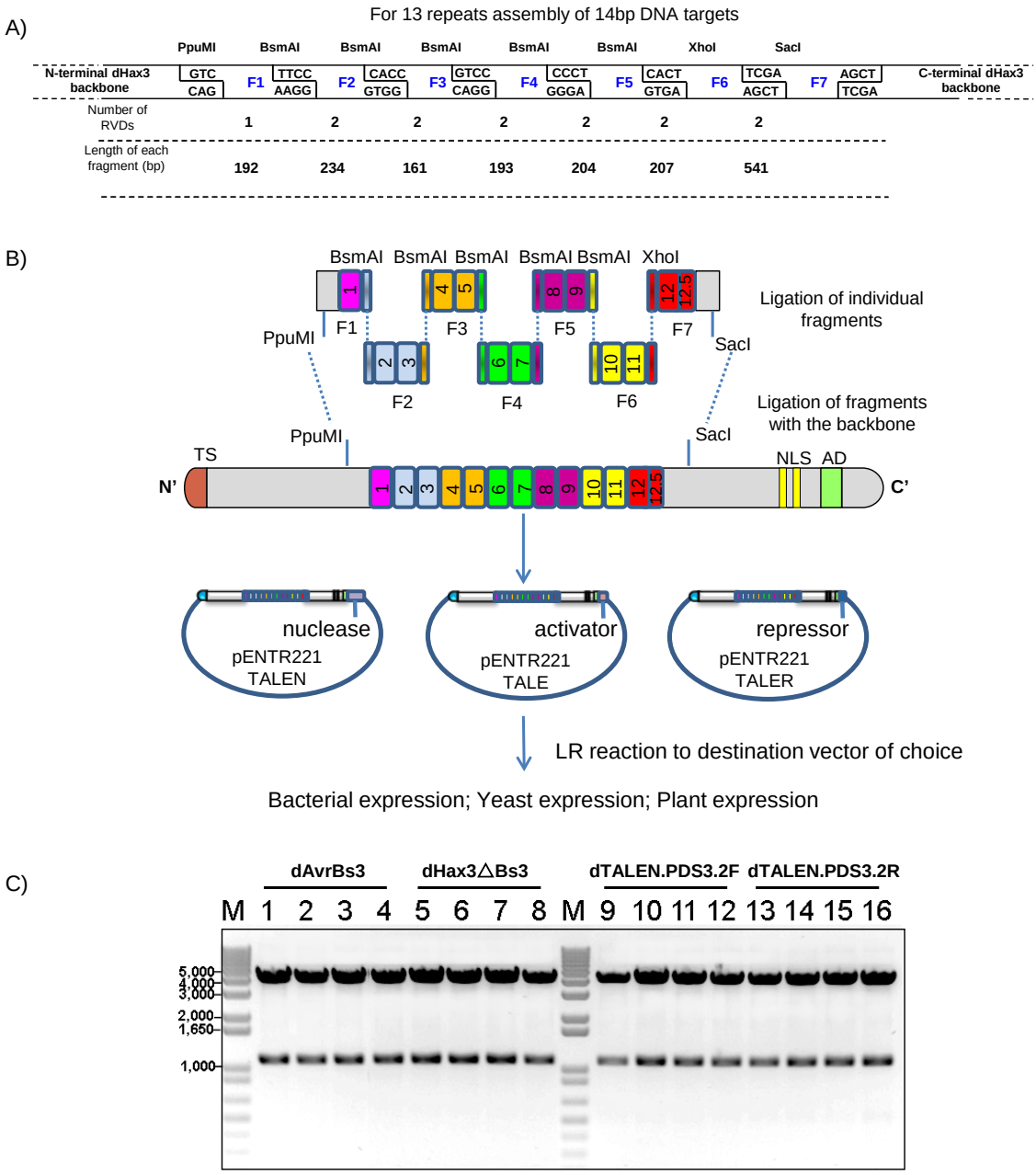


Figure 18. Construction of TALE-based transcription factors and nucleases by one sequential restriction ligation approach. **A)** Schematic representation of the *dHax3* repeat region showing the seven restriction fragments and the overhang sequence of each fragment produced by a restriction digestion. B) Overall assembly strategy showing the steps required to produce the final expression clone of TALE-TF or TALEN. The repeat region can be transferred between different *dHax3*-based backbones, with different effectors, using the *PpuM*I and *Sac*I restriction enzymes. C) Confirmation of the assembly clones by restriction digestion analysis. The *lower* bands, indicated by the *arrow*, are the products of *PpuM*I and *Xho*I digestion and demonstrate the successful and correct assembly of the repeats. All

restriction digestion positive clones showed the correct order and sequence of the repeats, indicating the extreme fidelity of this assembly approach. *M* 1-kb plus marker, *lane 1–4*, dAvr*Bs3*, *lane 5–8 dHax3*delta*Bs3*, *lane 9–12* dTALEN.*PDS*3.2F, and *lanes 13–16* dTALEN.*PDS*3.2R

Because the dTALEN and dTALE-TF backbone vectors are already in the Gateway-compatible vector pENTR221, an assembled dTALEN or dTALE-TF can be cloned in any Gateway-compatible destination vector for bacterial, yeast, plant, or mammalian expression (Figure 18 B). The LR Gateway reaction that moves the dTALEN or dTALE-TF clones from the pENTR221 vector into a plant expression vector can be avoided by the direct assembly of the repeats in the pK2GW7 expression vector. We performed the assembly of the dTALE repeats directly in the pK2GW7 plant expression vector. We first restricted the pK2GW7.*dHax3* (for TALE-TF backbone) and pK2GW7.*dHax3*.N (for TALEN backbone) clones with the unique *Sac*I and *PpuM*I and then assembled the repeats as described above. We generated correctly assembled TALE-TFs and TALENs in the destination expression vector in one step and with high efficiency. The ability to perform the assembly directly in the expression vector eliminates the need for a two-day step and expensive Gateway reagents. Users could have the choice of assembly directly either in the expression vector or in an entry Gateway-compatible clone that can be transferred into a variety of expression destination vectors for different species and different purposes. This assembly approach is relatively simple, rapid, straightforward, and reproducible, and thus can be utilized by researchers with different levels of molecular biology skills. Hence, it can facilitate the generation of genome-engineering reagents and their application in many laboratories and for various systems and organisms. Using our assembly approach, we efficiently generated more than 20 TALEs for site-specific transcriptional activation and cleavage purposes.

3.4.2 Functional testing of engineered dTALE-TFs in tobacco leaf transient expression assays

We tested the activity of the assembled dTALE-TFs *in vivo* in *Nicotiana benthamiana* leaves. We used our assembly strategy to assemble several dTALE-TFs in pENTR221. The pENTR221/dTALE-TF clones were then moved into the pK2GW7 plant expression vector by LR Gateway cloning. The pK2GW7/dTALE-TF clones were transformed into *Agrobacterium tumefaciens*, which was used to infiltrate *N. benthamiana* leaves. We used the WT *Bs3* promoter and generated different versions of it by inserting different DNA-binding boxes (Supplementary Sequence 3). We assembled different dTALE-TFs that would recognize these DNA-binding boxes. Tobacco leaves agroinfiltrated with pK2GW7/dAvr*Bs3* and *BS3::uidA* vectors were collected 48 hpi and assayed for GUS (β-glucuronidase) activity. We tested several dTALE-TFs that bind to different DNA-binding boxes inserted in the *Bs3* promoter, which is activated by Avr*Bs3* protein. To activate the WT *Bs3* promoter, we assembled a dTALE-TF (dAvr*Bs3*) with 12.5 repeats that specifically binds the 14-bp Avr*Bs3* target sequence. The assembled dAvr*Bs3*-TFs activated the expression of the WT *Bs3* promoter (Figure 19). Transcriptional activation did not occur following agroinfiltration with the *Bs3* promoter only or coinfiltration with other dTALE-TFs that do not contain a binding box in the *Bs3* promoter. These data indicate that the assembled dAvr*Bs3*-TF binds to the *Bs3* promoter and activates transcription *in vivo*.

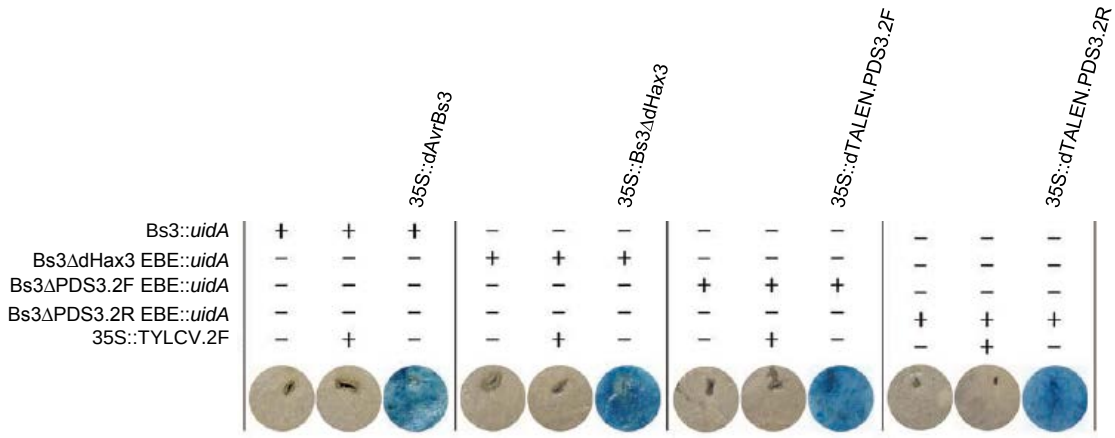


Figure 19. Validation of custom assembled dTALE-TFs in transcriptional activation of EBE containing promoters. Assembled dTALE-TFs, dAvr*Bs3*, *dHax3*, *PDS*3.2F, and *PDS*3.2R EBE activate WT *Bs3* promoter or modified *Bs3* promoter containing their respective binding EBE boxes (*lane 3*, *6*, *9*, and *12*). For a negative control, we used *Bs3* promoter only (*lane 1*, *4*, *7*, and *10*) or modified *Bs3* promoter. The 35S::dTALE.TLYCV.2F, which bind to a different EBE was used as a negative control as well (*lane 2*, *5*, *8*, and *11*).

3.4.3 Functional testing of engineered dTALEN activity in DNA cleavage assays *in vitro*

We tested the ability of the assembled dTALEN proteins to cleave their double-stranded DNA targets *in vitro*. Previous studies have shown that the nuclease activity of the FokI enzyme and the zinc finger hybrid nucleases require dimerization (Mani *et al.*, 2005; Miller *et al.*, 2007; Pruett-Miller *et al.*, 2009). To test the activity of our dTALENs, we designed target sequences containing two effector-binding elements in a tail-to-tail orientation and separated by 16-bp spacers (Li *et al.*, 2011a; Mahfouz *et al.*, 2011). We assembled the dTALEN heterodimers that bind to the phytoene desaturase cDNA from *A. thaliana*. The heterodimer pair was assembled in pENTR221 and then moved into Thioredoxin.His/pET32A by LR Gateway cloning. The linearized pCRII/*PDS*3 cDNA plasmid was gel-purified and used as substrate for the *PDS*3TALEN digestion reactions.

The thiroedoxin-tagged *PDS*3 dTALEN heterodimer pair proteins were expressed in *Escherichia coli* and affinity purified using Ni-NTA. The DNA cleavage activity of the proteins was tested *in vitro* on the pCRII/*PDS*3 cDNA plasmid (Mahfouz *et al.*, 2011). The heterodimer pair was active *in vitro* and resulted in cleavage of the *PDS*3 target. Digestion reactions performed using only one monomer of the *PDS*3 dTALENs (either *PDS*3 TALEN-F or *PDS*3 TALEN-R) did not show any cleavage activity (Figure 20).

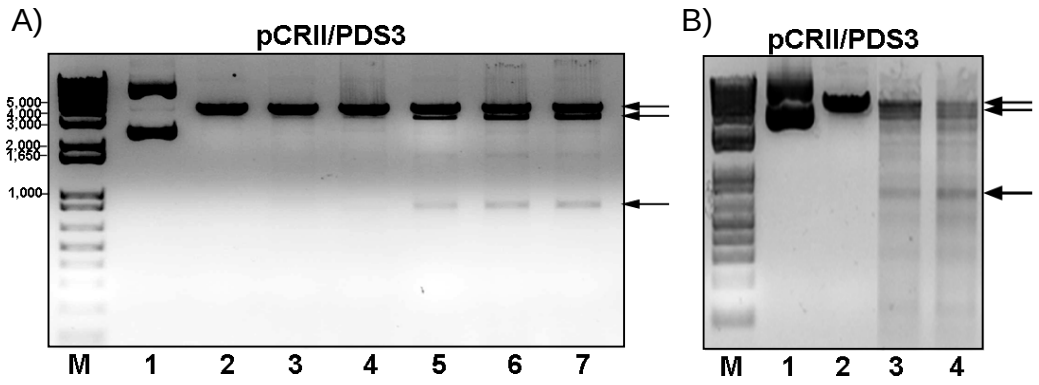


Figure 20. Validation of custom assembled TALEN activity using the DNA cleavage assays. TALEN heterodimer dTALEN.*PDS*3-1F/1R **a** and dTALEN.homodimer **b** digested target DNA of *PDS*3 cDNA fragment in the pCRII vector (pCRII/*PDS*3). Arrows indicate the digestion products of the TALEN proteins. *M* 1-kb plus marker. Lanes in **a** are *1* undigested pCRII/*PDS*3 plasmid, *2 Sac*I-digested pCRII/*PDS*3, *3* pCRII/*PDS*3 digested with dTALEN.*PDS*3-1F, *4* pCRII/*PDS*3 digested with dTALEN.*PDS*31R, *5–7*, pCRII/*PDS*3 digested with dTALEN.*PDS*3-1F/1R heterodimer. Lanes in **b** are *1* plasmid of pCRII/*PDS*3, *2 Sac*I-digested pCRII/*PDS*3, *3* and *4* pCRII/*PDS*3 digested with dTALEN.*PDS*3 homodimer.

3.5 Discussion

TALEs provide powerful and adaptable DNA-binding modules with great potential for a variety of biotechnological applications in agriculture and medicine. TALE RVDs dictate the DNA sequence specificity. The engineering of the RVDs of TALE repeats to bind to a user-selected DNA specificity is a challenging task. Because of their repetitive nature,

engineering by PCR approaches is complex and not reproducible. Moreover, commercial synthesis is feasible but relatively slow (8–12 weeks for a single TALE) and expensive (around 3,500 USD per TALE). To fully realize the potential of TALEs in different applications and organisms, a robust, efficient and inexpensive custom construction is required.

Here, we report the development of a rapid, robust, and highly efficient approach for dTALE repeat assembly in dTALE-TF and dTALEN backbones. To our knowledge, this approach is the simplest, quickest, and most efficient among the approaches reported in the literature. A library of 100 plasmids can be used to assemble 12 and 13 TALE repeats, thereby dictating DNA-binding specificity to any user-selected 13- or 14-bp targets, respectively. Our approach can also be used to construct dTALEs with a greater number of repeats. Because nine DNA fragments can be efficiently assembled using sequential restriction ligation cloning, our assembly approach can be used to assemble dTALEs with 18 repeats. Because the method uses sequence-verified clones, the integrity of the assembled clones is guaranteed.

Our dTALE assembly approach relies on the use of the type II restriction enzymes *PpuM*I, *BsmA*I, *BsmB*I, *Xho*I, and *Sac*I. These restriction enzymes are used in one sequential restriction-ligation reaction as indicated in Figure 18. A further advantage of our approach is that the assembly can be done in a variety of backbones, including activator, nuclease, repressor, etc. Additionally, the assembled repeats can be transferred from one backbone to another by *PpuM*I and *Sac*I restriction ligation. This is particularly advantageous for the simultaneous study of transcriptional activation and repression of a specific gene target (Mahfouz *et al.*, 2012). Our modular assembly tool kit is composed

of HD, NG, NI, and NK repeat modules, which have been shown to be highly specific in binding the C, T, A, and G nucleotides, respectively. This allows the design of highly specific dTALENs and dTALE activators and dTALEs fused with any functional domains. It was recently reported that the efficiency of dTALEs with NN repeat modules is much higher than those with NK repeat modules (Huang *et al.*, 2011). Accordingly, we also generated a library containing NN instead of NK so that users would have the choice of using either one or both. The number and identity of the repeats influence TALE target specificity (Boch *et al.*, 2009; Cermak *et al.*, 2011; Morbitzer *et al.*, 2011). Other factors, however, may also influence the binding *in vivo*. These factors include, but are not limited to, the chromatin status and the neighboring effects from the DNA sequence.

Because dTALENs require a dimerization step for their cleavage activity, two monomers in a homo- or heterodimers form are needed. Several studies have shown that dTALENs based on the *dHax3* backbone are highly efficient in generating DSBs in mammalian and plant cells. Given that two monomers (two heterodimers that bind 13 or 14 bp in a tail-to-tail orientation and that are separated by a 12–30-bp spacer region) are required, dTALENs with 12 or 13 repeats should be sufficient for the design of effective and highly specific TALE-based nucleases.

Several strategies have been previously described for dTALE assembly and for construction of chimeric activators and nucleases (Cermak *et al.*, 2011; Geissler *et al.*, 2011; Huang *et al.*, 2011; Li *et al.*, 2011b; Morbitzer *et al.*, 2011; Weber *et al.*, 2011; Zhang *et al.*, 2011a). Some of these previously reported protocols rely on PCR amplification of repeat modules followed by subsequent restriction ligations and transfer between more than one backbone vectors to achieve the final assembled clone. For

example, Zhang *et al.*, have reported a Golden Gate-based assembly approach. Their strategy involves two rounds of PCR amplification, restriction digestions and ligations of the assembled repeats in the final backbone (Zhang *et al.*, 2011a). Although this approach has allowed the assembly and functional testing of several dTALE-TFs it depends on PCR amplification and is therefore prone to sequence mutations and recombination between the repeats. Using a different approach, Li *et al.* have reported a modular construction of dTALENs by synthesis of eight sets of four RVD modules; these eight sets are used to assemble eight repeat subarrays, and the confirmed sequence of the eight repeat subarrays (for 16- or 24-bp targets) can be ligated together in a final dTALE backbone (Li *et al.*, 2011b). In addition, Weber *et al.* have used the native Avr*Bs3* scaffold to assemble dTALEs in two successive cloning steps; in the first step, five or six repeats are assembled in three preassembly vectors for positions 1-6, 7-12, and 13-17. Then, these three preassembled repeats are ligated together in a second cloning step to produce dTALEs with 17.5 repeats (Weber *et al.*, 2011). Similarly, Morbitzer *et al.* and Geissler *et al.* have reported dTALE assembly by two successive cut-ligation cloning steps that produced 17.5 repeats sufficient to confer high DNA-binding specificity (Geissler *et al.*, 2011; Morbitzer *et al.*, 2011).

Compared to these previously reported protocols, our tool kit represents a major improvement because it depends on a plasmid library of repeat modules to ensure the reliability of the sequence and the efficiency and reproducibility of a robust assembly. Moreover, our assembly strategy offers the following technical advantages: it does not require PCR amplification, which carries the risk of repeat recombination; it requires only one sequential restriction-ligation cloning step, which can be performed in one day

(including confirmation of the clones by restriction digestion analysis); and the assembled repeats can be transferred between different backbones with different functional domains, which facilitates the study of different genomic modifications at a particular locus.

Several other TAL effectors assembly approaches, developed after this study has been published, underscored the broad applicability of this work. For example, a more recent approach focused on high-throughput TALE repeat assembly methods. High-throughput solid phase assembly, called FLASH (Fast Ligation-based Automatable Solid-phase High-throughput) uses a solid-phase strategy to ligate several repeat units simultaneously. Using FLASH approach, more than 96 TALE arrays could be generated within one day (Reyon *et al.*, 2012). Other high-throughput methods included the ICA (Interactive Capped Assembly) (Briggs *et al.*, 2012) and an integrated chip method (Wang *et al.*, 2012) with the LIC (ligation-independent cloning) method, capable of generating over 600 TAL effectors in a single day (Schmid-Burgk *et al.*, 2013).

3.6 Significance

Our approach is particularly useful when a library of dTALE activators or repressors and dTALENs is needed for genome-wide applications. Restricted fragments of all possible RVDs can be ligated, and the library of dTALE activators, repressors, or nucleases can be transferred to a destination expression vector specific to a certain organism, including yeast, plant, mammalian cells, etc. The library of dTALEs in a plant expression vector can be transformed into *A. tumefaciens* and then into the plant species of interest. The generation of an overexpression TALE library in plants can be used, for example, to screen for phenotypes with useful agronomic traits. Following phenotypic selection and

heritability analysis, the TALE repeat sequences can be determined and their targets identified.

3.7 Materials and Methods

3.7.1 Design and efficient construction of TALE-TFs and TALENs using the *dHax3* and *dHax3*.N scaffolds

We previously reported the optimization of *dHax3* cDNA to reduce the homology between the 102-bp repeat units and the codon optimization for *in planta* expression (Mahfouz *et al.*, 2011). The *dHax3* and *dHax3*.N cDNAs, in the pENTR221 vector, were used as backbones to construct the TALE-TFs and TALENs, respectively. The *PpuM*I/*Sac*I fragment containing the repeat regions was divided into seven different fragments. Each fragment had two repeats and therefore two RVDs, except fragment 1, which had only one repeat. The sequence of each fragment and the primers used for amplification and cloning in pUC-minusMCS are provided in SI. These fragment clones were confirmed by sequencing, and their repeat RVDs were mutagenized to produce different RVDs. Moreover, because these fragments were short, we generated a whole library of all possible RVDs in each fragment by commercial synthesis (Blue Heron Bothell, WA, USA). As a result, we produced and amplified a library of fragments 1-7 containing all possible RVDs in every repeat in every fragment. We performed *in silico* assembly of the seven fragments, and the produced clones were identical to our *dHax3* and *dHax3*.N except for the user-selected RVDs. pUC-minusMCS was chosen to clone the library of all repeat fragments because it confers ampicillin resistance while the pENTR221 backbone assembly vector confers kanamycin resistance; these antibiotic markers facilitate the screening of the correctly assembled dTALE-TFs or dTALENs.

The correctly assembled dTALE-TF- or dTALEN-confirmed clones could be sub-cloned in any Gateway-based destination vector by LR recombination reactions for bacterial, yeast, plant, and mammalian expression.

3.7.2 One-step sequential restriction-ligation assembly protocol

Plasmids that contained the seven fragments were selected based on the desired TALE repeat sequence. Each fragment was restricted as follows: fragment 1 (F1) with *PpuM*I and *BsmA*I, purify target band of 192 bp; fragments 2, 3, 4, and 5 (F2, F3, F4, and F5) with *BsmA*I, purify target bands of 234, 161, 193, and 204 bp, respectively; fragment 6 (F6) first with *Xho*I and then with *BsmA*I, purify target band of 207 bp; fragment 7 (F7) with *Xho*I and *Sac*I, purify target band of 541 bp. Equi-molar amounts of fragments were ligated with the restricted backbone using T7 ligase at 23°C for 2 h or T4 ligase at 16°C for 6 h. A 2-µL volume of the ligation product was used to transform to TOP10 competent cells. Positive colonies were selected, and recombinant plasmid was purified and checked by restriction digestion with *Xho*I and *PpuM*I enzymes-the product band was expected to be about 1,190 bp. Sanger sequencing was used to further confirm the positive clones.

3.7.3 *Agrobacterium tumefaciens*-mediated transient expression and dTALE-TF activity assays

To test the activity of dTALE-TFs, we performed LR reactions between pENTR221/dTALE-TF entry clones and the pK2GW7 Gateway-compatible binary vector and generated overexpression clones pK2GW7/dTALE-TFs. These expression clones were transformed into A. *tumefaciens* GV3101 and used for transient expression in tobacco leaves. To assay the activity of dTALE-TFs *in planta*, we coinfiltrated or

separately infiltrated A. *tumefaciens* GV3101 harboring *BS3*::*uidA* and its dTALE activator into tobacco leaves. Infiltrated discs were collected 48 hpi. The GUS qualitative assay was performed by immersing leaf discs in GUS-staining solution (10 mM sodium phosphate, pH 7, 10 mM EDTA, 0.1% Triton X-100, 0.1% 5-bromo-4-chloro-3-indolyl-β-d-glucuronide, 1 mM potassium ferricyanide, 1 mM potassium ferrocyanide) at 37°C for 24 h. The discs were cleared in ethanol.

3.7.4 Bacterial expression and purification of dTALENs

The assembled dTALEN clones were confirmed by restriction digestions and sequencing. LR reactions were performed between the entry clones, pENTR221/dTALENs, and the Gateway-compatible pET32a expression vector (following the manufacturer's instructions) to generate pET32a.dTALENs. The expression clones were transformed into E. coli BL21, and protein expression was induced at 25°C for 5 h with 1 mM isopropyl β-d-1-thiogalactopyranoside. The TRX.6His.*dHax3*.N proteins were purified using Qiagen Ni-NTA agarose resin according to the manufacturer's instructions. The purified proteins, as homodimers or heterodimers for one specific target, were used for DNA cleavage assays as previously described (Mahfouz *et al.*, 2011).

CHAPTER 4

RNA-Guided Transcriptional Regulation in Plants via Synthetic dCas9 Based Transcription Factors

Reprinted with permission by John Wiley and Sons from:

Piatek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., Aouida, M. and Mahfouz, M.M. (2015) RNA-guided transcriptional regulation *in planta via* synthetic dCas9-based transcription factors. *Plant biotechnology journal* 13, 578-589. doi: 10.1111/pbi.12284.

4.1 Abstract

Targeted genomic regulation is a powerful approach to accelerate trait discovery and development in agricultural biotechnology. Bacteria and archaea use clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) regulatory systems for adaptive molecular immunity against foreign nucleic acids introduced by invading phages and conjugative plasmids. The type II CRISPR/Cas system has been adapted for genome editing in many cell types and organisms. A recent study used the catalytically inactive Cas9 (dCas9) protein combined with guide-RNAs (gRNAs) as a DNA-targeting platform to modulate gene expression in bacterial, yeast, and human cells. Here, we modified this DNA-targeting platform for targeted transcriptional regulation *in planta* by developing chimeric dCas9-based transcriptional activators and repressors. To generate transcriptional activators, we fused the dCas9 C-terminus with the activation domains of EDLL and TAL effectors. To generate a transcriptional repressor, we fused the dCas9 C-terminus with the SRDX repression domain. Our data demonstrate that dCas9 fusion with the EDLL activation domain (dCas9:EDLL) and the TAL activation domain (dCas9:TAD), guided by gRNAs complementary to selected promoter elements, induce strong transcriptional activation on *Bs3::uidA* targets in plant cells. Further, the dCas9:SRDX mediated transcriptional repression of an endogenous gene. Thus, our results suggest that the synthetic transcriptional repressor (dCas9:SRDX) and activators (dCas9:EDLL and dCas9:TAD) can be used as endogenous transcription factors to repress or activate transcription of an endogenous genomic target. Our data indicate that the CRISPR/dCas9 DNA-targeting

platform can be used in plants as a functional genomics tool and for biotechnological applications.

4.2 Introduction

The CRISPR/Cas9 system was reengineered for transcriptional regulation purposes by retaining the DNA-binding capabilities of the Cas9 protein while inactivating the nuclease cleavage domains: this is termed dead Cas9 (dCsas9) (Qi *et al.*, 2013). Tethering dCas9 protein to various effector domains (e.g., activators, repressors, or chromatin modifiers) has the potential to facilitate genomic screens through manipulating the expression time, abundance, and location of a particular mRNA. Moreover, such dCas9-based chimeras provide good platforms for recruitment of particular cellular TFs to specific genomic loci in order to enhance regulation of gene expression. The dCas9 protein can be tethered to various effector domains such as repressors (e.g., KRAB and SRDX domains) and activators (e.g., VP64, EDLL, and TAD domains) to elicit transcriptional changes. Targeting dCas9 protein to the promoter elements or coding regions of the target genes is sufficient to significantly repress gene expression in a site-specific manner through steric blockage of RNA polymerase II binding or inhibition of transcription elongation (Figure 21).

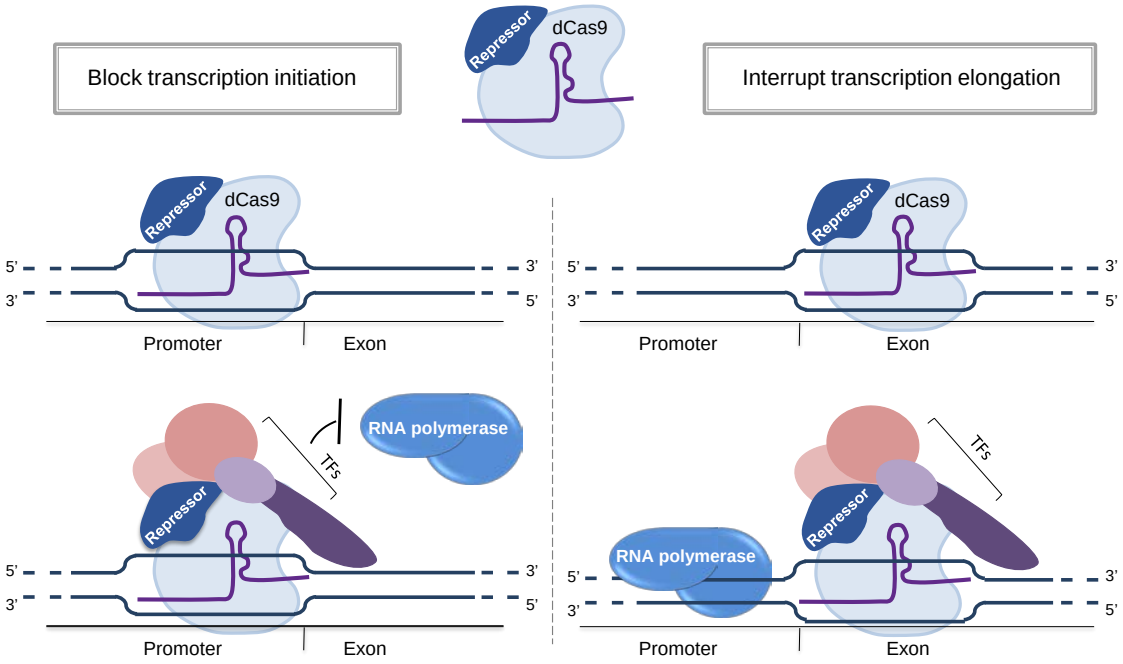


Figure 21. Schematic representation of the transcriptional repression mechanism. The dCas9 synthetic transcriptional repressor is guided to the target site by guide-RNAs (gRNAs). Binding to the target DNA sequence is mediated by dCas9, whereas the SRDX repressor domain recruits transcription factors that block *phytoene desaturase* (*PDS)* expression. Repression could occur due to a blockage of transcriptional initiation when dCas9:SRDX binds to the promoter region, or due to interruption of transcriptional elongation when binding takes place downstream of the transcriptional start site.

For instance, expression of dCas9 and sgRNA in bacteria efficiently repressed expression of reporter genes: 300-fold repression was observed when the dCas9-sgRNA complex was targeted to the mRFP gene (red fluorescent protein based reporter system) (Qi *et al.*, 2013) and over 100-fold reduction in fluorescence was observed when regions of a GFP reporter plasmid were targeted (Bikard *et al.*, 2013). Repression studies using dCas9 suggested that targeting the sgRNA to -10 and -35 elements in the promoter region produced the strongest gene modulation effects, regardless of the strand targeted. When the 5' UTR or coding region was targeted by the sgRNA, greater knockdown was

achieved when the non-template strand was targeted (this is termed the collision model) (Qi *et al.*, 2013). Furthermore, targeted repression was shown to be tunable and reversible and, when multiple sgRNAs were targeted to the same gene of interest, repression increased in an additive manner. When the CRISPR/dCas9 system is to be used for high-throughput screening experiments, any off-target effects can be determined using RNA sequencing. Qi *et al.* examined the specificity of the CRISPR/dCas9 system and found that only the target gene was affected, indicating that the system was highly precise. In addition to dCas9-mediated repression, dCas9 can be fused to activators to stimulate transcription of target genes. The dCas9 protein was converted to a transcriptional activator through N- and C- terminal fusions with the RNA polymerase omega (ω) subunit. The dCas9-ω fusion protein induced expression of the lacZ and GFP reporters when targeted at the optimal distance from the -35 promoter element (Bikard *et al.*, 2013). Modulation of gene expression by the CRISPR/dCas9 system is also effective in yeast and mammalian cells. Functional domains with the potential to recruit transcriptional factors were fused to dCas9 with the aim of improving the transcriptional modulation system. The dCas9 protein, alone or fused to a transcriptional repressor, effectively repressed gene expression in *Saccharomyces cerevisiae* (Gilbert *et al.*, 2013). Up to 53-fold repression was observed when dCas9 tethered to Mxi1 (a mammalian transcriptional repressor domain) was targeted against a reporter gene. Several studies demonstrate the potential of the CRISPR system for modulation of gene expression in human cell lines. Gilbert *et al.* used dCas9 coupled to the transcriptional activator VP16 and a single copy of the p65 activation domain (AD) to activate the expression of Gal4 UAS in an embryonic kidney reporter cell line (HEK293) expressing the Gal4 UAS-GFP

reporter. Reporter gene expression was successfully activated, with varying levels of activation observed with different fusion proteins (Gilbert *et al.,* 2013). These proof-of-principle experiments confirmed the effectiveness of the CRISPR/dCas9 system in mammalian cell cultures and showed that the system could control endogenous processes, such as cellular differentiation, and influence developmental pathways in human stem cells (Kearns *et al.*, 2014). Co-expression of dCas9 with multiple sgRNAs enabled synergistic activation of endogenous gene targets with no detectable off-target effects (Maeder *et al.*, 2013a; Perez-Pinera *et al.*, 2013a). Synergistic activation of gene expression by multiple sgRNAs was confirmed in other studies of endogenous human and mouse genes (Cheng *et al.*, 2013; Mali *et al.*, 2013a) and genome-wide expression profiling demonstrated the high specificity of the system (Cheng *et al.*, 2013). Similarly, fusions of dCas9 with KRAB, CS, or WRPW repressive chromatin modifier domains were constructed to increase efficacy of transcriptional repression. Reporter gene expression was repressed 15-fold and endogenous human gene expression was lowered by 60–80% using such fusion constructs when compared to dCas9 alone (Gilbert *et al.*, 2013). First report of CRISPR/dCas9 system to be used in plants is shown and discussed in Chapter 4. Expression of transiently expressed genes and endogenous genes was modulated in *Nicotiana benthamiana*. Various levels of gene repression or activation were reported, and these were affected by the position of the sgRNA with respect to the transcriptional start site, DNA target strand, and combinatorial effects of multiple sgRNAs (Piatek *et al.*, 2015).

4.3 Research Aim and Objectives

The aim of this study was to develop robust and efficient, RNA guided and dCas9 based platform for targeted gene regulation in plants. Objectives of this study included: generation of dCas9 based activators and repressors; validation of gene upregulation and downregulation in transient assays and that of endogenous genes; investigation of synergistic effect of the multiple gRNAs on gene activation or repression; verification of the influence of gRNA binding position and DNA strandedness in the promoter on gene expression profile; investigation into the competition factor between dCas9 based repressors and activators and their direct effect on gene expression level.

4.4 Results

4.4.1 Design and construction of dCas9-based synthetic transcriptional regulators

To test whether dCas9 can be used as a DNA-targeting module for site-specific transcriptional modulation in plants, we designed and constructed chimeric dCas9-based transcriptional activators and repressors. The human-codon-optimized dCas9 was PCR amplified from pdCas9-humanized plasmid (Qi *et al.*, 2013), (Supplementary Sequence 1), subcloned into pENTR-D/TOPO, and subsequently cloned into the pK2GW7 plant expression vector by LR-Gateway recombination cloning. The pK2GW7/dCas9 construct was used to generate dCas9 C-terminal fusions with functional transcriptional activation and repression domains. We constructed two chimeric transcriptional activators. First, we selected the EDLL domain from the ERF/EREBP family of transcriptional regulators because it functions as a strong activation domain, is transferable to other proteins, and active in proximal and distal positions from the target promoter (Tiwari *et al.*, 2012). To generate the dCas9:EDLL chimeric transcriptional activator, we fused a custom-

synthesized DNA fragment encoding the EDLL domain (EVFEFEYLDDKVLEELLDSEERKR) (Supplementary Sequence 2) in-frame to the C-terminus of dCas9 using *Mlu*I and *Eco*RI restriction enzymes (Figure 22). The second transcriptional activator was based on TAL effectors, which function as transcription factors *in planta* (Boch *et al.*, 2009; Morbitzer *et al.*, 2010). To generate the dCas9:TAD chimeric transcriptional activator, we fused a custom-synthesized DNA fragment encoding the *dHax3* TAL activation domain (amino acids 683–960) (Supplementary Sequence 3) in frame to the C-terminus of dCas9 using *Mlu*I and *Xho*I restriction enzymes (Figure 22). To generate a transcriptional repressor, we selected the SRDX motif because it is a potent and dominant repressor that could be fused to a variety of transcription factors and used for gene silencing (Hiratsu *et al.*, 2003). We constructed the chimeric dCas9 repressor by fusing the SRDX EAR motif (LDLDLELRLGFA) (Supplementary Sequence 4) to the dCas9 C-terminus using *Mlu*I and *Eco*RI restriction enzymes, to generate dCas9:SRDX (Figure 22).

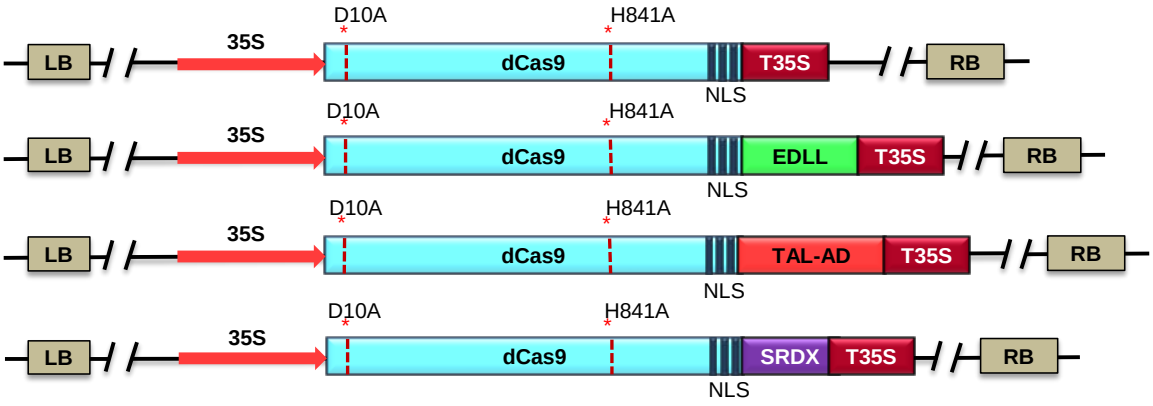


Figure 22. Schematic representation of dCas9 synthetic transcriptional regulators. The catalytically inactive Cas9 harboring the D10A and H841A nuclease-domain mutations was cloned into the pK2GW7 binary vector; dCas9 expression is driven by the 35S promoter and terminated by the 35S terminator. Only the T-DNA region is depicted.

The EDLL activation domain (shown in green) was fused to the dCas9 C-terminus to generate the dCas9:EDLL transcriptional activator. The dCas9 C-terminus was fused to the *dHax3* transcriptional activator-like (TAL) activation domain (shown in red) to generate the dCas9:TAL activation domain (TAD) transcriptional activator. The SRDX domain (shown in purple) was fused to the dCas9 C-terminus to generate the dCas9:SRDX transcriptional repressor.

To test the effects of the transcriptional activators, we constructed a target composed of the *Bs3* minimal promoter (Supplementary Sequence 5) driving the *uidA* reporter gene in the pKGWFS7 vector for transient transcriptional activation assays (Figure 23A). Targeting of dCas9 requires guide-RNA molecules; therefore, we constructed synthetic U6::gRNAs, that are a fusion of crRNA and tracrRNA, and subcloned them into the pYL156 binary vector (Supplementary Sequence 6 to 11). The synthetic gRNAs, 104 bp long with a 20nt sequence in the 5' end conferring target specificity, were designed to bind to several sites in the *Bs3* promoter on the sense and antisense strands. To test the activity of a synthetic transcriptional regulator on an endogenous target, we selected the *Nicotiana benthamiana phytoene desaturase* gene (*PDS*) (Supplementary Sequence 12). We generated three gRNAs that were capable of directing dCas9 to several positions in the *PDS* promoter or first exon (Figure 23B). The NGG PAM sequence (indicated in light blue), requirement for the *Streptococcus pyogenes* Cas9, was taken into consideration while designing all targets. To test the functionality of the transcriptional effectors, we separately transformed all binary constructs containing effectors, target, and gRNAs into *Agrobacterium tumefaciens*, and co-delivered them in different combinations into *Nicotiana benthamiana* leaves *via* agroinfiltration for transient-expression analysis (Figure 24).

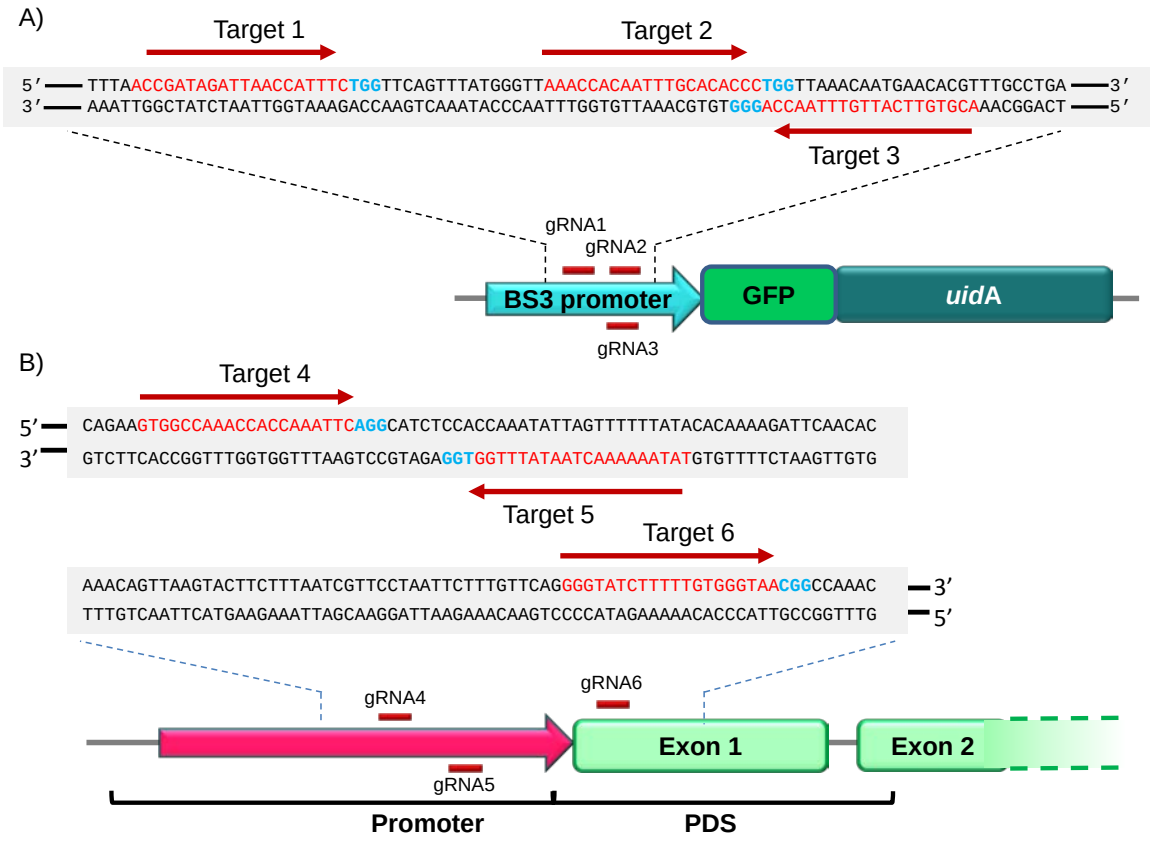


Figure 23. Schematic representation of transient and endogenous guide-RNA (gRNA) target sequences. (A) *Bs3* basal promoter driving *uidA* expression in the pKGWFS7 binary vector was used as a target for transient assays to test the synthetic transcriptional regulator activities on sense and antisense strands. The gRNA target sequences were selected on both sense and antisense strands preceding the protospacer-associated motif (PAM) NGG sequence (highlighted in light blue). (B) *Nicotiana benthamiana phytoene desaturase* (*PDS*) was used as an endogenous genomic target, and three gRNAs were designed to target the promoter region and first exon.

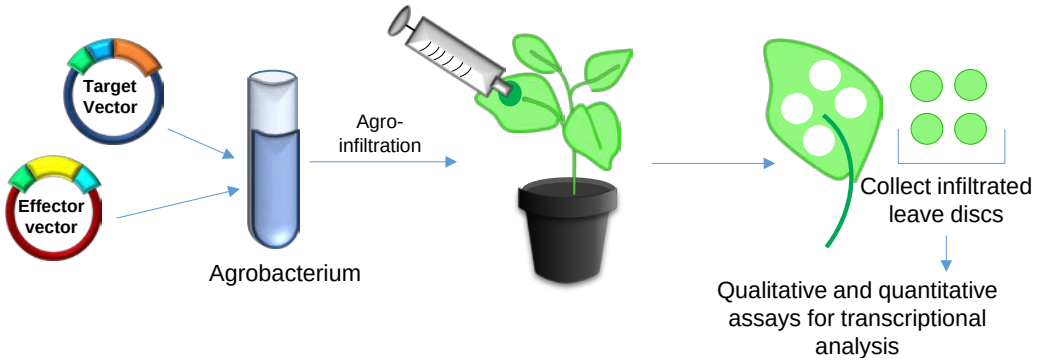


Figure 24. Overview of the experimental design. *Agrobacterium tumefaciens* GV3101 strain harboring guide-RNA (gRNA) target sequences (U6::gRNA) and dCas9 chimeric effector constructs (35S::dCas9, 35S::dCas9:TAD, 35S::dCas9:EDLL, and 35S::dCas9:SRDX) were mixed in different combinations and co-delivered into 3–4-week-old *Nicotiana benthamiana* leaves by agroinfiltration. At 36–48 hours post-infiltration, leaf discs were collected and subjected to qualitative and quantitative analysis. To test the activity of the transcriptional regulators on an *uidA* target gene, *Agrobacterium* GV3101 harboring *Bs3::uidA* in pKGWFS7 was mixed with effector and gRNA molecules.

4.4.2 The dCas9 synthetic activators mediate strong transcriptional activation

We assessed whether the chimeric dCas9:EDLL and dCas9:TAD transcriptional activators were guided by the gRNAs to the complementary sequence elements of the *Bs3* promoter, and whether they successfully activated expression of the *uidA* target gene (Figure 25).

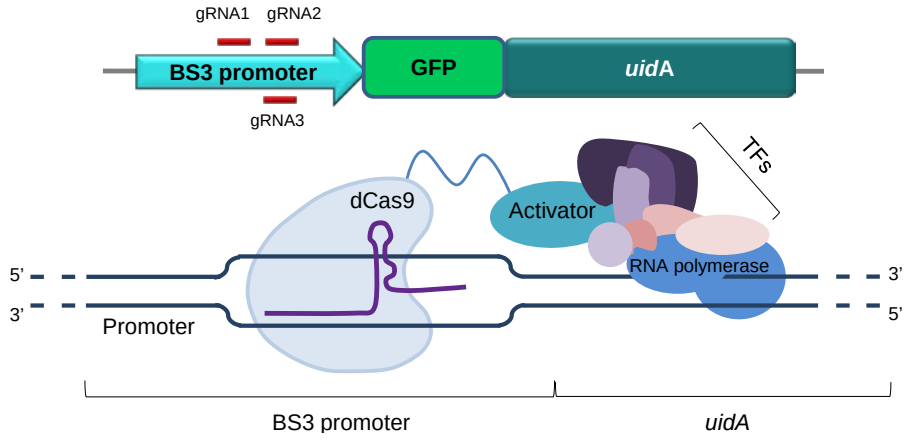


Figure 25. Schematic representation of the *Bs3*::*uidA* target and assembly of the CRISPR/dCas9 complex on the target DNA. Binding to the target DNA sequence is mediated by dCas9, whereas the transcriptional activation domains recruit the transcription factors involved in transcriptional machinery assembly and function.

The chimeric dCas9 transcriptional activators, *Bs3*::*uidA* target, and gRNAs were co-delivered in different combinations into 3–4-week-old *N. benthamiana* leaves *via* agroinfiltration. The gRNAs were targeted to different promoter elements on the sense and antisense strands of the *Bs3* promoter. Infiltrated leaf discs were collected 36–48 hours post-infiltration. Relative transcript abundance of *uidA* gene was then measured using the real time quantitative PCR method (Bruggeman *et al.*, 2014). Additionally, samples were also collected and subjected to qualitative and quantitative GUS analysis. The RT-qPCR results indicated that the dCas9:EDLL synthetic transcriptional activator significantly mediated site-specific transcriptional activation of the *uidA* reporter gene (Figure 26A). Several controls were used in this experiment, including gRNAs with non-complementary sequence to the *Bs3* promoter, *Bs3::uidA* target alone, and gRNA alone. Minimal background activation was observed when *Bs3*::*uidA* was separately agroinfiltrated into *N. benthamiana* leaves. Co-infiltration of dCas9:EDLL with gRNA6, which was non-complementary to the *Bs3* sequence, produced slightly higher level of

background expression than the *Bs3*::*uidA* control alone. Although background levels were observed with the control gRNA, the target gRNAs mediated significantly stronger and reproducible transcriptional activation.

To corroborate these transcriptional activation results with dCas9:EDLL function, we performed a similar set of experiments using the dCas9:TAD synthetic transcriptional activator. These results demonstrated that dCas9:TAD was guided to the *Bs3*::*uidA* promoter by gRNA2, and mediated significant site-specific transcriptional activation compared with that of control combinations (Figure 26B). These data indicate the versatility and reproducibility of the modified CRISPR/dCas9 system for targeted gene activation in plants.

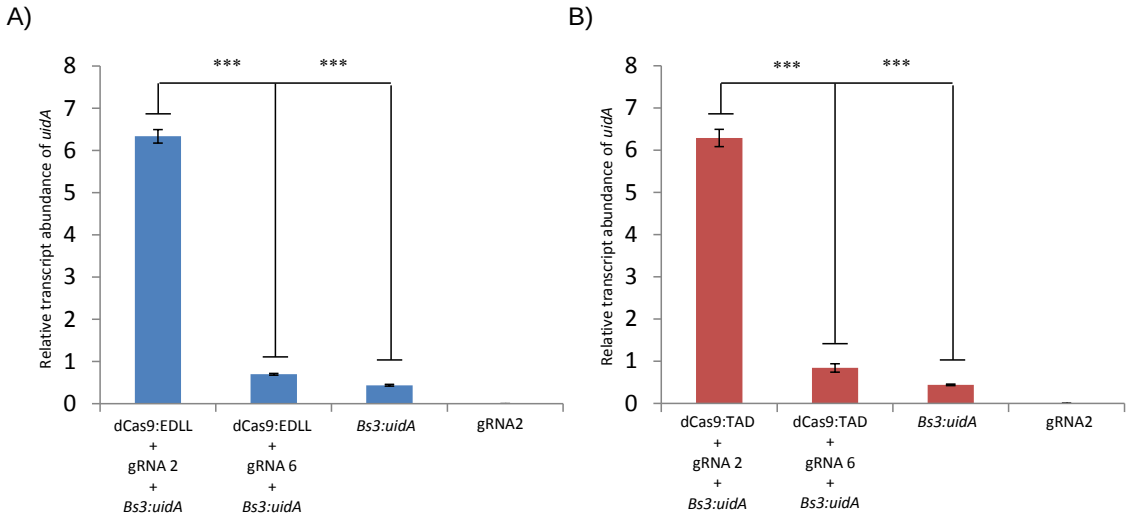


Figure 26. Transcriptional activation mediated by chimeric dCas9 transcriptional activators. (A) The dCas9:EDLL synthetic transcriptional activator strongly promoted transcription from the *Bs3* promoter compared with that of controls (*Bs3::uidA* and a non-specific guide-RNA (gRNA6)). (B) The dCas9:TAL activation domain (TAD) synthetic transcriptional activator promoted a higher level of transcription compared with that of controls (*Bs3::uidA* and a non-specific gRNA (gRNA6)). Quantitative analysis of *uidA* transcript levels are shown in both charts. Represented

results were normalized to the relative expression of the actin housekeeping gene. Error bars indicate standard error. Asterisks denote significantly different values according to the Student's test. * P<0.05; ** P<0.01; *** P<0.001.

4.4.3 Transcriptional activation is influenced by gRNA binding position in the promoter and enhanced by multiple gRNAs

The previous experiment demonstrated that chimeric CRISPR/dCas9 transcriptional regulators successfully activated transcription of a target gene *in planta*. Next, we assessed whether transcriptional activation was influenced by the gRNA-binding-site positions in the promoter relative to the transcriptional start site (TSS) (Hu *et al.*, 2014). Three gRNAs, complementary to *Bs3* promoter elements, were designed which bound to the promoter with variable distances from the TSS as follows: gRNA1, –297; gRNA2, –259; and gRNA3, –239. These gRNAs were co-delivered *via* agroinfiltration with each of the transcriptional activators (dCas9:EDLL and dCas9:TAD) and with *Bs3*::*uidA* in the pKGWFS7 target vector. We determined the level of transcriptional activation using the *N. benthamiana* transient-expression system. The RT-qPCR results and GUS quantitative and qualitative assays using dCas9:EDLL indicated that gRNA1 and gRNA2, which targeted the sense strand, mediated higher transcriptional activation compared with that of gRNA3, which targeted the antisense strand (Figure 27A and 27C). The results using dCas9:TAD indicated that transcriptional activation mediated by gRNA1 (target located further upstream of TSS on the sense strand) was lower than that of gRNA3 (target located on the antisense strand), whereas the highest level of transcriptional activation was mediated by gRNA2 (Figure 27B and 27D). These results indicate that the specific gRNA binding distance from the TSS might be important, and might influence transcriptional activity conferred by the CRISPR/dCas9 platform. Additional

characterization of these parameters using different promoters and widely spaced gRNA binding sites is required for further optimization and refinement of this promising targeted gene regulation technology.

To test whether multiple gRNAs could synergistically enhance transcriptional activation, we co-delivered dCas9:EDLL, *Bs3*::*uidA* target, and multiple gRNAs in different combinations into *N. benthamiana* leaves. We used the dCas9:EDLL and dCas9:TAD transcriptional activators and a combination of gRNAs that bound to the sense and antisense strands of the *Bs3* promoter.

In the dCas9:EDLL experiments, we observed different levels of transcriptional activation using different combinations of sense- and antisense-strand-binding gRNAs. For example, gRNA2 together with gRNA1 promoted the highest level of transcriptional activation. Mix of gRNA2 (targeting closer to the TSS on the sense strand) and gRNA3 (on the antisense strand) induced higher levels of *uidA* transcript than the mix of gRNA1 (further from the TSS) and gRNA3 (on the antisense strand). Surprisingly, mix of all three gRNA simultaneously did not induce the highest level of *uidA* transcript level suggesting that it might not be beneficial to add many gRNAs, targeted to sense and antisense strands, in close proximity. Overall, all four tested combinations significantly induced the *uidA* transcript levels when compared with the *Bs3::uidA* control alone (Figure 28A).

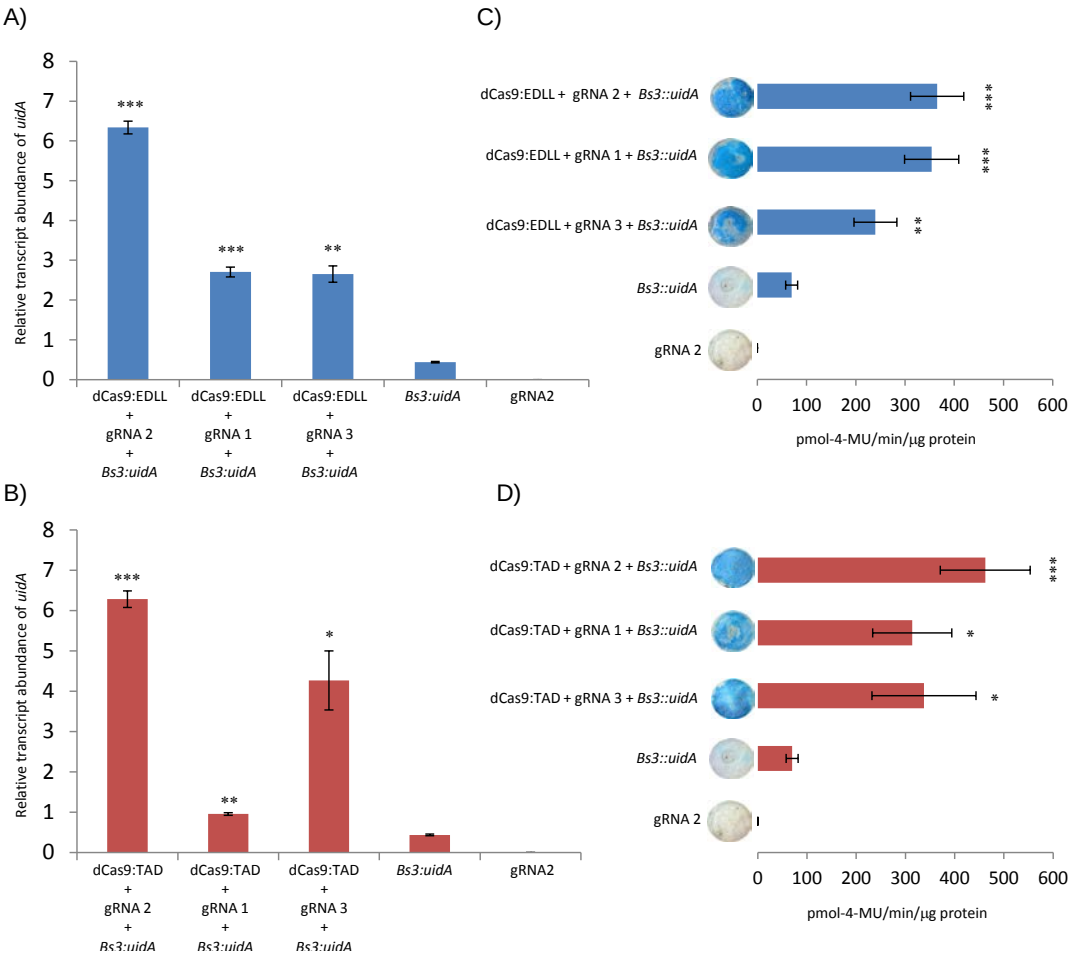


Figure 27. Positional effect of dCas9 promoter-binding elements on transcriptional activation of the *Bs3*::*uidA* target gene. (A) RT-qPCR analysis of *uidA* transcript aboundance mediated by dCas9:EDLL; gRNA 2 (targeting the sense strand closer to the transcriptional start site (TSS)) mediates stronger transcriptional activation of *uidA* gene than that of gRNA1 (targeting further away from the TSS) and gRNA3 (which targets the antisense strand). All three gRNAs separately induce significantly stronger transcriptional activation of *uidA* gene when compared with that of the *Bs3::uidA* control. Infiltration of gRNA2 alone did not yield any *uidA* transcripts. (B) RT-qPCR analysis of *uidA* transcript aboundance mediated by dCas9:TAD; the highest level of transcriptional activation is mediated by gRNA2, which targets the sense strand closer to the TSS. All gRNA targets mediate significantly higher levels of transcriptional activation compared with that of the *Bs3::uidA* control, however the lowest activation is achieved with the gRNA 1. Quantitative analysis of *uidA* transcript levels are shown in both charts. Represented results were normalized to the relative expression of the actin housekeeping gene. Error bars indicate standard error. Asterisks denote significantly different values according to the Student's test. * P<0.05; ** P<0.01; *** P<0.001. (C) Qualitative and quantitative

GUS assay mediated by dCas9:EDLL transcriptional activation. gRNA1 and gRNA2 target sequences on the sense strand mediate stronger transcriptional activation than that of gRNA3, which targets the antisense strand of the *Bs3* promoter, compared with that of the *Bs3::uidA* control. Infiltration of gRNA2 alone did not yield any GUS-positive signal. Statistical analysis using the Student-t test were done to confirm the significance of the *uidA* gene activation by the dCas9:EDLL, guided by gRNA2, gRNA1 and gRNA3, when compared to *Bs3::uidA* control, with p=0.000024, p=0.000042, p=0.0012 respectively. (D) Qualitative and quantitative GUS assay mediated by dCas9:TAD transcriptional activation; the highest level of transcriptional activation is mediated by gRNA2, which targets the sense strand of the *Bs3* promoter, closer to the TSS. All gRNA targets mediate significantly higher levels of transcriptional activation compared with that of the *Bs3::uidA* control. Statistical analysis using the Student-t test were done to confirm the significance of the *uidA* gene activation by the dCas9:TAD, guided by gRNA2, gRNA1 and gRNA3, when compared to *Bs3::uidA* control with p=0.00082, p=0.01, p=0.02 respectively. Error bars indicate standard error. Representative leaf discs used for the qualitative GUS assays are shown.

The same set of experiments was performed using dCas9:TAD. The highest transcriptional activation level was obtained using gRNA1 and gRNA2, which targeted the sense strand at proximal positions, similarly to that observed when using dCas9:EDLL. Combinations of the gRNAs targeting sense and antisense strands at positions closer to the TSS (gRNA2 and gRNA3) promoted higher transcriptional activation levels compared with those of gRNAs targeting positions further upstream of the TSS (gRNA1 and gRNA3) (Figure 28B). These results were consistent with those observed using the dCas9:EDLL transcriptional activator. Our data indicate that using sense-strand-binding gRNAs promoted higher transcriptional activation levels of the *uidA* target gene compared with those obtained using other gRNA combinations targeting sense and antisense strands simultaneously.

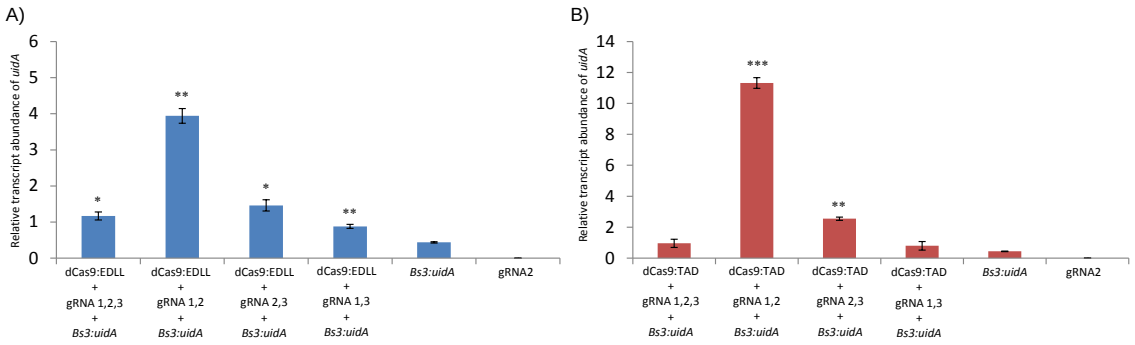


Figure 28. Synergistic effects of dCas9 complexes on transcriptional regulation of the *Bs3* promoter. (A) Using dCas9:EDLL, the highest transcriptional activation level was observed with gRNA1 and gRNA2, which targeted the *Bs3*-promoter sense strand when compared with to *Bs3::uidA* control. Combining gRNAs that targeted sense and antisense strands promoted higher transcriptional activation levels when targets were in proximal positions (gRNA2 and gRNA3) compared with those in distal positions (gRNA1 and gRNA3). Co-infiltration of all three gRNAs did not promote the highest activation level, however markedly induced transcription when compared to *Bs3::uidA* control. (B) Using dCas9:TAL activation domain (TAD), the highest transcriptional activation level was obtained using gRNA1 and gRNA2, which targeted the sense strand at proximal positions, similarly to that observed when using dCas9:EDLL. Lower transcriptional activation levels were obtained using gRNA1 and gRNA3, which targeted the sense and antisense strands at distal locations. Co-infiltration of all three gRNAs activated the *GUS* gene at a similar level as that observed in the experiment using dCas9:EDLL and all three gRNAs. Quantitative analysis of *uidA* transcript levels are shown in both charts. Represented results were normalized to the relative expression of the actin housekeeping gene. Error bars indicate standard error. Asterisks denote significantly different values according to the Student's test. * P<0.05; ** P<0.01; *** P<0.001.

Next, we investigated whether the dCas9 chimeric activators could enhance expression levels of the endogenous *N. benthamiana PDS* gene. We used the chimeric dCas9:EDLL and dCas9:TAD constructs with three gRNA molecules designed to target the sense and antisense strands of the *PDS* promoter and first exon. RT-qPCR analysis indicated that the *PDS* expression level were markedly higher in cells transiently expressing dCas9:EDLL transcriptional activator compared with relative *PDS* transcript abundance

of mock control. Co-delivery of all three gRNAs synergistically enhanced *PDS* expression levels (Figure 29A). Agroinfiltration of single gRNAs with dCas9:TAD significantly increased the *PDS* transcript levels only when guided to the sense strand of exon 1 (gRNA6), or when guided by all three gRNAs simultaneously. Similarly as in dCas9:EDLL mediated activation, multiple gRNAs induced a significant synergistic effect on the level of *PDS* transcript abundance when compared with gene expression driven by a single gRNA molecules (Figure 29B).

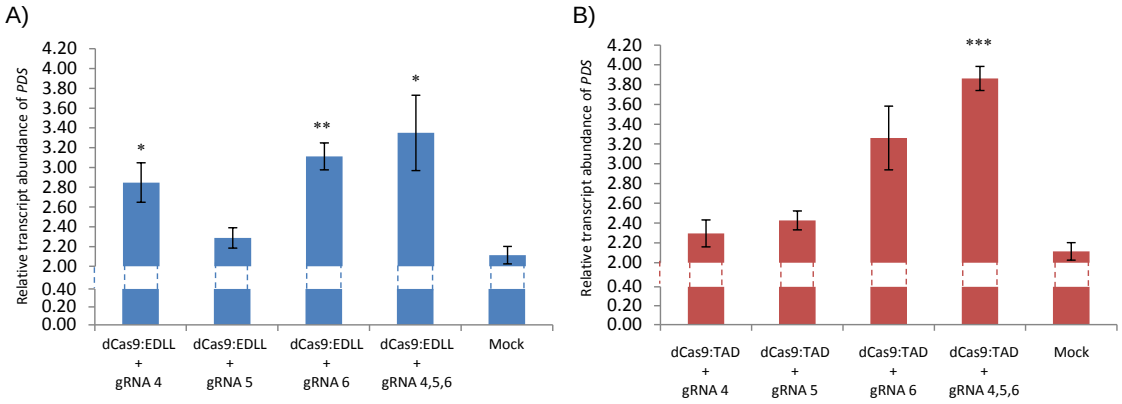


Figure 29. RT-qPCR analysis of endogenous *PDS* transcriptional activation using dCas9 synthetic transcriptional activators. (A) Analysis of *PDS* expression levels activated by dCas9:EDLL. Significantly higher levels of *PDS* transcript, in comparison to the *PDS* expression levels in mock control, were observed when guiding the dCas9:EDLL activator by single guide-RNAs (gRNAs) targeting the sense (gRNA4) strand in the promoter and the sense strand of exon 1 (gRNA6). Co-delivery of all three gRNAs did synergistically enhanced *PDS* expression levels when compared with the mock control. (B) Analysis of *PDS* expression levels activated by dCas9:TAD. dCas9:TAD activator only significantly increased the *PDS* transcript levels when guided to the sense strand of exon 1 (gRNA6), or when guided by all three gRNAs simultaneously when compared with the mock control. Quantitative analysis of *PDS* transcript levels are shown in both charts. Represented results were normalized to the relative expression of the actin housekeeping gene. Error bars indicate standard error. Asterisks denote significantly different values according to the Student's test. * P<0.05; ** P<0.01; *** P<0.001.

4.4.4 Transcriptional repression of *phytoene desaturase* by dCas9:SRDX

The CRISPR/dCas9 system has been used to functionally repress several target genes in bacterial and mammalian cells (Qi *et al.*, 2013). Here, we tested if dCas9 fused to the repressor domain SRDX (dCas9:SRDX) could mediate site-specific repression of plant genes. We constructed a chimeric dCas9:SRDX synthetic transcriptional repressor and three gRNA molecules designed to target and bind to the promoter region and first exon of the genomic *phytoene desaturase* in *N. benthamiana*. We co-delivered dCas9:SRDX together with the designed gRNAs *via* agroinfiltration of 3–4-week-old *N. benthamiana* leaves. Infiltrated leaf samples were collected 36–48 hours post-infiltration, and gene expression levels were determined by RT-qPCR analysis. Previous reports indicated that dCas9 alone could repress target genes, possibly by interfering with the transcriptional machinery or due to structural hindrance of RNA PolII (Figure 30A) (Gilbert *et al.*, 2013; Larson *et al.*, 2013; Qi *et al.*, 2013). Therefore, we tested if dCas9 alone could induce transcriptional repression *in planta*. The results showed that, *PDS* transcript levels were markedly reduced when dCas9 was targeted to the sense and antisense strands on the promoter region of *PDS* gene as well as the sense strand of the first exon, when compared to the *PDS* expression levels in control conditions. The strongest repression of the *PDS* gene was observed when dCas9 was guided by all three gRNAs simultaneously (Figure 30B)

Next, we tested the activity of dCas9:SRDX repressor on *PDS* transcription. dCas9:SRDX markedly reduced the gene transcript abundance when dCas9:SRDX was targeted to the sense strand on the promoter region of *PDS* gene as well as the sense strand of the first exon, when compared to the *PDS* expression levels in the mock control

(Figure 30C). dCas9:SRDX clearly repressed *PDS* transcription when combined with all three gRNAs targeting exon 1 and the promoter region sense and antisense strands. These data demonstrate that the dCas9 alone and dCas9:SRDX synthetic transcriptional repressor can effectively mediate transcriptional repression of genomic *PDS* target *in planta*.

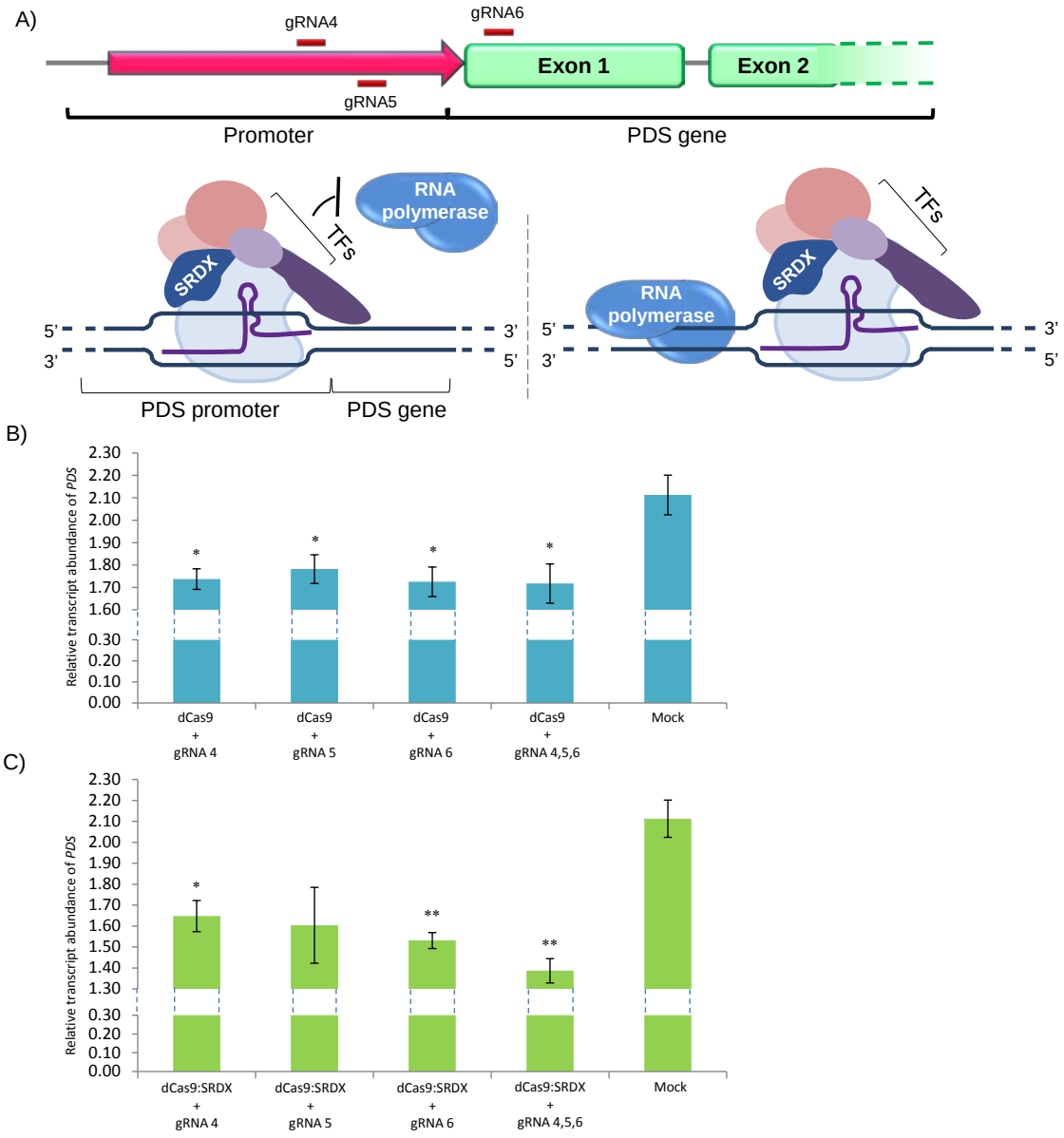


Figure 30. Transcriptional repression of genomic phytoene desaturase in *Nicotiana benthamiana* mediated by dCas9 and dCas9:SRDX. (A) Schematic representation of the transcriptional repression mechanism. The dCas9 synthetic transcriptional repressor is guided to the target site by guide-RNAs (gRNAs). Binding to the target DNA sequence is mediated by dCas9, whereas the SRDX repressor domain recruits transcription factors that block *phytoene desaturase* (*PDS)* expression. Repression could occur due to a blockage of transcriptional initiation when dCas9:SRDX binds to the promoter region, or due to interruption of transcriptional elongation when binding takes place downstream of the transcriptional start site. (B) RT-qPCR analysis of *PDS* expression levels repressed by dCas9. *PDS* expression pattern using dCas9 alone is significantly reduced when dCas9 is targeted to the sense and antisense strands on the promoter

region of *PDS* gene as well as the sense strand of the first exon, when compared to the *PDS* expression levels of the mock control. The strongest repression of the *PDS* gene is observed when dCas9 is guided by all three gRNAs simultaneously. (C) RT-qPCR analysis of *PDS* expression levels repressed by dCas9:SRDX. *PDS* expression pattern using dCas9:SRDX is more prominent than of that when using dCas9 alone. dCas9:SRDX significantly reduces the gene transcript abundance when dCas9:SRDX is targeted to the sense strand on the promoter region of *PDS* gene as well as the sense strand of the first exon, when compared to the *PDS* expression levels in the mock control. Similarly to the dCas9 alone experiment, dCas9:SRDX reduces *PDS* transcript levels the most when guided by all three gRNAs simultaneously. Quantitative analysis of *PDS* transcript levels are shown in both charts. Represented results were normalized to the relative expression of the actin housekeeping gene. Error bars indicate standard error. Asterisks denote significantly different values according to the Student's test. * P<0.05; ** P<0.01; *** P<0.001.

4.4.5 Synthetic transcriptional repressor interferes and/or competes with transcriptional activation machinery

The SRDX domain is a dominant repressor that retains its function when fused to transcription factors containing activation domains (Heyl *et al.*, 2008; Hiratsu *et al.*, 2003; Mahfouz *et al.*, 2012; Takada, 2013). Therefore, we analyzed the effect of dCas9:SRDX on transcriptional activation mediated by dCas9:EDLL and dCas9:TAD synthetic transcriptional activators by co-delivering repressor and activator constructs simultaneously with single or multiple gRNAs targeting the *Bs3* promoter driving the *uidA* reporter gene in transient assays. This system can test dCas9:SRDX transcriptional interference with dCas9:EDLL and dCas9:TAD, because a single gRNA can guide either the repressor or activator to the *Bs3::uidA* target. RT-qPCR data show that co-delivery of dCas9:SRDX with dCas9:EDLL and gRNA2 induced significantly lower transcriptional activation levels of *uidA* gene when compared with those normally mediated by dCas9:EDLL (Figure 31A). Similarly, co-delivery of dCas9:SRDX significantly reduced

the normal transcriptional activation mediated by dCas9:TAD when guided by gRNA2 (Figure 31B).

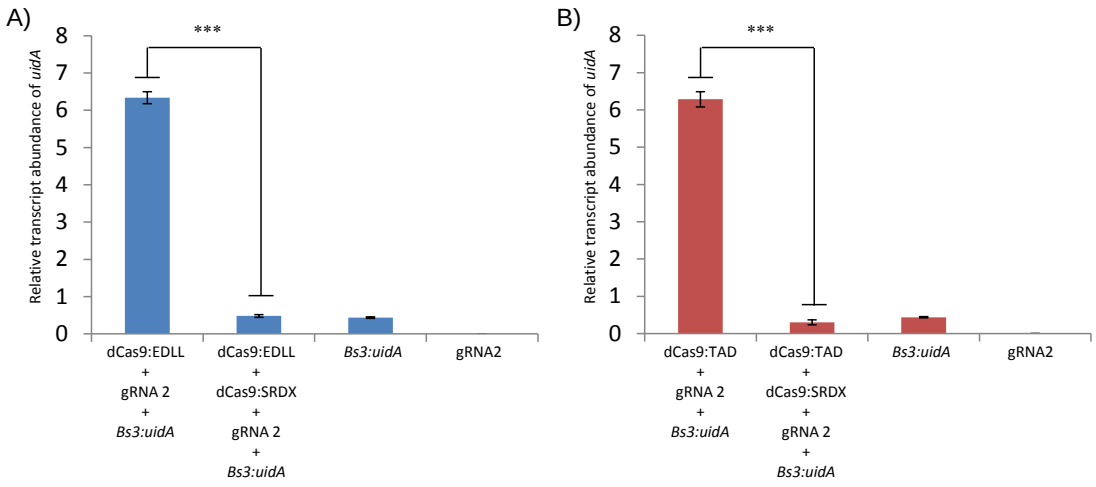


Figure 31. Analysis of dCas9:SRDX transcriptional repressor interference with dCas9:EDLL and dCas9:TAD transcriptional activators. (A) Co-delivery of dCas9:EDLL and gRNA2 induced high expression levels of the transient *uidA* target; these levels decreased significantly in the presence of the dCas9:SRDX transcriptional repressor. (B) Co-delivery of dCas9:SRDX along with dCas9:TAD and gRNA2 also significantly reduces the transcriptional activation conferred by dCas9:TAD and gRNA2. Quantitative analysis of *uidA* transcript levels are shown in both charts. Represented results were normalized to the relative expression of the actin housekeeping gene. Error bars indicate standard error. Asterisks denote significantly different values according to the Student's test. * P<0.05; ** P<0.01; *** P<0.001.

Both experiments were confirmed by the data obtained from the qualitative and quantitative GUS stain assays (Figure 32). These results indicate that dCas9:SRDX interferes with transcriptional activation mediated by both of the synthetic transcriptional activators.

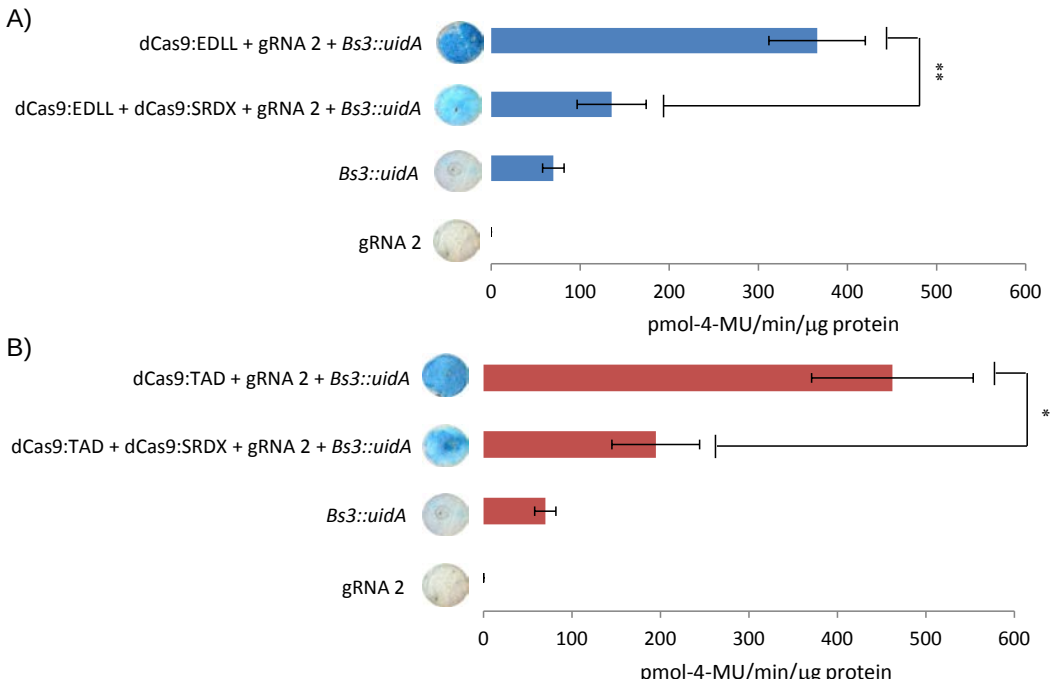


Figure 32. Analysis of dCas9:SRDX transcriptional repressor interference with dCas9:EDLL and dCas9:TAD transcriptional activators. (A) Co-delivery of dCas9:EDLL and gRNA2 induced high expression levels of the *uidA* target gene; these levels decreased dramatically in the presence of the dCas9:SRDX transcriptional repressor. Statistical analysis using the Student-t test were done to confirm the significance of the *uidA* gene repression by the dCas9:SRDX with p=0.0013. (B) Co-delivery of dCas9:SRDX along with dCas9:TAD and gRNA2 also significantly reduces the transcriptional activation conferred by dCas9:TAD and gRNA2. Statistical analysis using the Student-t test were done to confirm the significance of the *uidA* gene repression by the dCas9:SRDX with p=0.02. Error bars indicate standard error. Representative leaf discs used for the qualitative GUS assays are shown.

Our previous experiments showed that using multiple gRNAs resulted in significant activation of the target gene (Figure 29). Therefore, we tested if dCas9:SRDX could reduce the transcriptional activation mediated by multiple gRNAs by co-delivering dCas9:SRDX with either dCas9:EDLL or dCas9:TAD guided with two or three different gRNAs. This approach enables formation of the CRISPR/dCas9 activation and repression complexes at every promoter element. The results exhibited a significant reduction in transcriptional activation of *uidA* gene when gRNA1 and gRNA2 were used together as

well as when gRNA1, gRNA2 and gRNA3 were used simultaneously, indicating the dCas9:SRDX interference with transcriptional activation (Figure 33).

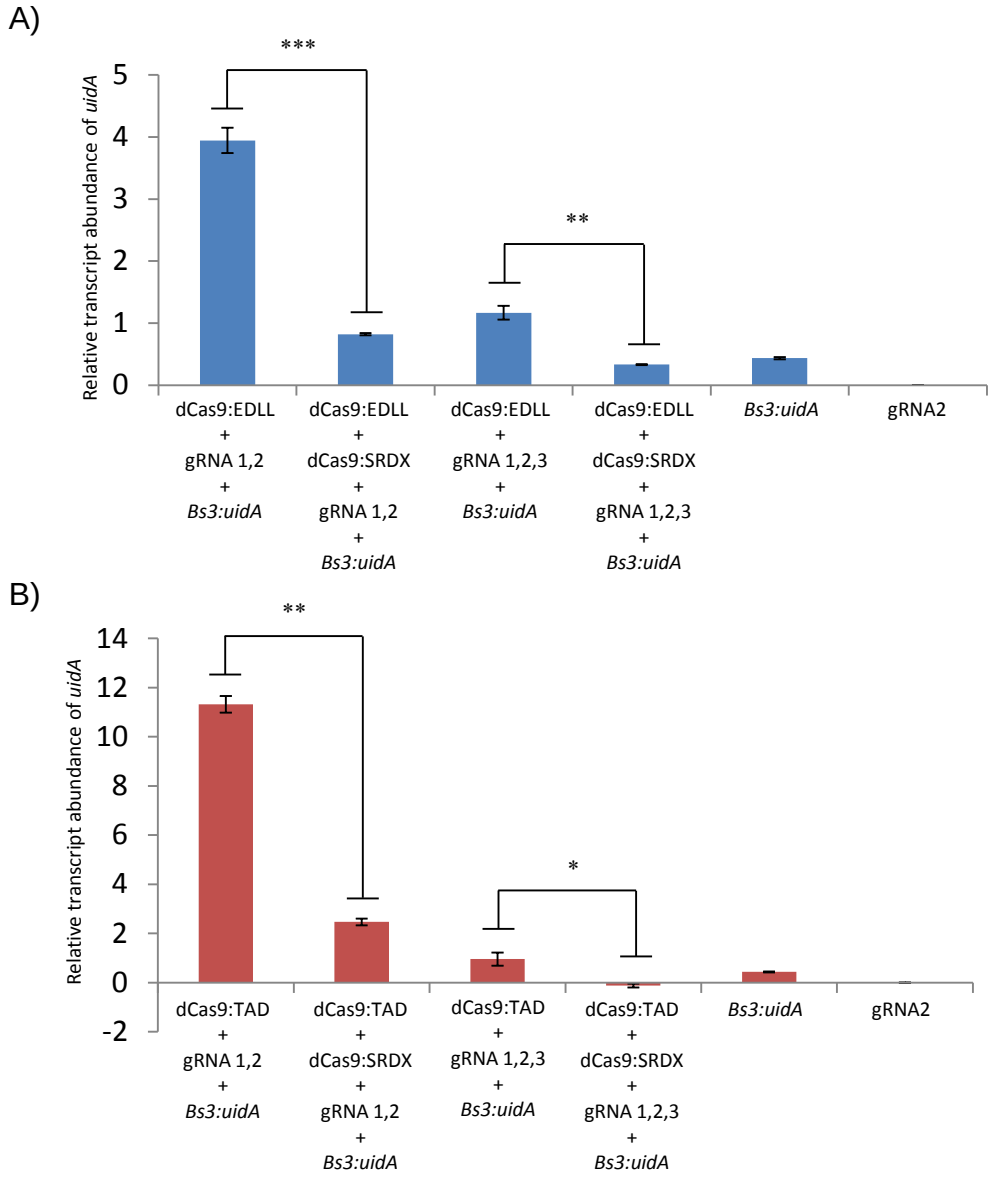


Figure 33. Competition between chimeric dCas9 activators and repressor using multiple combinations of guide-RNAs (gRNAs). (A) The dCas9:SRDX transcriptional repressor markedly reduced dCas9:EDLL-induced transcriptional activation using gRNA1 and gRNA2, which targeted the *Bs3* promoter sense strand. Co-delivery of gRNAs targeting both sense and antisense strands of the *Bs3* promoter (gRNA1, gRNA2, and gRNA3) resulted in competition and also significant repression of *uidA* gene. (B) The dCas9:SRDX transcriptional repressor competed with and reduced

dCas9:TAD-induced transcriptional activation. The repression of transcriptional activation conferred by dCas9:SRDX was significant when multiple gRNAs targeting both sense and antisense strands were used. Quantitative analysis of *uidA* transcript levels are shown in both charts. Represented results were normalized to the relative expression of the actin housekeeping gene. Error bars indicate standard error. Asterisks denote significantly different values according to the Student's test. * P<0.05; ** P<0.01; *** P<0.001.

Both experiments were confirmed by the data obtained from the qualitative and quantitative GUS stain assays (Figure 34).

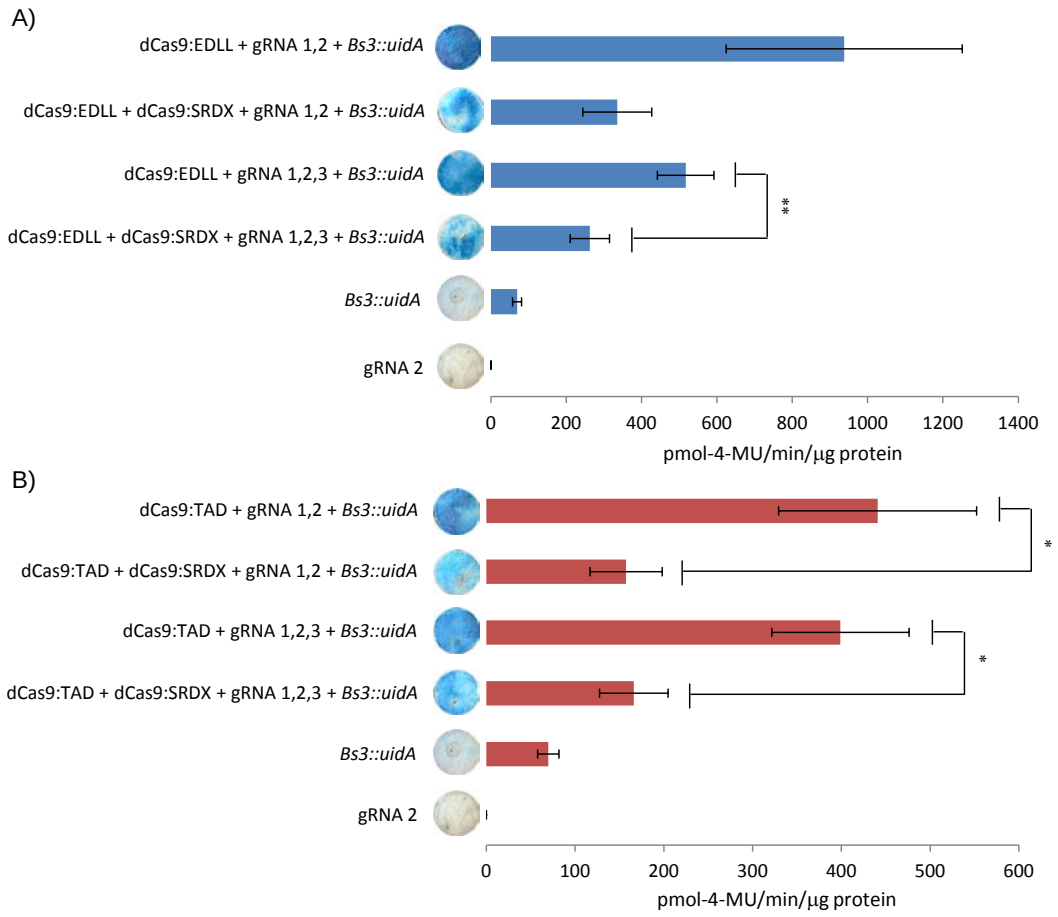


Figure 34. Competition between chimeric dCas9 activators and repressor using multiple combinations of guide-RNAs (gRNAs). (A) The dCas9:SRDX transcriptional repressor reduced dCas9:EDLL-induced transcriptional activation using gRNA1 and gRNA2, which targeted the *Bs3*-promoter sense strand. Co-delivery of gRNAs targeting both sense and antisense strands of the *Bs3* promoter (gRNA1, gRNA2, and gRNA3) resulted in competition and significant repression

of the *uidA* transcript. Statistical analysis using the Student-t test were done to confirm the significance of the *uidA* gene repression by the dCas9:SRDX with p=0.008. (B) The dCas9:SRDX transcriptional repressor competed with and reduced dCas9:TAD-induced transcriptional activation when guided by gRNA 1 and gRNA2 with p=0.04. The repression of transcriptional activation conferred by dCas9:SRDX was also significant when multiple gRNAs targeting both sense and antisense strands were used with p=0.01. Statistical analysis were done using the Student-t test. Error bars indicate standard error. Representative leaf discs used for the qualitative GUS assays are shown.

4.5 Discussion

The CRISPR/dCas9 system successfully targeted DNA in bacterial and human cells (Bikard *et al.*, 2013; Qi *et al.*, 2013), and dCas9 chimeras fused to transcriptional domains mediated targeted transcriptional regulation (Perez-Pinera *et al.*, 2013a). Previous work showed that dCas9 functioned as a transcriptional repressor by interfering with the transcriptional machinery or by hindering RNA PolII (Qi *et al.*, 2013). In the present work, we showed that the CRISPR/dCas9 system can be used as a platform for targeted gene regulation *in planta*. We employed the catalytically inactive dCas9 combined with gRNA molecules that target specific DNA sequences in the promoter of target genes. We optimized the CRISPR/dCas9 system for *in planta* expression by subcloning dCas9 into a binary vector driven by the tandemly duplicated 35S constitutive promoter (pK2GW7).

We generated two dCas9 synthetic transcriptional activators. The EDLL domain of the ERF/EREBP family of plant transcription factors was selected because it is a strong transcriptional activator, easily transferable to other proteins, and binds positions either proximal or distal to the TATA box (Tiwari *et al.*, 2012). The *dHax3* TAL effector protein (TAD) of phytopathogenic *Xanthomonas* spp. was selected due to extensive characterization and strong activation of plant target genes. We selected the *Bs3*

promoter, of the resistance (R) gene from pepper plants (Romer *et al.*, 2009), as a target fused to *uidA* of a GUS reporter vector. Transcriptional activation levels were measured by RT-qPCR and qualitatively and quantitatively using GUS transient-expression assays. gRNAs complementary to either the *Bs3* promoter sense or antisense strand were designed and constructed into the pYL156 binary vector under control of the *U6* promoter.

Real time quantitative PCR analysis of the *uidA* gene transcript indicated that dCas9:EDLL and dCas9:TAD both mediated 14-fold transcriptional activation compared with that of *Bs3::uidA* control (Figure 26), whereas quantitative GUS assays revealed that dCas9:EDLL mediated 5-fold and dCas9:TAD mediated 6-fold transcriptional activation when guided by gRNA2 compared with that of *Bs3::uidA* control (Figure 27C and 27D). Using a control that targeted DNA outside of the *Bs3* promoter, we observed some level of transcriptional activation with dCas9:EDLL and dCas9:TAD, respectively (Figure 26). This can be explained by the fact that the high copy numbers of the transiently expressed target are scanned by the CRISPR/dCas9 synthetic transcriptional activators to find complementary sequences of gRNAs (Hu *et al.*, 2014; Jinek *et al.*, 2014; Sternberg *et al.*, 2014) and thus the process of scanning for and binding the complementary target could mildly activate gene expression. This interpretation will require further studies and analysis. These data demonstrate successful targeting and transcriptional regulation of plant gene expression as determined by the activation of *uidA* gene detected by the relative transcript abundance as well as qualitative and quantitative presence of the GUS enzyme.

We assessed whether gRNA molecules targeted to the antisense strand could induce activation of *Bs3::uidA* target similar to those targeting the sense strand. The results showed that transcriptional activation of gRNAs targeting the sense strand closer to the TSS was higher than that of gRNAs targeting the antisense strand, possibly due to structural flexibility with respect to the transcriptional machinery and RNA polymerase (Figure 27). We then tested the effects of gRNA binding relative to the TSS on transcriptional activation of *Bs3::uidA* target driven by dCas9:EDLL and dCas9:TAD. The gRNA2 bound closest to the TSS on the sense strand and mediated higher expression compared with that of gRNA1 by two fold when delivered with dCas9:EDLL and 6 fold when delivered with dCas9:TAD. This difference was more significant with dCas9:TAD than with dCas9:EDLL. These data indicate that different dCas9 synthetic transcriptional activators could require targeting to different positional distances from the TSS for optimum transcriptional activation levels, primarily due to structural requirements for the chimeric dCas9 fusions.

We assessed potential synergistic effects of multiple gRNAs, when used simultaneously, targeted to sense and antisense strands on transcriptional activation of *Bs3::uidA* target (Figure 28). Using dCas9:EDLL, co-delivery of gRNA1 and gRNA2 targeting the sense strand resulted in the strongest transcriptional activation of *Bs3::uidA* target compared with those of other combinations. Similar results were obtained when using dCas9:TAD, (Figure 28B). The weakest transcriptional activation mediated by dual gRNAs was obtained with gRNA1 targeting the proximal sense strand and gRNA3 targeting the antisense strand. These data indicate that gRNAs targeted to the promoter antisense strand might structurally hinder RNA PolII binding and the transcriptional initiation

machinery. Co-delivery of multiple gRNAs that assemble transcriptional complexes on different promoter regions appears to elevate transcriptional activation. However, it may not be necessary to simultaneously use several gRNAs for optimum transcriptional activation; the number of gRNAs and dCas9 complexes may reach a saturation point beyond which transcriptional activity could be attenuated due to structural interference of neighboring complexes.

These experiments demonstrated CRISPR/dCas9 transcriptional activation of transient targets. Next, we generated a transcriptional repressor to specifically downregulate a genomic target. We fused the SRDX dominant repression domain, which belongs to the ERF family of transcription factors (Kagale and Rozwadowski, 2011; Ohta *et al.*, 2001), to the dCas9 C-terminus. The SRDX domain downregulated target genes when fused to transcription factors or TAL effectors (Mahfouz *et al.*, 2012). The dCas9:SRDX synthetic transcriptional repressor was co-delivered with gRNAs targeting different positions on the *PDS* promoter and exon 1 in *Nicotiana benthamiana*. The RT-qPCR results showed that dCas9 alone and dCas9:SRDX mediated transcriptional repression of genomic *PDS* (Figure 30).

Our study provides the first evidence that the CRISPR/dCas9 platform can be used for targeted plant genome regulation. We designed and generated two synthetic transcriptional activators and one transcriptional repressor, and demonstrated targeted transcriptional regulation on transient and genomic targets *in planta*. Several parameters require further optimization, including generating shorter functional versions of dCas9 that could resolve some of the observed effects of structural interference.

4.6 Significance

CRISPR/dCas9 transcriptional regulation system carries advantages over the existing platforms including the amenability to multiplexing and possibility of using viruses to systemically deliver gRNAs. Generation of dCas9 synthetic transcriptional regulators in model plants such as *Arabidopsis* could facilitate functional analysis of genes under different nutritional, hormonal, or environmental conditions (Lowder *et al.*, 2015). Profiling of abiotic stress responses through identification of downstream processes and subsequent control by synthetic transcriptional regulators will lead to the generation of new synthetic genetic circuits and molecular pathways that can be utilized for the generation and use of novel allelic variants for breeding in crops.

The gRNAs can be delivered using symptomless viral vectors, like tobacco rattle virus (TRV), that mediate systemic plant infection. This approach could enable construction of a functional genomics platform applicable for many plant species. A trait-discovery platform could be generated by co-delivery of multiple gRNAs targeting single, multiple, or all members of a gene family (Figure 35). The present study opens exciting possibilities to address important questions in plant biology and gene regulation, and provides a foundation for future development of potentially powerful applications in agricultural biotechnology.

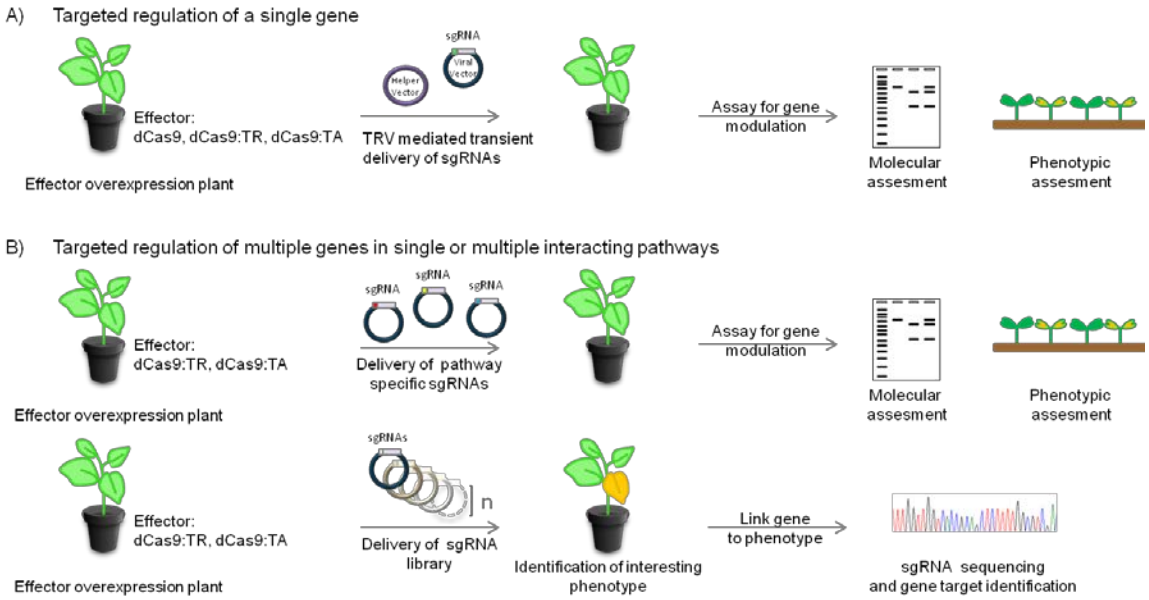


Figure 35. Modalities for targeted interrogation of gene functions in plants. A) Targeted genome regulation using synthetic transcriptional regulators and epigenomic modulators can be applied to interrogate the gene function. For example, plants overexpressing effectors, such as dCas9:TR or dCas9:TA, could be used in transient functional analysis by employing the viral mediated sgRNA delivery. Viral mediated delivery of sgRNA, using TRV is a system of choice because TRV has wide host range, is symptomless, multiplies and quickly moves from one cell to another (systemic infection) as it is continuously expressing the sgRNA. Subsequently, targeted gene modulation can be analyzed using different molecular and phenotypic assays. B) Platform described in A) can be used to simultaneously study the function of multiple genes (multiplexing), in which case multiple sgRNA would be simultaneously delivered. Moreover, sgRNA libraries can be used to perform genome wide interrogation of gene functions under different developmental and physiological cues, e.g. analyzing gene functions under stress conditions including salt and heat. Causal genes controlling a certain phenotype can be identified by isolating and sequencing the sgRNA.

4.7 Materials and Methods

4.7.1 Plasmid construction

Human-codon-optimized dCas9 was obtained from AddGene (Product number 44246) (https://www.addgene.org/44246/). The dCas9 was amplified from the original plasmid with primers 5'dCas9-F and 3'dCas9-R (Supplementary Table 1) and subcloned into the

pENTR-D/TOPO vector (Invitrogen). The dCas9 clone was verified by Sanger sequencing and subcloned into the destination vector pK2GW7 by the Gateway-LR reaction using LR Clonase II (Invitrogen). Expression of dCas9 was driven by a constitutive, tandemly duplicated, cauliflower mosaic virus 35S promoter.

To generate chimeras with transcriptional activators and repressors, in-frame fusions of EDLL, TAD, and SRDX domains with the dCas9 C-terminus were custom-synthesized by BlueHeronBio. To generate the dCas9:EDLL transcriptional activator, we custom-synthesized a fragment containing the dCas9 C-terminus fused to the EDLL domain flanked by *Mlu*I and *EcoR*I restriction sites (Supplementary Sequence 2). The EDLL restriction fragment was subcloned to dCas9/pK2GW7 digested with *Mlu*I and *EcoR*I. A similar strategy was used to generate the dCas9:SRDX transcriptional repressor using *Mlu*I and *EcoR*I (Supplementary Sequence 4). Due to the presence of an *EcoR*I restriction site within the TAD sequence, we custom-synthesized a fragment containing the dCas9 C-terminus fused to TAD flanked by *Mlu*I and *Xho*I restriction sites (Supplementary Sequence 3). Then, *Mlu*I and *Xho*I were used to directly exchange the EDLL domain in dCas9/pK2GW7 with TAD.

A chimeric RNA containing CRISPR RNA and *trans*-activating RNA was used in this study. Chimeric RNA cores were custom-synthesized and expressed under control of the *U6 Arabidopsis thaliana* RNA polymerase III promoter. All chimeric RNA cores containing the *U6* promoter, 20-nucleotide targets, and gRNA sequences were custom-synthesized (Supplementary Sequence 6-11). 20-nucleotide targets were designed for *Bs3* promoter and *PDS* endogenous gene (http://solgenomics.net/ Niben044Scf00014185). Chimeric RNAs containing specific targets were PCR amplified with the forward primer

carrying *BamH*I restriction site and the reverse primer carrying the *Xba*I recognition sequence (Supplementary Table 1). The PCR product integrity was confirmed by Sanger sequencing. These fragments were cloned into the multiple cloning site (MCS) of the pYL156 destination vector using restriction-ligation reactions and the *BamH*I and *Xba*I sites. The *Bs3* promoter was fused to *uidA* in the pKGWFS7 vector as described previously (Li *et al.*, 2013b).

4.7.2 Agroinfiltration

All dCas9 chimeric transcriptional regulators, U6:gRNAs and *Bs3*:*uidA* constructs were separately transformed into *A. tumefaciens* strain GV3101 by electroporation. *Agrobacteria* were separately cultured overnight in 5 mL LB medium containing gentamicin (50 μg/mL), specitnomycin (100 μg/mL) and rifampicin (50 μg/mL) antibiotics for bacteria containing the effector constructs, and in gentamicin (50 μg/mL), kanamycin (50 μg/mL) and rifampicin (50 μg/mL) antibiotics for bacteria containing the gRNA constructs. Subsequently, *Agrobacteria* were separately cultured in 20 mL LB, and grown to OD600 = 1.0-1.5. *Agrobacteria* were collected and re-suspended in infiltration buffer (10 mM $MgCl_2$, 5 mM MES, 0.1 mM acetosyringone). To study target binding and activation or repression, combinations of *Agrobacteria* containing effectors, transient targets, and gRNAs were grown separately and then mixed together, to equal OD_{600}=0.6, and co-delivered into 3-4-week-old *Nicotiana benthamiana* leaves by agroinfiltration.

4.7.3 Qualitative and quantitative GUS assays

Agroinfiltrated leaf discs were collected 36-48 hours post-infiltration, immersed in GUS staining buffer [10 mM NaH2PO4, 10 mM EDTA, 0.1% Triton-X100, 0.1% X-gluc, 1

mM K4Fe(CN)$_6$, 1 mM K3Fe(CN)6], and kept at 37° C for 24 hours. The following day, discs were de-stained with 70% ethanol. For quantitative assays, proteins were extracted from two leaf discs using 150 μL of GUS extraction buffer (50 mM NaH2PO4, 10 mM EDTA, 10 mM β-mercaptoethanol, 0.1% Triton-X100, 0.1% SDS). Protein concentrations in the sample extracts were quantified by Bradford assay. For the fluorometric assay, 90 μL of assay buffer [50 mM NaH2PO4, 10 mM EDTA, 10 mM β-mercaptoethanol, 0.1% Triton-X100, 0.1% SDS, 5 mM 4-methylumbelliferone β-D-galactopyranoside (MUG)] was mixed with 10 μL of sample and incubated at 37°C for 60 min. Reactions were stopped by adding 900 μL of 0.2 M sodium carbonate (pH 9.5). Fluorescence was measured in a Tecan microplate reader at 360 nm (excitation) and 465 nm (emission) with 4-methyl-umbelliferon (MU) dilutions as standard. In all experiments, background fluorescence (autofluorescence) was subtracted from the sample fluorescence measurements.

4.7.4 RNA isolation and real time quantitative PCR

Total RNA was isolated from two leaf discs using the RNeasy Plant Mini Kit (Qiagen) according to manufacturer's instructions. DNA digestion was performed on column during the RNA extraction with the use of RNase-Free DNase Set (Qiagen). 1 μg of total RNA was used for reverse transcription with SuperScript First-Strand Synthesis System for RT-qPCR (Invitrogen) to generate cDNA. qPCR was performed, as previously described (Bruggeman *et al.*, 2014), using the Power SYBR® Green PCR Master Mix (Invitrogen) with the following conditions: hold step at 95 °C for 10 min, denature step at 95 °C for 15 sec, anneal/extend step at 60 °C for 1 min. Primers used for RT-qPCR are listed in Supplementary Table 1.

CHAPTER 5

Understanding the Functional Roles of Rice Serine/Arginine-rich Proteins under Stress Conditions Using CRISPR/Cas9 System

5.1 Abstract

A serine/arginine-rich (SR) family of proteins are splicing factors and significant players in constitutive and alternative splicing of pre-mRNA. The interaction of SR proteins with RNA and other proteins underlines their importance in splice site recognition and assembly of the spliceosome. SR proteins are also known to play a role in plant growth and development in response to environmental cues, by controlling their own alternative splicing patterns. Alternative splicing regulates abiotic stress responses, and genes with regulatory functions are particularly prone to alternative splicing events. Recent studies indicate the link between the SR proteins and alternative splicing in response to abiotic stresses including abscisic acid (ABA).

In this study we set out to characterize the functional roles of all *Oryza sativa* ssp. *japonica* cv. *Nipponbare* SR proteins. We use the power of the CRISPR/Cas9 technology to generate a comprehensive rice SR knockout mutant collection harnessing the multiplexing feature of the CRISPR toolkit. Through the knockout mutant collection attempt to understand the roles of SR proteins in alternative splicing modulation and regulation in response to abiotic stress. We intend to identify downstream transcripts targeted by the SR proteins. Such understanding may help engineer plants for better stress tolerance when exposed to unfavorable environmental cues.

5.2 Introduction

Eukaryotic gene structure has been established with the discovery that genes comprise coding -exonic and noncoding -intronic regions. It has been shown that introns are excised from the primary RNA transcript and exons are stitched (spliced) together in the nucleus (Chow *et al.*, 1977). Later, it has been shown that the complexity of the proteome increases when the primary transcripts containing multiple introns and exons undergo splicing in various ways, generating few protein variants from a single gene. Also, splicing of the pre-mRNA seems to play a role in tagging the primary transcript with a destination message directing it to cytoplasm, nonsense-mediated mRNA decay or governing the localization in the cytoplasm (Le Hir *et al.*, 2000). Primary transcript splicing is regulated in spatiotemporal manner, depending on the cell type, developmental stage and environmental cues.

Alternative splicing (AS) varies from constitutive splicing (CS) in a way that, in CS only one type of splice sites are used to carry out splicing process in the primary RNA transcript. AS on the other hand uses different splice sites to produce two or more mature transcripts from pre-mRNA, resulting in greater proteome complexity. It is said that more than 95% of human genes (Pan *et al.*, 2008) and over 60% of *Arabidopsis thaliana* genes (Filichkin *et al.*, 2010; Zhang *et al.*, 2010b) undergo AS. Specific type of AS, namely intron retention is the most prevalent type present in plants, especially *Arabidopsis* (Ner-Gaon *et al.*, 2004). pre-mRNA splicing takes place in the nucleus where a large complex, called the spliceosome, processes the primary transcript. Many different components constitute the spliceosome including five different types of small nuclear ribonucleoproteins - snRNPs (U1, U2, U4, U5 and U6), whose role is to recognize the

splice sites and pre-mRNA branch point (Reddy, 2004). snRNPs are highly conserved across animal and plant kingdoms. snRNPs interact with many other proteins in the spliceosome whose major role is to provide assembly and stability of the complex and regulation of splicing process itself (Bessonov *et al.*, 2008).

Serine/arginine-rich SR proteins are known as major regulators that constitute the dynamic spliceosome complex. SR proteins were first discovered in *Drosophila melanogaster*, amphibian oocytes and later in humans. SR proteins were later identified in plants with the identification of SR proteins through antibodies against the conserved RS domain (Lazar *et al.*, 1995; Lopato *et al.*, 1996a) and complementation assays on S100 HeLa splicing deficient cell extracts (Lopato *et al.*, 1996a). Bioinformatics analysis confirmed that SR proteins are more abundant in plant species when compared to humans, other metazoans and unicellular eukaryotes. It has been reported that *Arabidopsis* has eighteen SR proteins encoded in its genome (Barta *et al.*, 2010), twenty four in *Oryza sativa* (Iida and Go, 2006), seventeen in *Brachypodium* (International Brachypodium, 2010). For comparison there are seven SR genes in *C. elegans* (Longman *et al.*, 2000) and twelve SR genes in humans (Manley and Krainer, 2010). Greater diversity of SR family in plants may be due to the different patterns in intron recognition between plants and metazoans possibly because plants have relatively shorter introns and are profuse in uracil (Duque, 2011). Moreover, during evolution plants have gone through genome duplications. For instance, twelve out of eighteen *Arabidopsis* genes encoding SR proteins are located on the duplicated parts of the genome and it is not known whether they evolved to play different roles or are just redundant (Reddy and Shad Ali, 2011).

In plants SR family of proteins is divided into six subfamilies. Three subfamilies are orthologs of human SR proteins and three subfamilies are specific to plant kingdom. SR subfamilies together with their structural features and assigned genes from *Arabidopsis* and rice are presented in Table 1.

Table 1. SR protein subfamilies, their structural features and *Arabidopsis* and rice assigned accessions. The SR subfamily is characterized by a conserved amino acid motif in their RRM domain (SWQDLKD) and SR domain rich in serine and arginine. The RSZ subfamily in addition to RRM and RS domains possesses a linking zinc knuckle. The SC subfamily is characterized by a RRM and SR domains. The SCL subfamily is characteristic of plant specific SR protein with RRM domain and N-terminal extension rich in such amino acids like arginine, proline, serine, glycine and tyrosine. Another plant specific RS2Z subfamily is characterized by RRM and SR domains, two linking zinc knuckles and an additional serine-proline rich acidic C-terminus domain. Finally the last plant specific subfamily, the RS subfamily, is characterized by two RRM domains (lacking the conserved SWQDLKD amino acid motif), followed by the RS domain. Table adapted from (Barta *et al.*, 2010).

A) Mammalian orthologs

Subfamily and Structure	Gene symbol	Accession
SR subfamily 	At-SR30	At1g09140
	At-SR34	At1g02840
	At-SR34a	At3g49430
	At-SR34b	At4g02430
	Os-SR32	Os03g22380
	Os-SR33a	Os05g30140
	Os-SR33	Os07g47630
	Os-SR40	Os01g21420
RSZ subfamily 	At-RSZ21	At1g23860
	At-RSZ22	At4g31580
	At-RSZ22a	At2g24590

	Os-RSZ21a	Os06g08840
	Os-RSZ21	Os02g54770
	Os-RSZ23	Os02g39720
SC subfamily 	At-SC35	At5g64200
	Os-SC34	Os08g37960
	Os-SC32	Os07g43050
	Os-SC25	Os03g27030

B) Plant specific

Subfamily and Structure	Gene symbol	Accession
SCL subfamily 	At-SCL28	At5g18810
	At-SCL30	At3g55460
	At-SCL30a	At3g13570
	At-SCL33	At1g55310
	Os-SCL25	Os07g43950
	Os-SCL26	Os03g25770
	Os-SCL30a	Os02g15310
	Os-SCL30	Os12g38430
	Os-SCL28	Os03g24890
	Os-SCL57	Os11g47830
RS2Z subfamily 	At- RS2Z32	At3g53500
	At- RS2Z33	At2g37340
	Os-RS2Z36	Os05g02880
	Os-RS2Z37	Os01g06290
	Os-RS2Z38	Os03g17710
	Os-RS2Z39	Os05g07000
RS subfamily 	At-RS31a	At2g46610
	At-RS31	At3g61860
	At-RS40	At4g25500

	At-RS41	At5g52040
	Os-RS29	Os04g02870
	Os-RS33	Os02g03040

All SR proteins across different species share common structural features. The N-terminus of the SR protein contains one or two RNA Recognition Motifs (RRM) that interact with RNA molecule through the recognition of short sequences such as the exonic splicing enhancers (ESEs) or silencers (ESSs). The C-terminus of the SR protein is rich in arginine and serine amino acids dipeptides and it is defined as the RS domain. The RS domain plays important role in interactions with other proteins in the spliceosome complex, and is involved in determining the subcellular localization of the SR protein, through interaction with the transportin-SR (Kataoka *et al.*, 1999). Phosphorylation status of SR proteins determines their subcellular localization and biological function. Apart from being the key component of the splicing machinery, SR proteins are known to play diverse molecular functions. For example, SR proteins are involved in RNA metabolism, transport of mRNA into cytoplasm (Huang *et al.*, 2003), mRNA translation and stimulation of protein synthesis (Bedard *et al.*, 2007; Sanford *et al.*, 2004) and promotion of nonsense-mediated mRNA decay (NMD) (Zhang and Krainer, 2004). Moreover, recent studies indicate that some SR proteins uphold genome stability (Li and Manley, 2005), take part in transcription elongation processes (Lin *et al.*, 2008) and assist in progression of cell cycle (Zhong *et al.*, 2009).

Current knowledge about SR proteins originates from limited collection of *Arabidopsis* overexpression lines including that of SR30 (Lopato *et al.*, 1999) and RS2Z33 (Kalyna *et al.*, 2003) and knockout line of SR45 (Ali *et al.*, 2007). According to the current

nomenclature, SR45 protein is no longer considered as a member of the SR family (Barta *et al.,* 2010), due to its unusual protein features - one RRM domain bordered by two RS domains (Golovkin and Reddy, 1999), despite the fact that it is a splicing factor and has been shown to reconstitute splicing in a splicing deficient background of cell extracts (Ali *et al.*, 2007). *Arabidopsis* possesses one SR45 and SR45a protein and rice has two SR45 like genes encoded in its genome.

Expression patterns of SR proteins vary in different tissues and under different developmental stages (Lazar *et al.*, 1995; Lopato *et al.*, 1996b) in *Arabidopsis*. Their subcellular localization is mostly restrained to the nucleus, however there is evidence that some SR proteins, including RSZ22, can shuttle between the nucleus and cytoplasm indicating the involvement in mRNA transport or translation (Rausin *et al.*, 2010). In addition, expression profiles of SR genes demonstrate that they undergo AS themselves and produce multiple transcripts per gene with the exception of RSZ22a and SCL28 that only have one mRNA transcript each (Palusa *et al.*, 2007). It remains to be determined whether their splicing isoforms are functional under stress conditions. Many of the alternative splice variants may encode shorter proteins that differ in function, or transcripts that contain a premature termination codon (PTC). On average 50% of such truncated transcripts go through NMD pathway (Palusa and Reddy, 2010).

Multiple transcripts per gene exhibit intron retention and perhaps some functional variants aid the plants to respond quickly to biotic and abiotic stresses. It remains to be investigated how intron retention isoforms are involved in stress responses. For example, it has been shown previously that genes responding to environmental cues or involved in regulatory functions are targeted by alternative splicing (Ali and Reddy, 2008). Some of

the examples include: the *waxy* gene in rice (Larkin and Park, 1999) and gene encoding the HSFA2 factor in *Arabidopsis* (Sugio *et al.*, 2009) alternatively spliced upon heat stress; *AOX* gene in rice alternatively spliced upon salt stress (Kong *et al.*, 2003); and probably most prominent example is that of DREB2 type transcription factor in *Arabidopsis* alternatively spliced during cold and drought stress.

Since SR proteins are involved in and regulate AS events it is highly probable that they have direct influence on how plants respond to environmental cues, albeit no extensive studies have been undertaken to examine this notion, specifically in crop plants. One study reports that there are no dramatic changes in overall transcript abundance of SR proteins when *Arabidopsis* plants were exposed to stress (Palusa *et al.*, 2007). The only notable change was that of SCL33 transcript, which was significantly repressed under salt, temperature and abscisic acid (ABA) stress (Palusa *et al.*, 2007). However, exposure to temperature, salinity and high light irradiation, has influenced alternative splicing pattern of some of *Arabidopsis* SR genes (Filichkin *et al.*, 2010; Lykke-Andersen *et al.*, 2001; Palusa *et al.*, 2007; Tanabe *et al.*, 2007). Because functional analysis of these splice isoforms are not present, it is hard to speculate whether they have any direct roles in stress responses. Nonetheless, current hypothesis is that different SR isoforms possibly have downstream gene targets and affect their splicing, hence regulate stress responses indirectly. To date only one SR gene, namely SR30, was reported to generate higher levels of transcript encoding full length protein under heat and salt stress (Filichkin *et al.*, 2010). Recently, a study that investigates the influence of stress phytohormone ABA on SR gene expression has identified several candidates of SR genes that are involved in ABA-mediated stress responses, including SR34, SR34b, SCL30a, SCL28, SCL33, RS40

and two SR-like genes: SR45 and SR45a (Cruz *et al.*, 2014). ABA is a plant hormone that plays important role in plant physiological and developmental stages. ABA mediated signaling is important in modulating stress responses to environmental cues and plant pathogens. ABA levels are known to rise in response to drought, cold, heat and high salinity. Palusa *et al.* reported that addition of exogenous ABA changes the alternative splicing pattern of three SR genes, namely SR34, SR34b and SCL33 (Palusa *et al.*, 2007) and Chen *et al.* identified ABA and salt sensitive phenotypes in RS40 and RS41 knockout mutants (Chen *et al.*, 2013b). To add, Carvalho *et al.* examined glucose signaling in *Arabidopsis* seedlings in SR45 mutant background and described that SR45 protein negatively regulates glucose signaling through downregulation of ABA pathway (Carvalho *et al.*, 2010). This experimental evidence implies that SR proteins play a role in some processes controlled by hormones, particularly ABA through involvement with key elements of the ABA signal transduction machinery.

To understand how SR proteins regulate stress responses in plants it would be highly important to pinpoint their downstream targets. Regulation of self splicing has been shown through *Arabidopsis* overexpression lines of *sr30* gene (Lopato *et al.*, 1999) and *rs2z33* (Kalyna *et al.*, 2003) and *rs2z36* and *sr33* in rice (Isshiki *et al.*, 2006). Another example includes SR45 protein that has been reported to regulate splicing of several other SR genes and in *sr45* mutant direct link has been shown between knocking out *sr45* gene and upregulation of flowering factor and ABA-related genes in *Arabidopsis* (Ali *et al.*, 2007; Carvalho *et al.*, 2010). Yet, full elucidation and wide-ranging analysis of splicing targets of SR proteins and gene expression profiles of these and other genes under abiotic stress maintain to be determined.

5.3 Research Aim and Objectives

The aim of this study is to understand the relationship between alternative splicing events regulated by SR proteins and responses to abiotic stress in rice plants. Main objectives of this study include: generation of comprehensive rice SR knockout mutant resource using CRISPR/Cas9 system, screening of rice SR mutant phenotypes resistant or sensitive to abiotic stress including salt and ABA, identification of gene expression profiles and alternative splicing events in identified candidates.

5.4 Results

5.4.1 Design and construction of CRISPR/Cas9 reagents targeting rice SR genes

To functionally knockout each gene in the SR family we designed and constructed specific single guide gRNAs (sgRNAs) targeting 20 nt sequence in each SR gene in rice genome. Harnessing the multiplexing feature of CRISPR system, we also designed constructs that targeted many SR genes simultaneously. To deliver single or multiple targets we used the polycistronic tRNA-gRNA (PTG) system (Xie *et al.*, 2015). This system is comprised of tandem motifs of tRNA followed by sgRNA, expressed from *U3* promoter. Internal tRNA processing RNases (RNase P and RNase Z) cleave the primary transcript at the 5' leader and 3' trailer recognition sites in the tRNA molecule, resulting in excision of separate, individual sgRNAs (Figure 36).

We custom synthesized each individual PTG construct with *U3::tRNA-sgRNA* scaffold to target each individual SR gene, and *U3::tRNA-sgRNA-[tRNA-sgRNA]*$_n$*-tRNA-sgRNA* scaffold to target multiple SR genes simultaneously (Table 2). We added *Bsa*I restriction enzyme recognition sites flanking each PTG construct. Next, we cloned our PTG

constructs into a destination vector pRGEB32 carrying Cas9 gene, using the *Bsa*I restriction enzyme.

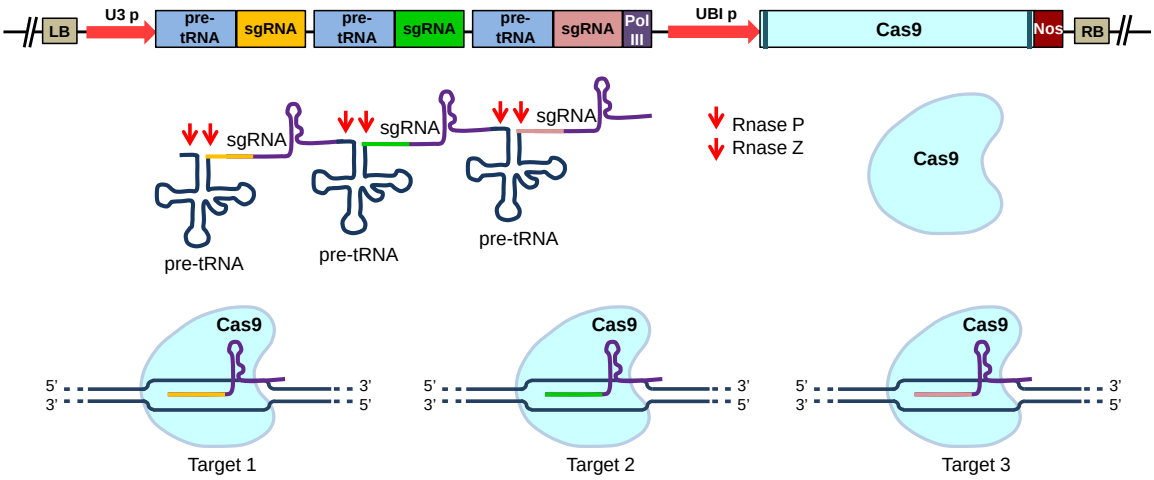


Figure 36. The PTG system for multiplex genome editing of rice SR genes with CRISPR/Cas9 technology. Polycistronic tRNA-gRNA (PTG) is expressed from the U3 promoter and comprises tandem t-RNA and sgRNA components. Transcript product is cleaved by endogenous enzymes RNase P and RNase Z separating mature sgRNAs from tRNAs. Cas9 expressed from *UBI* promoter is subsequently directed to single or multiple target sites by mature sgRNAs through RNA-DNA complementation.

Table 2. SR gene targets.

Construct name	sgRNA sequence	Gene symbol
PTG 1	AAGATCCCTCCGAGGCCACC	Os-SR32
PTG 2	AAAATTCCTCCAAGGCCTCC	Os-SR33a
PTG 3	GGAACCTCCCGGGCGACATC	Os-SR33
PTG 4	CATAAGGATATAGATCTTGC	Os-SR40
PTG 5	CGTGACGGCGCGGGAGCTCG	Os-RSZ21a
PTG 6	GGATCCCCGAGTGACTTCCG	Os-RSZ21
PTG 7	GAACCTGGATCCGCGCGTGA	Os-RSZ23
PTG 8	CCTCGGCCGCTACGGCCCAC	Os-SC25

PTG 9	ATGTCGCGCTTCGGCCGCTC	Os-SC32
PTG 10	CTTCGGGAGGTCGGGGCCGC	Os-SC34
PTG 11	GAGGACCTTCGTCGGCCATT	Os-SCL25
PTG 12	AGTCCTCGTGGGCGCTATGG	Os-SCL26
PTG 13	CCGGGTACCGGAGCCGGAGC	Os-SCL28
PTG 14	AGTCCCCCTAGGAGGGGATA	Os-SCL30/Os-SCL30a
PTG 15	GGAGCCGCAGCCCCAGCAAG	Os-SCL57
PTG 16	TGTTTTAACTGTGGGATTGA	Os-RS2Z36/Os-RS2Z38
PTG 17	TTCCTCGCGCACCCGGACCC	Os-RS2Z37
PTG 18	CCCGGGATGCAAATGATGCG	Os-RS2Z39
PTG 19	GCTATCCTTTTGGCCCTGGG	Os-RS29
PTG 20	GAGCGCCTCTTCAGCAAATA	Os-RS33
PTG 21	CTCCGCCTCGCGCTCCTCCT	Os-SR45-1
PTG 22	TCCGTCTCCGTCTCCGGCGA	Os-SR45-2
PTG 23	AAGATCCCTCCGAGGCCACC AAAATTCCTCCAAGGCCTCC GGAACCTCCCGGGCGACATC CATAAGGATATAGATCTTGC	Os-SR32 Os-SR33a Os-SR33 Os-SR40
PTG 24	CGTGACGGCGCGGGAGCTCG GGATCCCCGAGTGACTTCCG GAACCTGGATCCGCGCGTGA	Os-RSZ21a Os-RSZ21 Os-RSZ23
PTG 25	CCTCGGCCGCTACGGCCCAC ATGTCGCGCTTCGGCCGCTC CTTCGGGAGGTCGGGGCCGC	Os-SC25 Os-SC32 Os-SC34
PTG 26	GAGGACCTTCGTCGGCCATT AGTCCTCGTGGGCGCTATGG CCGGGTACCGGAGCCGGAGC	Os-SCL25 Os-SCL26 Os-SCL28

PTG 27	AGTCCCCCTAGGAGGGGATA GGAGCCGCAGCCCCAGCAAG	Os-SCL30a/Os-SCL30 Os-SCL57
PTG 28	TGTTTTAACTGTGGGATTGA TTCCTCGCGCACCCGGACCC CCCGGGATGCAAATGATGCG	Os-RS2Z36/Os-RS2Z38 Os-RS2Z37 Os-RS2Z39
PTG 29	GCTATCCTTTTGGCCCTGGG GAGCGCCTCTTCAGCAAATA	Os-RS29 Os-RS33
PTG 30	CTCCGCCTCGCGCTCCTCCT TCCGTCTCCGTCTCCGGCGA	Os-SR45-1 Os-SR45-2

5.4.2 Confirmation of Cas9 induced mutation and mutation types at sgRNA target sites

Synthetic Cas9-sgRNA constructs targeting SR genes were delivered into rice using the *Agrobacterium*-mediated transformation of calli induced from mature seed (Hiei and Komari, 2008). Transformed calli were selected on hygromycin containing media and transgenic rice plantlets were regenerated. Three weeks after rice plants were established in soil, leaf blade samples were collected for DNA extraction and confirmation of mutation at the target site (Figure 37).

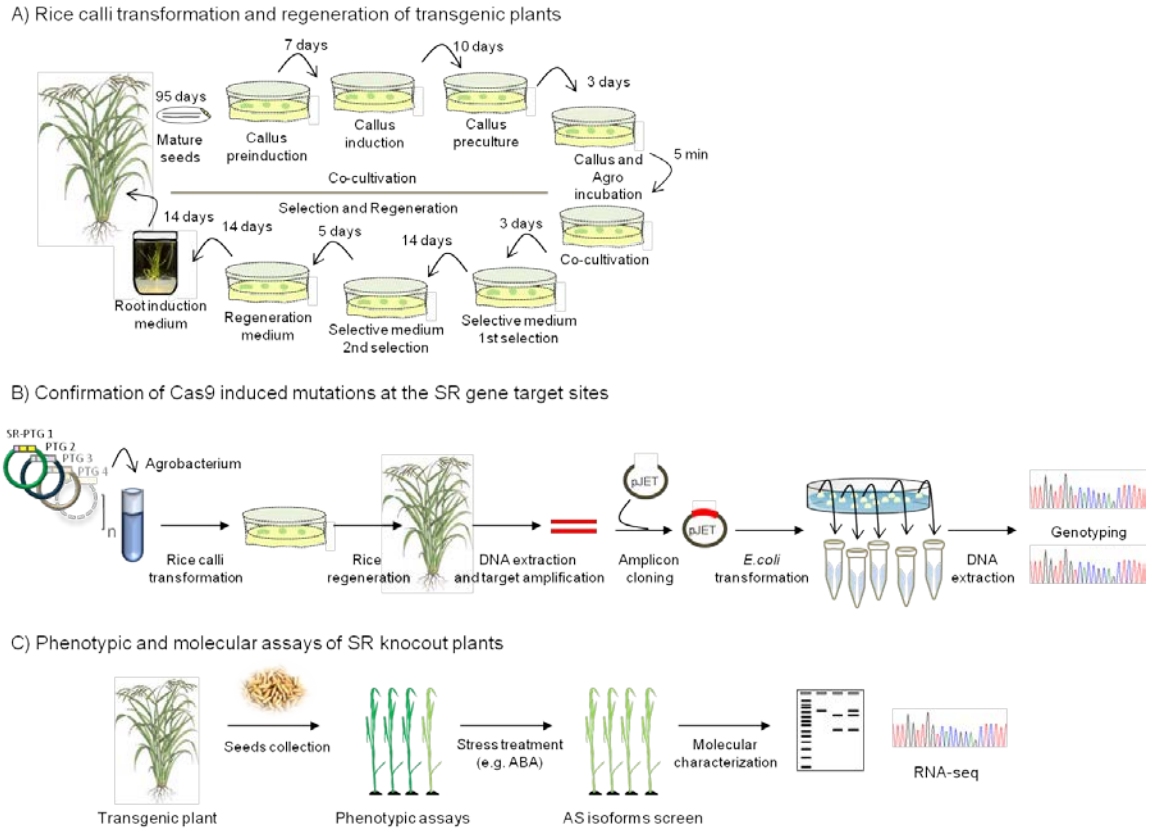


Figure 37. Schematic representation of rice transformation and experimental setup.A) Schematic representation of rice calli transformation using *Agrobacterium* carrying designed constructs targeting SR genes. B) Methodology for confirmation of various types of mutations induced by Cas9 at the SR genes target sites. C) Representation of phenotypic and molecular assays for functional analysis of SR proteins.

Plants were first screened for the presence of the transgene through PCR with sgRNA specific primers (Supplementary Table 5). Following transgene confirmation SR target site was amplified using target specific primers. Nearly all regenerated plants tested positive for the transgene presence, marking the transformation efficiency at about 99%. InDel mutations at the target sites were confirmed by Sanger sequencing (Table 3). A total of 126 plant samples were harvested and subsequently a total of 210 target sites were sequenced to check for mutation type. Cumulatively, 114 plants exhibited various types of InDel mutations on target, and 12 plants [4 plants for Os-SR32 (PTG1) and 8

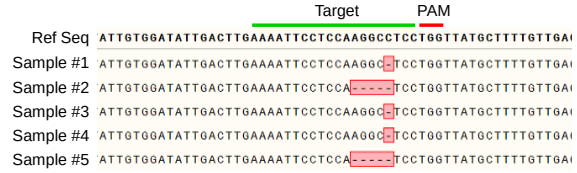
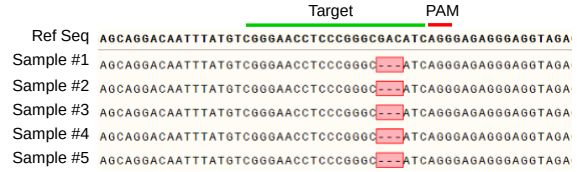
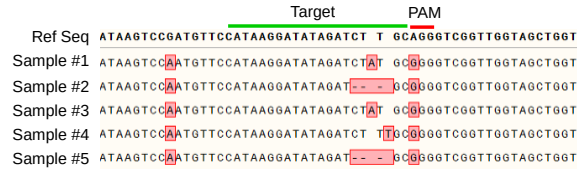
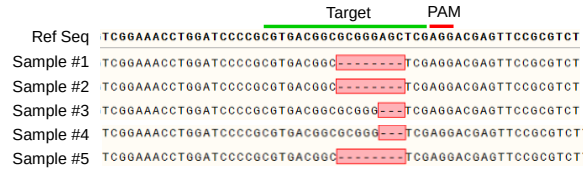
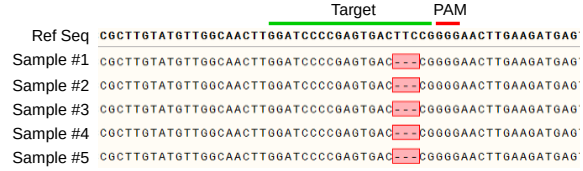
plants for Os-SR45-1 (PTG21)], showed no mutation. However, a mutation for Os-SR32 (which is also a part of the multiplexing construct) was observed in five different plants carrying the PTG23 construct that targets multiple genes simultaneously from the SR subfamily (Table 2 and 3).

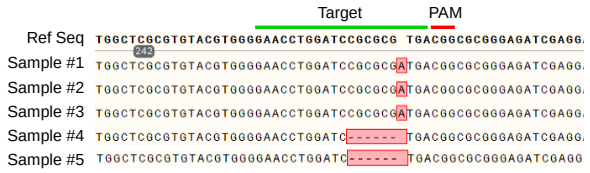
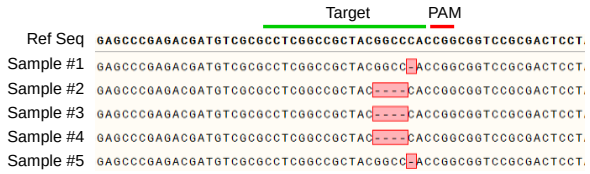
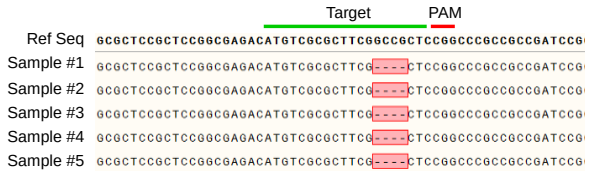
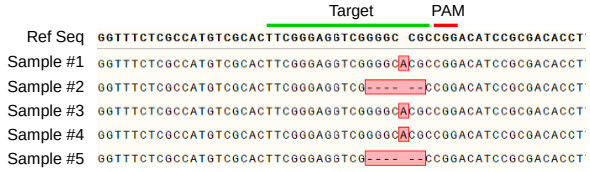
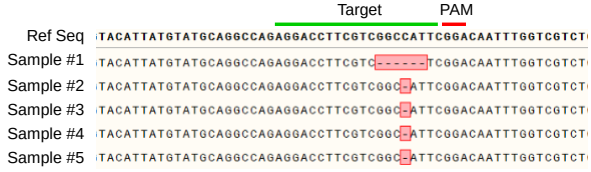
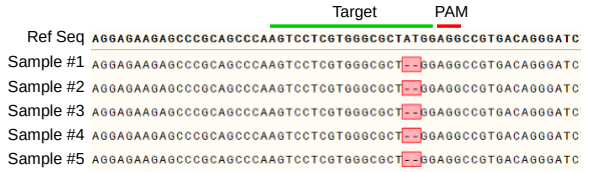
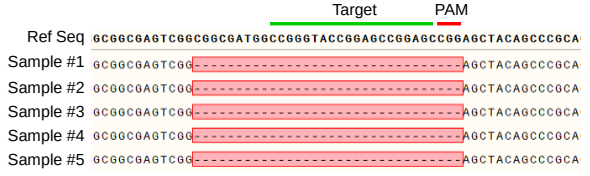
Out of 210 sequenced targets 26.2% showed no mutation, 17.6% were monoallelic, 56.2% were biallelic. Out of 56.2% of biallelic mutations, 41.4% were biallelic heterozygous and 14.8% were biallelic homozygous. On average the population of plant's cells carry either majority of wild type (unmodified) alleles or monoallelic mutations, yet our results indicate that the percentile of biallelic mutations is more than the half of all the samples genotyped, making the Cas9 targeting efficiency very high in rice. Moreover, the homozygosity achieved in the first generation of regenerated rice plants makes this crop an excellent model to study gene function through generation of functional knockouts.

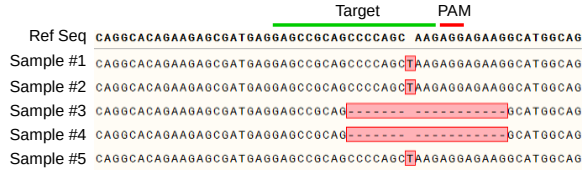
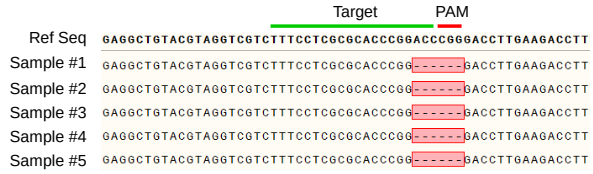
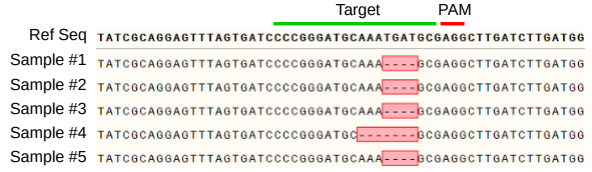
Table 3 reports all observed mutation types per target site and shows one example of the mutation (amongst others) observed within the plant. For example, four different plants with Os-SR32a (PTG2) target site showed different types of biallelic mutations, both heterozygous and homozygous (reported in Table 3), but only one type of mutation is presented in Table 3 as an example. The predominant indel mutations constitute deletions of usually up to ten nucleotides, however deletions as big as 31, 45 and 54 nucleotides have been observed. Insertions occur less frequently and on average are 1nucleotide long, however a 49 nucleotide insertion has also been observed. In addition, most of the mutations arise on the target sequence, three base pairs upstream of PAM sequence, where the Cas9 nuclease is known to cut double-stranded DNA. On occasion mutations

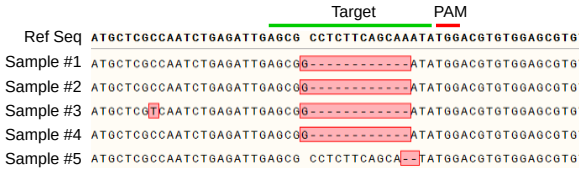
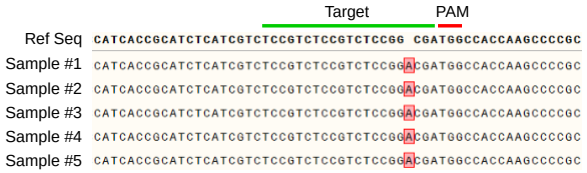
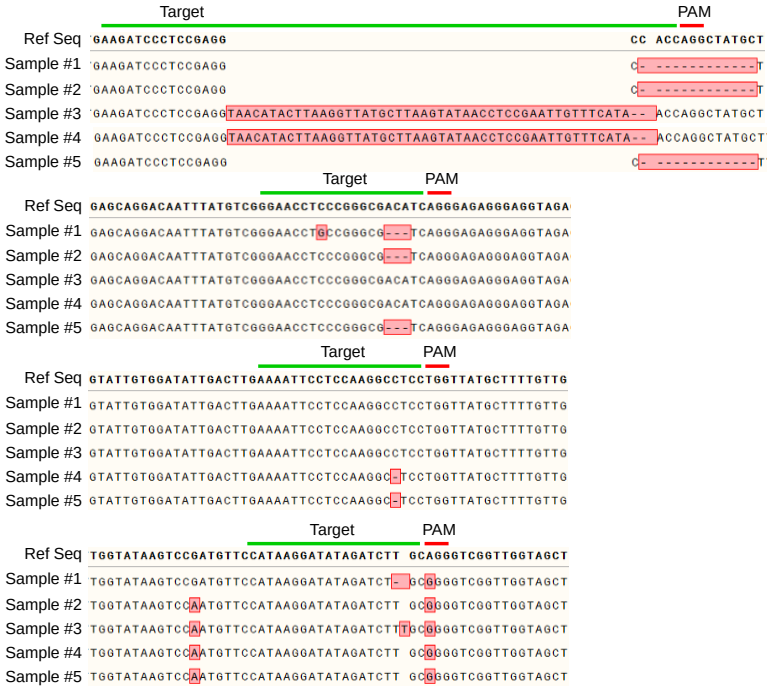
span PAM sequence, especially deletions longer than eight nucleotides, and in few examples mutations were noted to happen outside the target sequence. Different types of mutations depend on the efficiency and robustness of the NHEJ repair machinery.

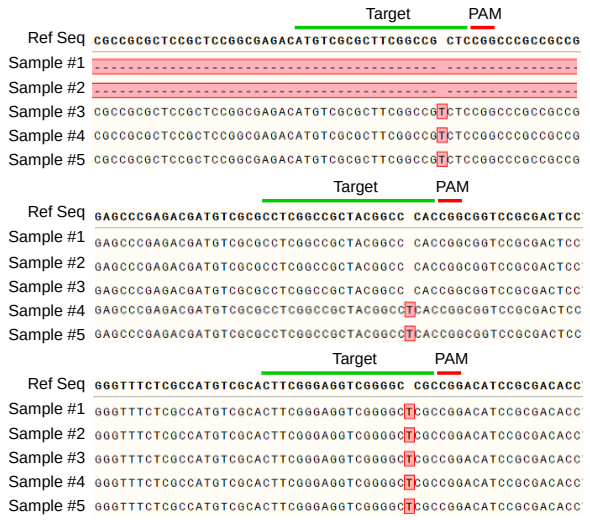
Table 3. Cas9 induced mutation types at SR target sites.

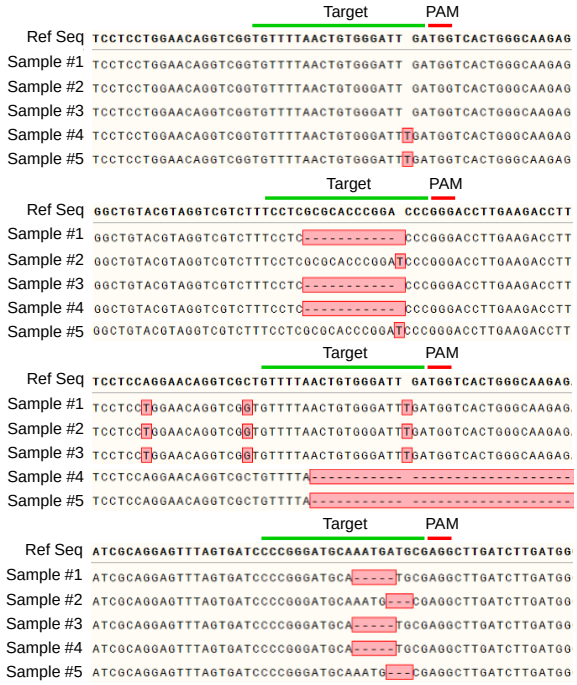
Target gene	Observed mutation types	Mutation example
Os-SR32 (PTG1)	No mutation	N/A
Os-SR32a (PTG2)	Biallelic-heterozygous Biallelic-homozygous	
Os-SR33 (PTG3)	Biallelic-heterozygous Biallelic-homozygous	
Os-SR40 (PTG4)	Biallelic-heterozygous	
Os-RSZ21a (PTG5)	Biallelic-heterozygous	
Os-RSZ21 (PTG6)	Biallelic-heterozygous Biallelic-	

	homozygous	
Os-RSZ23 (PTG7)	Biallelic-heterozygous	
Os-SC25 (PTG8)	Biallelic-heterozygous Biallelic-homozygous	
Os-SC32 (PTG9)	Biallelic-homozygous	
Os-SC34 (PTG10)	Biallelic-heterozygous	
Os-SCL25 (PTG11)	Biallelic-heterozygous Biallelic-homozygous	
Os-SCL26 (PTG12)	Biallelic-heterozygous Biallelic-homozygous	
Os-SCL28 (PTG13)	Biallelic-heterozygous Biallelic-homozygous	

Os-SCL30/ Os-SCL30a (PTG14)	Biallelic-heterozygous Biallelic-homozygous Monoallelic	
Os-SCL57 (PTG15)	Biallelic-heterozygous Biallelic-homozygous	
Os-RS2Z36/ Os-RS2Z38 (PTG16)	Biallelic-heterozygous Biallelic-homozygous Monoallelic	
Os-RS2Z37 (PTG17)	Biallelic-heterozygous Biallelic-homozygous	
Os-RS2Z39 (PTG18)	Biallelic-heterozygous	
Os-RS29 (PTG19)	Biallelic-heterozygous	

Os-RS33 (PTG20)	Biallelic-heterozygous Biallelic-homozygous	
Os-SR45-1 (PTG21)	No mutation	N/A
Os-SR45-2 (PTG22)	Biallelic-heterozygous Biallelic-homozygous	
Os-SR32/ Os-SR33a/ Os-SR33/ Os-SR40 (PTG23)	Biallelic-heterozygous Monoallelic	

Os-RSZ21/ Os-RSZ21a/ Os-RSZ23 (PTG24)	Biallelic- heterozygous Monoallelic	
Os-SC25/ Os-SC32/ Os-SC34 (PTG25)	Biallelic- heterozygous Biallelic- homozygous Monoallelic	
Os-SCL25/ Os-SCL26/ Os-SCL28 (PTG26)	Biallelic- heterozygous Biallelic- homozygous Monoallelic	

Os-SCL30/ Os-SCL30a/ Os-SCL57 (PTG27)	Biallelic- heterozygous Monoallelic	
Os-RS2Z36/ Os-RS2Z37/ Os-RS2Z38/ Os-RS2Z39 (PTG28)	Biallelic- heterozygous Monoallelic	
Os-RS29/ Os-RS33/ (PTG29)	Biallelic- heterozygous Monoallelic	

Os-SR45-1/ Os-SR45-2 (PTG30)	Biallelic- heterozygous Biallelic- homozygous Monoallelic	

5.5 Significance and Perspectives

Alternative splicing greatly increases the transcriptome and proteome complexity in plants and metazoans. Particularly, in plants, alternative splicing events are frequently changing under biotic or abiotic stresses. Few studies suggest that plant SR proteins play crucial roles in generating various splice isoforms in response to stress. Our SR knockout resource in rice will provide a great means by which the function of each SR protein can be elucidated and possibly identified to play an important role in stress responses. Moreover, many other biological questions can now be addressed by studying our rice mutant resource. For instance we would be able to investigate how SR genes are regulated in response to stress, what mRNA isoforms have their functional pools. Are any SR isoforms directed to NMD pathway, what is their localization in the nucleus, and what are the downstream targets of different SR splice isoforms? Moreover, is there any nucleo-cytoplasmic movement of SR proteins, other than reported by (Rausin *et al.*, 2010), indicating their involvement in other steps of gene expression processes.

5.6 Materials and Methods

5.6.1 Plant material and seed sterilization

Oryza sativa spp. *japonica* cv. *Nipponbare* plants were used to generate the SR knockout mutant resource. Seeds were surface sterilized in 96% EtOH for 1 min, then in 4% sodium hypochlorite with two drops of Tween-20 for 1 h on rotation and washed with sterile water. 100 seeds were used to generate calli for each separate transformation event.

5.6.2 Callus induction and preculture

Twenty five rice seeds were placed on 2N6 media (Hiei and Komari, 2008) and kept in dark at 30°C for seven days for callus preinduction. Next, scutella was cut from seeds, placed on 2N6 media and kept in dark at 30°C for ten days for callus induction (25 per plate). After, globular and yellowish in color callus was cut into small pieces (0.5 to 1mm in diameter) and cultured for another three days in continuous illumination at 32°C in 2N6 media. 110 calli pieces were placed in a single plate.

5.6.3 *Agrobacterium* inoculum

Different Cas9-sgRNA constructs targeting SR genes were separately transformed into *Agrobacterium tumefaciens* strain EHA105 by electroporation. The bacterial cells were separately cultured on AB media plates containing kanamycin (50 μg/mL) and chloramphenicol (25 μg/mL), at 28°C three days prior to rice calli transformation. Subsequently, bacterial cells were resuspended in AAM media (Hiei and Komari, 2008) directly from the plate and the OD_{600} was adjusted to 0.33. *Agrobacteria* were cultured in AAM media for 30 min prior to calli transformation.

5.6.4 Rice transformation and regeneration

About 650 calli pieces were placed in sterile water and then in *Agrobacterium* suspension, carrying desired Cas9-sgRNA construct, for two minutes for transformation step. Next, calli were drained on filter paper and cultured on 2N6-As medium (Hiei and Komari, 2008) in dark at 25°C for three days for co-cultivation.

Transgenic calli were selected on first selective 2NBKCH40 medium containing hygromycin (40 μg/mL) and ticarcillin/clavulanic (200 μg/mL) antibiotic for two weeks under continuous illumination at 32°C. Next, proliferated calli were transferred to second selective nN6CH50 medium also containing hygromycin (50 μg/mL) and ticarcillin/clavulanic (200 μg/mL) antibiotic for five days under continuous illumination at 32°C. Subsequently, resistant calli, yellowish in color, were transferred to N6RH50 regeneration media containing hygromycin (50 μg/mL) and ticarcillin/clavulanic (100 μg/mL) and incubated at 30°C under continuous illumination until the emergence of shoots. Then shoots were transferred to N6FH50 rooting media and cultured in magenta vessel at 30°C under continuous illumination until roots were long enough to establish the seedling in soil. Plants were grown to maturity in greenhouse conditions at 30°C, 12h light and 12h dark. Plant selection and regeneration protocol was followed from (Hiei and Komari, 2008) with slight modifications.

CHAPTER 6

Targeted transcriptional regulation and editing.

Concluding Remarks, Challenges and Future Directions

Piatek, A., and Mahfouz, M.M. (2016) Targeted Genome Regulation via Synthetic Programmable Transcriptional Regulators. *Critical reviews in biotechnology*. doi: 10.3109/07388551.2016.1165180.

TALEs and CRISPR/Cas9 are promising tools for the targeted editing and regulation of endogenous genes in their native context. The CRISPR/Cas9 system is of particular interest as it overcomes many of the pitfalls associated with the ZF- and TALE-based technologies.

Efficiency and specificity is paramount for any genome editing or genome regulation technology. CRISPR offers similar high levels of efficiency to TALE, and its design and implementation are simpler. However, there are concerns regarding the specificity of the CRISPR system. Mismatches between the DNA target sequence and RNA molecule are tolerated, which increases the potential for off-target effects in both Cas9 and dCas9 systems. These limitations require further exploration; however, as the coordinated effects of multiple TFs are needed to produce significant changes in expression, off-target effects are less problematic for targeted genome regulation than for genome editing. Furthermore, any off-target effects would be largely limited to regulatory portions of the genome.

Specificity of the CRISPR/dCas9 system can be examined by predicting the genomic off-target sites based on the nucleotide similarity to the intended target with mismatches allowed. Also, specificity of the CRISPR/dCas9 system can be examined directly by whole transcriptome sequencing. One study compared Cas9 activity and dCas9-KRAB repressor activity when the same sgRNA was used that had mismatches to the DNA target site. Findings suggested that activity of dCas9-KRAB was more sensitive to mismatches than Cas9 mediated cleavage activity (Gilbert *et al.*, 2014). Several approaches were proposed with the aim of increasing CRISPR specificity. For example, higher GC content in the target region might increase binding strength, which in turn

might strengthen stabilization of the complex and reduce the likelihood of binding at non-target sites (Fu *et al.*, 2013; Pattanayak *et al.*, 2013). The effect of novel sgRNA architectures on off-target effects was also examined. A 5′-truncated sgRNA and sgRNA with a truncated 3′ end and with two guanine nucleotides added before the complementarity region at the 5′ end both yielded improved on-target effects (Cho *et al.*, 2014; Fu *et al.*, 2014). Mismatches and sgRNA architecture should therefore be considered when designing sgRNAs to efficiently regulate gene expression at single and multiple targets.

The dCas9 protein offers a novel platform for the recruitment of effector domains, thus CRISPR/dCas9-mediated activation or repression represent promising synthetic transcription factor opportunities. However, CRISPR/dCas9 usefulness goes beyond transcriptional regulation applications. CRISPR/dCas9 was used to fluorescently tag endogenous proteins in order to visualize *in vivo* expression (Chen *et al.*, 2013a). In one example, EGFP-tagged endonuclease-deficient Cas9 protein was used to track elongation of repetitive telomere elements *in vivo* (Chen *et al.*, 2013a). This highlights the potential for using the CRISPR/dCas9 system to study chromosome dynamics in different developmental stages in various organisms and in response to environmental stimuli. Recently, CRISPR/dCas9 was repurposed for use in an optogenetic system for spatial and temporal control of gene transcription (Polstein and Gersbach, 2015). Light-activation provides another level of control to the CRISPR/dCas9 system, allowing fine temporal and reversible control. One potential use of such a light-controlled system is the activation or repression of key regulatory genes in a time- and space-controlled manner.

The CRISPR/dCas9 system could potentially be used in a broad range of applications (Figure 38). For example, in human cells, CRISPR is capable of binding and cutting methylated DNA (Hsu *et al.*, 2013) and it is therefore likely that dCas9 could be used for epigenome editing (Laufer and Singh, 2015). Fusions of dCas9 with histone acetyltransferase (dCas9-p300) were recently developed and shown to activate the promoter and proximal and distal enhancers of targeted genes (Hilton *et al.*, 2015). The activity of dCas9-p300 was higher than that of dCas9-VP64. This study suggested that the CRISPR/dCas9 system could be used to manipulate epigenetic marks and thereby facilitate comprehensive studies of the association between the epigenome and transcriptional control. Bound CRISPR proteins can also be used in chromatin co-purification experiments (Waldrip *et al.*, 2014), which further highlights the potential uses of CRISPR/dCas9 in understanding epigenetic regulation, chromatin-associated proteins, and chromatin dynamics.

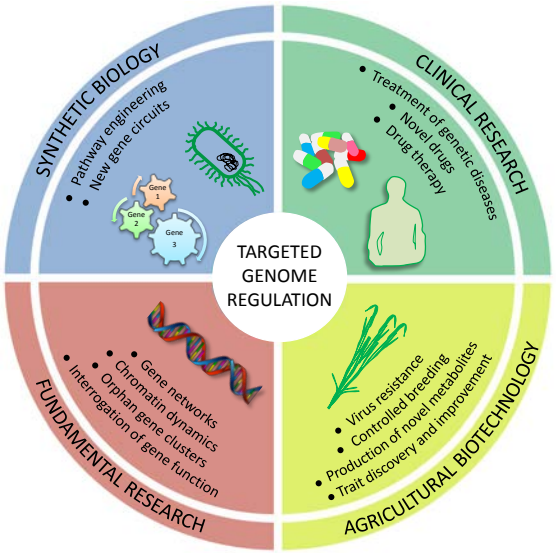


Figure 38. Applications for targeted genome regulation. Targeted genome regulation mediated by synthetic activators, repressors or chromatin modifiers have broad applications from addressing fundamental biological research questions to plant biotechnology and genomic medicine. Synthetic biology applications involve pathway engineering (e.g. for targeted production of metabolites, including biofuels) and manipulating genetic circuits in single cells, tissues and systems level. In clinical research targeted functional analysis of genes and genomes would expedite the discovery and development of novel drugs and understand the molecular bases of genetic diseases, that could lead to the development of personalized drugs and gene therapies. For agricultural biotechnology applications, targeted genome regulation could develop plants tolerant to biotic and abiotic stresses, thereby

increasing yield and prevent crop losses. In fundamental research reagents for targeted genome regulation would enable targeted functional analysis in native context at genomic and epigenomic levels, including chromatin dynamics under different physiological, developmental and stress cues.

The dCas9 protein has the potential to bind virtually any sequence on the genome, and screening using single sgRNA libraries targeting every promoter region in the genome will facilitate high-throughput functional genomics (Shalem *et al.*, 2015). This will facilitate the discovery of novel gene functions, gene redundancy, new metabolic pathways, and pathways of endogenous gene regulation. Such libraries were constructed for CRISPR activation and CRISPR repression genome-wide screens in mammalian cells (Gilbert *et al.*, 2014; Konermann *et al.*, 2015). Promoter architecture, cis regulatory elements, and distant enhancers might also be identified by using the CRISPR system. Determining interactions between different regulatory signals and spatiotemporal gene regulation will now be possible using not only the multiplexing feature of the CRISPR platform, but also the ability of the system to up- or down-regulate multiple gene expression within one cell at the same time (Zalatan *et al.*, 2015). Implementing multiplex strategies and orthogonal dCas9 proteins will increase the specificity of the CRISPR system and allow for precise manipulation of gene expression profiles and genetic circuits (Kiani *et al.*, 2014; Nissim *et al.*, 2014). Taken together, the amenable features of the CRISPR system suggest it will be the preferred genome and epigenome regulation tool for biomedical research and gene therapy.

CRISPR-based technologies also facilitate research in synthetic biology and plant biotechnology. Targeted gene regulation using CRISPR has particular potential for elucidating the gene expression networks underlying various plant environmental

responses which has important implications for agronomic trait development and crop improvement. Profiling of abiotic stress responses through identification of downstream processes and subsequent control by synthetic transcriptional regulators will lead to the generation of new synthetic genetic circuits and molecular pathways that can be utilized for the creation and use of novel allelic variants for breeding in crops.

The rapidly developing field of genome editing and regulation provides researchers with a wide variety of tools for gene discovery and functional analysis. These include older natural technologies such as gene overexpression and RNAi as well as newer synthetic targeted systems such as ZFs, TALEs and CRISPRs. The novelty and broad utility of the CRISPR/Cas9 system has gleaned much attention, and this technique is now routinely used to investigate biological processes in various cell types in a wide range of species. The speed at which CRISPR techniques are developing is remarkable, testifying to the power of this technology to revolutionize biotechnological and biological research. Recent developments of the CRISPR/dCas9 system will advance functional genomic studies, accelerate the development of therapeutic applications in genomic medicine, and facilitate plant breeding for enhanced agronomic traits.

APPENDICES

6.1 List of Publications

Authors	Title	Journal	Impact Factor	Notes
Piatek A, Mahfouz MM.	Targeted genome regulation *via* synthetic programmable transcriptional regulators.	Critical Reviews in Biotechnology	7.178 (2014)	In press (Part of Chapter 6)
Piatek A, Ali Z, Baazim H, Li L, Abulfaraj A, Al-Shareef S, Aouida M, Mahfouz MM.	RNA-guided transcriptional regulation *in planta via* synthetic dCas9-based transcription factors.	Plant Biotechnology	5.752 (2015)	Published (Chapter 4)
Li L, Atef A, Piatek A, Ali Z, Piatek M, Aouida M, Sharakuu A, Mahjoub A, Wang G, Khan S, Fedoroff NV, Zhu JK, Mahfouz MM.	Characterization and DNA-binding specificities of *Ralstonia* TAL-like effectors.	Molecular Plant	6.605 (2013)	Published (Chapter 2)
Li L, Piatek MJ, Atef A, Piatek A, Wibowo A, Fang X, Sabir JS, Zhu JK, Mahfouz MM.	Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification.	Plant Molecular Biology	3.518 (2012)	Published (Chapter 3)
Mahfouz MM, Piatek A, Stewart CN Jr.	Genome engineering *via* TALENs and CRISPR/Cas9 systems: challenges and perspectives.	Plant Biotechnology	5.752 (2014)	Published
Ali Z, Abul-faraj A, Li	Efficient virus-mediated	Molecular Plant	6.337	Published

L, Ghosh N, Piatek M, Mahjoub A, Aouida M, Piatek A, Baltes NJ, Voytas DF, Dinesh-Kumar S, Mahfouz MM	genome editing in plants using the CRISPR/Cas9 system		(2015)	
Aouida M, Li L, Mahjoub A, Alshareef S, Ali Z, Piatek A, Mahfouz MM	Transcription activator-like effector nucleases mediated metabolic engineering for enhanced fatty acids production in *Saccharomyces cerevisiae*	Journal of Bioscience and Bioengineering	1.884 (2015)	Published

6.2 Supplementary Information for Chapter 2

Name	Description
Supplementary Sequence 1	Codon Optimized RTLs.
Supplementary Sequence 2	Fragments for cloning of RTL backbones.
Supplementary Sequence 3	*Ralstonia* di-repeats sequences for ordered assembly.
Supplementary Sequence 4	DNA fragments used for the generation and cloning of the RTL-repeat concatemers in the *dHax3* backbone.
Supplementary Sequence 5	Minimal *Bs3* and *Bs4* promoter sequences.
Supplementary Figure 1	Assembly and cloning of *Ralstonia* RVDs to *Xamthomonas dHax3* backbone
Supplementary Table 1	List of primers used in this study

Supplementary Sequence 1. Codon Optimized RTLs.

YP00375049_codon_optimized:

ATGGCCGCTCTTGGCTATTCTAGGGAGCAAATTAGGAAACTTAAGCAGGAATCTTTATCTGGTGTCGCAAA
GTATCACGCTCCTCTAACCAGACATGGGTTCACACATACTGACATTTGTCGTATATCAAGACGTTGGCAAT
CCCTCCGAATGGTAGCCAAAAACTACCCTAAATTAATTGCAGCTTTACCTGATCTAACACGAACTCATATA
GTTGACATTGCTCGTCAGCGATCCGGAGATCTAGCTCTCGAAGCTCTATTACCTGTTGCTACGGCTCTCGC
CGCGGCGCCGCTTCGTCTCAGAGCATCCCAAATCGCAATCATAGCTCAGTGTGGAGAGAGGCCAGCTATAC
TCGCTCTGCATAGATTAAGGAGAAAGCTCACGGGTGCTCCACTTAACCTCACACCCCAACAAGTCGTCGCT
ATAGCAGCGAATACCGGCGGTAAGCAGGCACTTGGTGCGATCACCACTCAGTTGCCGATTTTAAGAGCTGC
ACCCTATGAGCTCTCTCCAGAGCAGGTGGTGGCCATTGCTAGCAATAATGGTGGAAAACCTGCTCTCGAAG
CGGTTAAGGCTCAACTTTTGGAGCTCAGAGCGGCTCCTTACGAACTGTCGCCGGAGCAAGTGGTCGCGATC
GCGAGCAATAATGGGGGAAAGCCAGCACTCGAAGCAGTTAAAGCTTTATTACTTGCCCTGAGAGCCGCACC
GTATGAACTAAGTACAGAACAGGTTGTGGCCATAGCGAGCAACAACGGAGGAAAGCCGGCGTTGGAAGCCG
TTAAAGCTTTACTTTTAGAATTGAGAGCAGCACCTTATGAACTGAGTACTGGCCAAGTTGTAGCAATAGCT
TCCAATGGAGGTGGCAGACAGGCTTTAGAAGCGGTTAGAGAGCAATTACTTGCACTAAGGGCTGTACCATA

CGAGCTTAGTACTGAACAGGTTGTGGTCATTGCCAATTCCATTGGGGGGAAACAGGCTTTGGAAGCTGTGA

AGGTGCAGTTACCGGTTTTACGTGCCGCCCCATACGAGCTGAACACAGAACAGGTTGTGGCTGTTGCAAGT

AACAAAGGTGGAAAGCAAGCTTTGGAGGCAGTTGGTGCACAGTTACTAGCATTGAGAGCAGTTCCATATGC

CCTCACAACAGCACAAGTTGTGGCGATTGCATCTAATGATGGGGGCAAACAGGCGTTGGAGGCGGTTGGTG

CTCAATTGTTAGTGTTGAGGGCCGTACCTTATGAATTGACTACTGCTCAAGTCGTAGCTATTGCTAGCAAC

GATGGTGGTAAGCAAACACTTGAAGTCGCGGGGGCTCAGTTATTGGCTCTTAGAGCTGTTCCATACGAACT

AAGTACTGAGCAGGTTGTTGCGATCGCGAGCAACAATGGGGGTAAGCAGGCGCTAGAAGCAGTTAAGACTC

AGCTCCTTGCTTTGCGTACAGCCCCTTATGAACTATCTACGGAACAAGTGGTAGCCATAGCTTCAAATAAT

GGCGGCAAACAAGCGCTCGAGGCCGTTAAGGCTCAGCTTCCTGCTCTCAGGGCCGCACCATATGAGTTGTC

TACTGAGCAGGTCGTTGCAATAGCAAGCAATAATGGTGGTAAGCAAGCACTCGAAGCTGTTAAGGCGCAAC

TTTTGGTCCTCAGGGCGGCTCCCTACGGTCTTAGTACGGCACAAGTTGTGGCCATCGCTGCTAATAATGGT

GGAAAGCAAGCTCTCGAGGCGGTACGTGCTTTGCTTCCTGTTCTCCGAGTTGCTCCCTATGAATTGAGCAC

GACCAGAGTGGTGTCAATCGCGTGCATCGGTGGAAGGCAGGCTTTGGAAGCTATAAAAACACATATGCCAG

CTTTAAGACAGGCACCATACTCTCTCAGCCACGAGCGAGTTGTGGCAATTGTGTGTAATGGAGGACGAAGT

GCCCTTGAAGCCGCACGACATGGACTCCCTGTTAGAGACATAAGAAGAATTCGTAATAGGAAGACGTCAGA

TGCAACTCTACCTGCCCCTGTATTGGGTCCAACGCCACAGGAATTGGTTGCCGTCCTGCACTTCTTTAGGA

CACATCAACAGCCAAGGCAAGCCTTCGTGGATGCACTTACTGCATTCCAGACAACCAGAAGGGCACTGCTT

AGGCTATTATCCAGTGCTGGAGTCACTGAAATTGAGGCCCTGGGTGGGATGATTCCTGATGCTGCCGAACG

TTGGCAGAGATTGTTAGGTCGTCTCGGGATCCGTCCAGCCACCGACACTGTGGTCACGTCACCTGATCCAA

TGCAGGGATTTGCACAATCCCTGGAGAGAAGCTTGATGTCATCTGGAACTACGGAGCAGTCAGCTAGCCCA

TCTCAACACAGATGGCCAGCGGGAACAGCAAAAACGCCAGAATCAGCCCGTCACAGACTAGACAACGCAGC

TCAACCTCCCATGCAATGGCCTGATCAACTCGTGTGGAGTCAAAGGAGGAGGTGCACCGCTAGCTCACAGA

CTCCTCCGCCTGGAGCCGCTAGCGTGCCGGCTAACCTTCAGTGGGGAGCACGAGCTCAGAGAACAAGTTCT

AGACTCCAGCCTGAACTGAGGCCTATGCCTGCGAGGGTTGTGCCAGCGTCAGCACAACCTAGTCCTTTAGG

AGACCTTGAGTTTGGACTCCCTGATCCTGGTACCCCTACAGCAGCTGATTTAGCCCTGGACCTTGACTGGC

TTCTCCAACTACTAGACCTC

NP5199361_codon_optimized:

ATGAGAATTGGTAAAAGCTCTGGGTGGCTGAATGAATCCGTCAGTCTTGAATACGAGCATGTTAGCCCTCC

TACTCGTCCTAGAGATACAAGACGTAGACCTCGAGCTGCGGGAGACGGAGGGCTAGCTCATTTACATAGGA

GGCTTGCTGTCGGATATGCCGAAGATACTCCGAGGACGGAAGCTCGTTCTCCTGCCCCAAGACGTCCTCTC

CCCGTGGCACCCGCGTCAGCGCCTCCTGCACCTTCTCTCGTTCCAGAGCCTCCTATGCCGGTATCCCTTCC

AGCTGTAAGTAGCCCACGTTTCTCTGCCGGATCGTCAGCTGCCATCACCGACCCTTTTCCTAGTTTGCCTC

CTACGCCGGTTCTTTATGCAATGGCTCGTGAGCTTGAGGCATTAAGCGACGCTACTTGGCAGCCTGCGGTG

CCATTGCCTGCAGAACCACCTACTGATGCACGAAGAGGAAACACAGTTTTCGATGAAGCTTCTGCTAGTTC

CCCGGTAATCGCTTCAGCTTGTCCGCAGGCTTTTGCATCACCTCCTCGTGCCCCTAGGTCGGCTCGTGCTC

GAAGAGCACGAACAGGTGGTGATGCCTGGCCCGCACCAACTTTCCTTTCACGTCCAAGTTCTAGCCGTATC

GGGCGAGACGTGTTCGGAAAGTTGGTTGCTCTCGGGTACTCGAGAGAGCAAATCAGAAAGCTTAAGCAAGA

ATCTCTCAGTGAAATAGCTAAGTATCACACCACCCTCACGGGACAAGGATTCACGCACGCAGATATCTGCC

GAATCAGCAGGCGACGACAGTCACTTAGAGTCGTGGCCCGAAATTACCCTGAATTGGCAGCTGCATTACCT

GAATTGACTAGGGCGCATATAGTTGATATTGCTAGACAGAGGTCTGGAGATTTGGCATTGCAAGCATTGCT

TCCAGTTGCTACCGCCTTGACAGCAGCCCCACTCAGATTGTCCGCTTCTCAAATCGCTACTGTGGCTCAGT

ATGGTGAAAGGCCTGCTATCCAAGCACTCTACAGATTACGTCGTAAACTTACCCGTGCTCCTCTCCACCTC

ACTCCCCAACAAGTGGTGGCCATTGCATCCAATACGGGTGGTAAAAGAGCTCTTGAAGCAGTATGTGTTCA

GCTGCCCGTGCTTCGTGCCGCCCCATATAGGCTTAGTACGGAACAGGTGGTAGCTATTGCATCAAATAAGG

GTGGAAAACAGGCACTTGAAGCAGTCAAAGCACATCTATTAGACCTCCTTGGGGCCCCGTATGTCCTCGAT

ACTGAACAAGTGGTCGCTATCGCTTCACACAATGGTGGGAAACAGGCTCTGGAAGCGGTCAAGGCTGACCT

TCTCGATCTTAGGGGAGCTCCGTATGCGCTAAGTACTGAGCAAGTCGTCGCAATAGCTAGTCATAACGGGG

GAAAGCAAGCACTTGAAGCTGTCAAAGCGGATCTCCTAGAGTTGCGTGGAGCTCCCTACGCCTTATCTACG

GAACAAGTCGTCGCCATTGCCTCGCACAATGGAGGAAAACAGGCTTTAGAAGCTGTTAAGGCTCATTTGTT

GGATCTAAGAGGTGTTCCCTATGCGCTCTCGACTGAACAAGTCGTGGCTATAGCTAGCCATAATGGTGGGA

AGCAGGCCCTTGAGGCGGTTAAAGCTCAGCTCCTCGATCTCAGAGGTGCTCCCTATGCACTGTCTACTGCC

CAAGTCGTAGCTATAGCCTCAAATGGAGGGGGGAAGCAGGCACTGGAGGGGATAGGAGAGCAACTTCTTAA

ATTGAGGACCGCCCCTTATGGGCTTAGCACTGAGCAAGTAGTTGCCATAGCTTCACACGATGGCGGTAAGC

AAGCACTCGAAGCAGTTGGTGCTCAGTTAGTTGCATTGAGGGCCGCTCCTTACGCTTTAAGTACTGAGCAG

GTTGTTGCGATCGCGTCGAATAAAGGAGGGAAACAGGCTTTAGAAGCGGTTAAGGCCCAACTTCTTGAATT

GAGAGGTGCACCTTACGCCCTCTCAACAGCACAAGTTGTTGCAATTGCGTCACATGATGGTGGTAATCAGG

CTTTGGAGGCAGTAGGTACTCAATTGGTGGCACTGCGAGCAGCTCCTTACGCATTGTCGACAGAACAAGTA

GTTGCCATTGCATCTCACGATGGAGGTAAGCAAGCGCTTGAAGCGGTTGGAGCACAGCTCGTCGCTTTGAG

GGCTGCGCCCTACGCCCTGAACACCGAACAAGTCGTGGCTATTGCTTCATCACATGGAGGCAAGCAGGCTC

TCGAGGCGGTTAGGGCGTTGTTTCCTGATCTAAGGGCTGCACCGTATGCCTTATCTACAGCTCAACTTGTT

GCTATCGCTTCGAACCCGGGAGGAAAACAAGCACTCGAGGCTGTAAGAGCTCTTTTTAGGGAACTCCGAGC

CGCTCCATATGCACTCTCGACGGAACAAGTTGTGGCAATAGCCTCCAACCATGGAGGAAAACAGGCACTAG

AAGCAGTCCGTGCACTGTTCCGTGGATTGAGAGCTGCTCCATATGGTCTTAGCACAGCACAAGTTGTCGCC

ATCGCAAGTAGTAACGGGGGAAAGCAAGCTCTCGAAGCCGTTTGGGCCCTTTTACCAGTCTTAAGGGCAAC

CCCTTATGACCTGAACACAGCTCAAATCGTTGCGATCGCAAGCCATGACGGAGGTAAACCGGCTTTAGAGG

CTGTATGGGCCAAACTTCCGGTCCTTAGGGGTGCTCCTTATGCCCTGTCGACAGCTCAGGTTGTCGCTATC

GCTTGTATTTCAGGGCAACAGGCACTTGAAGCTATCGAGGCACATATGCCTACTCTCAGACAGGCTTCTCA

TTCATTATCTCCTGAACGAGTTGCAGCAATCGCCTGCAATGGAGGACGAAGTGCTGTTGAAGCAGTAAGAC

AAGGACTCCCAGTCAAAGCTATTAGAAGAATTAGAAGAGAGAAAGCCCCGGTGGCTGGTCCTCCACCAGCA

TCACTTGGTCCTACCCCACAAGAGCTTGTGGCTGTGCTTCACTTTTTTCGAGCACACCAACAACCTAGACA

AGCATTTGTTGATGCATTGGCCGCATTTCAAGCAACGAGACCCGCTCTTCTTAGATTATTATCGTCAGTAG

GGGTTACCGAGATCGAGGCATTGGGAGGAACAATTCCTGACGCTACTGAGAGGTGGCAAAGGCTCCTCGGT

AGACTGGGGTTTAGGCCGGCCACAGGAGCGGCTGCTCCCAGCCCGGATAGTTTGCAGGGTTTCGCTCAGTC

GCTTGAACGTACTTTGGGTTCGCCTGGGATGGCAGGACAATCAGCTTGCTCCCCTCATAGGAAGAGGCCTG

CTGAAACCGCAATAGCTCCTAGAAGTATTCGACGTTCACCAAATAACGCGGGGCAACCTTCTGAGCCGTGG

CCTGATCAATTAGCTTGGCTACAGCGACGAAAACGAACGGCCCGATCTCACATAAGGGCTGATTCCGCTGC

GTCTGTGCCTGCAAATCTGCACTTAGGTACTAGAGCACAATTTACTCCAGACAGACTAAGGGCCGAACCAG

GGCCTATTATGCAGGCGCATACCTCTCCTGCTTCAGTCAGTTTCGGCTCTCATGTAGCTTTCGAACCTGGA

TTACCTGATCCTGGAACTCCTACTTCAGCGGATCTCGCTTCGTTCGAAGCCGAACCATTTGGTGTGGGTCC

GCTCGACTTTCACTTGGATTGGCTTCTCCAAATTCTTGAAACT

CAQ18687_codon_optimized:

ATGCGTAGAAGAACAGCGGTCGGTCGAGTTCCTGGTGCTTCGCGTTCCGGAACTTCACCCCTAGTGCTTTC

CCAACCGCTTAGCCGTGTGAGTGCCTCACCACAACCTGCACGTTCCTCGGCTATAAGTTCTCCGAATTTTT

CAGCTGGAAATCCAACCGCCTTCGCAAATCCTAGCCCGAGCTTGCCCCCTACTCCAGTTTTACCACCCACA

CCCGTCCTTTACGCAATGGCTAGAGAGTTGGAGGAACTTCATAATGCTACATGGCAGCCCGCTGTCCCCCT

GACAGCGGAACCATTGGCCGATGCTAGAAGGGACAATACTGTGGTTGATGGTGGAAGTGGATCTAGTCCTG

CCATCGCATCAGCTCGTCCGCAGGCATTCGCTGGACCGCCAAGAGCGCCTAGATCAGCCAGAGCTAGAAGA

GCTCGTACAGCGGGAGATGCTTGGCCAGCTCCGGCTTATTTGGGTCCTAGCCCTAGTCCGTCACCTTCGCC

TTCTCCATCGCCAAGAATCGCTCCGGACCTTAGTGGGAAACTTGCAGCATTGGGATACAGCAGGGAGCAGA

TTCGTAAGCTAAAACAAGAATCATTAGCTGAGGTGGCTAAATATCATGCAACGCTAGCTGGACAAGGATTT

ACTCATGCAGATATTTGCAGAATTTCTAGGCGTTGGCAGTCGCTACGTGTAGTTGCTAACAACTATCCTGA

ACTCATGGCAGCATTACCTAGGTTGACAACTGCTCAGATAGTTGATATTGCTAGGCAAAGGAGCGGTGACT

TGGCTTTACAGGCACTCCTGCCGGTGGCTGCGGCTCTCACTGCCGCACCTTTGGGGCTCTCTGCGTCTCAA

ATTGCTACTGTGGCTCAATACGGTGAGAGGCCAGCTATTCAAGCACTGTATAGACTGAGGAGGAAGCTGAC

TCGAGCACCGTTAGGGCTTACACCACAACAGGTCGTGGCTATAGCAAGCAACACAGGAGGGAAGCAGGCCC

TTGAAGCTGTTACGGTTCAACTTCGTGTGCTAAGAGGAGCACGTTATGGCTTATCGACCGAACAAGTGGTA

GCTGTTGCAAGCAACAAGGGAGGCAAGCAAGCTCTGGCGGCAGTTGAGGCTCAGTTACTTAGATTGAGGGC

TGCCCCGTATGAACTGTCTACAGAGCAAGTTGTGGCCGTAGCCAGTCATAACGGCGGAAAACAAGCTCTCG

AGGCTGTTCGTGCTCAACTCTTGGATTTGAGAGCTGCACCATACGAGCTTTCGACGGAGCAAGTGGTTGCG

ATTGCTTCACATAATGGTGGCAAACAGGCATTGGAGGCTGTGAGGGCTCAGCTTCCTGCGCTAAGAGCAGC

CCCTTACGGTCTCAGTACTGAGCAGGTTGTGGCTATCGCATCTCACAACGGTGGAAAGCAAGCCTTGGAGG

CTGTCAAAGCCCAACTCCCTGTTCTTAGGAGAGCTCCATACGGTTTGTCGACAGAACAAGTAGTCGCCATC

GCATCGAACGGGGGGGGAAAGCAAGCTTTAGAAGGAATCGGTAAACAATTACAAGAGCTCAGAGCTGCGCC

ACATGGGTTGAGCACAGCGCAAGTAGCAACTATCGCATCAAGTATCGGAGGAAGGCAGGCGCTGGAGGCAG

TTAAAGTCCAACTCCCTGTTCTAAGAGCTGCTCCATATGGCTTGACGATCGCACAGGTCGTTGCGGTTGCG

AGTCACAATGGTGGTAAACAAGCTCTCGAGGCTATTGGGGCGCAATTGCTGGCATTACGTGCAGCTCCTTA

CGCACTGTCAACAGCACAAGTGGTCGCCATAGCCTCGAACGATGGTGGAAAACAGGCATTAGAAGAAGTAG

AAGCGCAGCTCCTTGCACTTAGGGCTGCCCCATACGAATTGTCTAGTGCTCAGGTGATTGCGATAGCTGCA

GGTAATGATGGAGGAGGGAAGCAGGCATTGGAGGCTGTAAAAGCCCAGTTGCCTGTGCTTAGAGCTGCGCC

CTACGAGCTCAGTGTCGCTCAGGTGGTGACAATAGCTTCTCACAACGGGGGTAAGCAGGCGCTGGAAGCTG

TCCGAGCGCAGCTCCTAGCATTGCGTGCGGCTCCTTATGGTCTTTCGACCGCACAAGTCGTTGCTGTTGCT

GGCAGGAATGGCGGGAAACAGGCTCTCGAGGCTGTTCGAGCTCAGCTTCCAGCTCTCAGAGCAGCACCATA

TGGATTATCAATTGCCCAAGTTGTAGCTGTTGCTTCAAGAAGCGGTGGAAAACAGGCTCTTGAAGCGGTCA

GAGCCCAACTTCTTGCGCTACGTGCTGCTCCTTACGGGCTTAGTACAGCTCAAGTGGTGGCGGTTGCTTCG

GGCTCAGGAGGGAAGCCGGCGTTGGAAGCAGTTAGGGCTCAACTACTTGCACTTAGGGCTGCCCCATACGG
CTTAAGTACTGCGCAAGTTGTTGCTGTCGCTTCCCATGACGGTGGAAAACCAGCCCTGGAGGCTGTACGAA
AACAACTCCCGGTTCTTAGAGGAGTTCCGCATCAACTCAGCATTGCTCAAGTTATAGCTATAGCTTGTATC
GGTGGTAGGCAAGCTCTAACTGCTATTGAAATGCACATGCTTGCATTAAGGGCTGCTCCTTATAATCTTTC
GCCTGAGAGAGTCGTCGCTATCGTCTGTATCGGCGGCAGGTCCGCAGTTGAGGCCATCAGACATGGCCTTC
CTGTCAAGAGAATACAGAGAATCCGAAGACGAAAAGCTCCCGAAGCAACCCCGCCAGCAGGGCCTTTCGGG
CCTACACCACAAGAACTCGTGGCCGTCCTTCATTTCTTTAGAGCTCATAGACAACCTCGTCATGCATTTGT
TGATGCGCTGACAGAGTTCCAAATCACTAGAGCTGCACTCTTAAGACTGCTGTCAAGCGCTGGTGTGACTG
AGATTGAGGCTCTTGGAGGCACGATCCCAGATGCTGCTGAAAGATGGCAACGACTTCTCGGCCGACTTGGT
ACCAGACAGGCTATTGGAGTAGCTGCTCCCTCACCAGATTCGTTGCAAGGTTTCGCTCAGTCTCTTGAAAG
ATCTATGCTTTCGCCGGGGATAGCAGAACAATCAGCATCTCCACCCCCAAGAGAGAGGCCAGCTGAAACCG
CTATTGCACCCAGAAGTAATACCGGTAGGCCATCTGAGCCATGGTTGGACCAACCTGCTTGGTCTCAACAT
CGTAAGAGAACTGTGAGGTCTCGAATGAGAGCTGATGCAGTGGCAAACATGCCTGCCAATCCTCATCGTGA
CGCCAGAGCTCAGTTTACGCCTGGATGCCTCCCTCCTGCTCCTGGGCCCATGCCTGCCCACATCTCTCCAG
CCAGTGCTGGTTCTAGTTCTCACGTCGCTTCTGGTAGTGTTCTACCTGACCCTGGTACCCCTACTGGAAGC
GACCTTGCAATGCTCGAGGCGGAGTCATTTGGAGCCGCTGCTCTTGCCTTTGATTTTGATAGCTTTTTACA
GATGCTCGATGTT

Supplementary Sequence 2. Fragments for cloning of RTL backbones.

For RTL1-YP backbone:

YP N-terminus fragment (L1+F1-*Apa*I-*Nde*I-*Spe*I clone):

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTAT
AATGCCAACTTTGTACAAAAAAGCAGGCTCCATGGCCGCTCTTGGCTATTCTAGGGAGCAAATTAGGAAAC
TTAAGCAGGAATCTTTATCTGGTGTCGCAAAGTATCACGCTCCTCTAACCAGACATGGGTTCACACATACT
GACATTTGTCGTATATCAAGACGTTGGCAATCCCTCCGAATGGTAGCCAAAAACTACCCTAAATTAATTGC
AGCTTTACCTGATCTAACACGAACTCATATAGTTGATATTGCTCGTCAGCGATCCGGAGATCTAGCTCTCG
AAGCTCTATTACCTGTTGCTACGGCTCTCGCCGCGGCGCCGCTTCGTCTCAGAGCATCCCAAATCGCAATC
ATAGCTCAGTGTGGAGAGAGGCCAGCTATACTCGCTCTGCATAGATTAAGGAGAAAGCTCACGGGGGCCCC

ACTAAACCTGAGCACCGAGCAGGTGGTGGCCGTCGCCAGCCACAATGGCGGCAAGCAGGCACTGGAGGCGG
TGAGGGCGCAACTGCTGGACCTGCGCGCCGCACCCTATGAACTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATA
TGAGCTCTCTCTCTCTCTCTACTAGT

YP C-terminus fragment (*Nde*I-F2-F3-L2-*Spe*I clone):

CTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGC
ACTGTTGCTGGAACTGCGCGCAGCGCCATATGAGCTCAGCCACGAGCGAGTTGTGGCAATTGTGTGTAATG
GAGGACGAAGTGCCCTTGAAGCCGCACGACATGGACTCCCTGTTAGAGACATAAGAAGAATTCGTAATAGG
AAGACGTCAGATGCAACTCTACCTGCCCCTGTATTGGGTCCAACGCCACAGGAATTGGTTGCCGTCCTGCA
CTTCTTTAGGACACATCAACAGCCAAGGCAAGCCTTCGTGGATGCACTTACTGCATTCCAGACAACCAGAA
GGGCACTGCTTAGGCTATTATCCAGTGCTGGAGTCACTGAAATTGAGGCCCTGGGTGGGATGATTCCTGAT
GCTGCCGAACGTTGGCAGAGATTGTTAGGTCGTCTCGGGATCCGTCCAGCCACCGACACTGTGGTCACGTC
ACCTGATCCAATGCAGGGATTTGCACAATCCCTGGAGAGAAGCTTGATGTCATCTGGAACTACGGAGCAGT
CAGCTAGCCCATCTCAACACAGATGGCCAGCGGGAACAGCAAAAACGCCAGAATCAGCCCGTCACAGACTA
GACAACGCAGCTCAACCTCCCATGCAATGGCCTGATCAACTCGTGTGGAGTCAAAGGAGGAGGTGCACCGC
TAGCTCACAGACTCCTCCGCCTGGAGCCGCTAGCGTGCCGGCTAACCTTCAGTGGGGAGCACGAGCTCAGA
GAACAAGTTCTAGACTCCAGCCTGAACTGAGGCCTATGCCTGCGAGGGTTGTGCCAGCGTCAGCACAACCT
AGTCCTTTAGGAGACCTTGAGTTTGGACTCCCTGATCCTGGTACCCCTACAGCAGCTGATTTAGCCCTGGA
CCTTGACTGGCTTCTCCAACTACTAGACCTCTGAACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAA
GCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGCTCTCTCT
CTCTCTCTACTAGT

For RTL2-NP backbone:

NP N-terminus fragment (L1+F1-*Sal*I-*Nde*I-*Spe*I clone):

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTAT
AATGCCAACTTTGTACAAAAAAGCAGGCTCCATGAGAATTGGTAAAAGCTCTGGGTGGCTGAATGAATCCG
TCAGTCTTGAATACGAGCATGTTAGCCCTCCTACTCGTCCTAGAGATACAAGACGTAGACCTCGAGCTGCG
GGAGACGGAGGGCTAGCTCATTTACATAGGAGGCTTGCTGTCGGATATGCCGAAGATACTCCGAGGACGGA

AGCTCGTTCTCCTGCCCCAAGACGTCCTCTCCCCGTGGCACCCGCGTCAGCGCCTCCTGCACCTTCTCTCG

TTCCAGAGCCTCCTATGCCGGTATCCCTTCCAGCTGTAAGTAGCCCACGTTTCTCTGCCGGATCGTCAGCT

GCCATCACCGACCCTTTTCCTAGTTTGCCTCCTACGCCGGTTCTTTATGCAATGGCTCGTGAGCTTGAGGC

ATTAAGCGACGCTACTTGGCAGCCTGCGGTGCCATTGCCTGCAGAACCACCTACTGATGCACGAAGAGGAA

ACACAGTTTTCGATGAAGCTTCTGCTAGTTCCCCGGTAATCGCTTCAGCTTGTCCGCAGGCTTTTGCATCA

CCTCCTCGTGCCCCTAGGTCGGCTCGTGCTCGAAGAGCACGAACAGGTGGTGATGCCTGGCCCGCACCAAC

TTTCCTTTCACGTCCAAGTTCTAGCCGTATCGGGCGAGACGTGTTCGGAAAGTTGGTTGCTCTCGGGTACT

CGAGAGAGCAAATCAGAAAGCTTAAGCAAGAATCTCTCAGTGAAATAGCTAAGTATCACACCACCCTCACG

GGACAAGGATTCACGCACGCAGATATCTGCCGAATCAGCAGGCGACGACAGTCACTTAGAGTCGTGGCCCG

AAATTACCCTGAATTGGCAGCTGCATTACCTGAATTGACTAGGGCGCATATAGTTGATATTGCTAGACAGA

GGTCTGGAGATTTGGCATTGCAAGCATTGCTTCCAGTTGCTACCGCCTTGACAGCAGCCCCACTCAGATTG

TCCGCTTCTCAAATCGCTACTGTGGCTCAGTATGGTGAAAGGCCTGCTATCCAAGCACTCTACAGATTACG

TCGTAAACTTACCCGTGCTCCTCTCCACCTGTCGACCGAGCAGGTGGTGGCCGTCGCCAGCCACAATGGCG

GCAAGCAGGCACTGGAGGCGGTGAGGGCGCAACTGCTGGACCTGCGCGCCGCACCCTATGAACTCAGTACC

GAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCATATGAGGCTCTCTCTCTCTCTCTACTAGT

NP C-terminus fragment (*Nde*I-F2-F3-L2-*Spe*I clone):

CTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGC

ACTGTTGCTGGAACTGCGCGCAGCGCCATATGAGTTATCTCCTGAACGAGTTGCAGCAATCGCCTGCAATG

GAGGACGAAGTGCTGTTGAAGCAGTAAGACAAGGACTCCCAGTCAAAGCTATTAGAAGAATTAGAAGAGAG

AAAGCCCCGGTGGCTGGTCCTCCACCAGCATCACTTGGTCCTACCCCACAAGAGCTTGTGGCTGTGCTTCA

CTTTTTTCGAGCACACCAACAACCTAGACAAGCATTTGTTGATGCATTGGCCGCATTTCAAGCAACGAGAC

CCGCTCTTCTTAGATTATTATCGTCAGTAGGGGTTACCGAGATCGAGGCATTGGGAGGAACAATTCCTGAC

GCTACTGAGAGGTGGCAAAGGCTCCTCGGTAGACTGGGGTTTAGGCCGGCCACAGGAGCGGCTGCTCCCAG

CCCGGATAGTTTGCAGGGTTTCGCTCAGTCGCTTGAACGTACTTTGGGTTCGCCTGGGATGGCAGGACAAT

CAGCTTGCTCCCCTCATAGGAAGAGGCCTGCTGAAACCGCAATAGCTCCTAGAAGTATTCGACGTTCACCA

AATAACGCGGGGCAACCTTCTGAGCCGTGGCCTGATCAATTAGCTTGGCTACAGCGACGAAAACGAACGGC

CCGATCTCACATAAGGGCTGATTCCGCTGCGTCTGTGCCTGCAAATCTGCACTTAGGTACTAGAGCACAAT

TTACTCCAGACAGACTAAGGGCCGAACCAGGGCCTATTATGCAGGCGCATACCTCTCCTGCTTCAGTCAGT

TTCGGCTCTCATGTAGCTTTCGAACCTGGATTACCTGATCCTGGAACTCCTACTTCAGCGGATCTCGCTTC

GTTCGAAGCCGAACCATTTGGTGTGGGTCCGCTCGACTTTCACTTGGATTGGCTTCTCCAAATTCTTGAAA

CTTGAAACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAA

CAGGTCACTATCAGTCAAAATAAAATCATTATTTGCTCTCTCTCTCTCTCTACTAGT

For RTL3-CAQ backbone:

CAQ N-terminus fragment (L1+F1-*Sal*I-*Nde*I-*Spe*I clone):

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTAT

AATGCCAACTTTGTACAAAAAAGCAGGCTCCATGCGTAGAAGAACAGCGGTCGGTCGAGTTCCTGGTGCTT

CGCGTTCCGGAACTTCACCCCTAGTGCTTTCCCAACCGCTTAGCCGTGTGAGTGCCTCACCACAACCTGCA

CGTTCCTCGGCTATAAGTTCTCCGAATTTTTCAGCTGGAAATCCAACCGCCTTCGCAAATCCTAGCCCGAG

CTTGCCCCCTACTCCAGTTTTACCACCCACACCCGTCCTTTACGCAATGGCTAGAGAGTTGGAGGAACTTC

ATAATGCTACATGGCAGCCCGCTGTCCCCCTGACAGCGGAACCATTGGCCGATGCTAGAAGGGACAATACT

GTGGTTGATGGTGGAAGTGGATCTAGTCCTGCCATCGCATCAGCTCGTCCGCAGGCATTCGCTGGACCGCC

AAGAGCGCCTAGATCAGCCAGAGCTAGAAGAGCTCGTACAGCGGGAGATGCTTGGCCAGCTCCGGCTTATT

TGGGTCCTAGCCCTAGTCCGTCACCTTCGCCTTCTCCATCGCCAAGAATCGCTCCGGACCTTAGTGGGAAA

CTTGCAGCATTGGGATACAGCAGGGAGCAGATTCGTAAGCTAAAACAAGAATCATTAGCTGAGGTGGCTAA

ATATCATGCAACGCTAGCTGGACAAGGATTTACTCATGCAGATATTTGCAGAATTTCTAGGCGTTGGCAGT

CGCTACGTGTAGTTGCTAACAACTATCCTGAACTCATGGCAGCATTACCTAGGTTGACAACTGCTCAGATA

GTTGATATTGCTAGGCAAAGGAGCGGTGACTTGGCTTTACAGGCACTCCTGCCGGTGGCTGCGGCTCTCAC

TGCCGCACCTTTGGGGCTCTCTGCGTCTCAAATTGCTACTGTGGCTCAATACGGTGAGAGGCCAGCTATTC

AAGCACTGTATAGACTGAGGAGGAAGCTGACTCGAGCACCACTAGGCCTGTCGACCGAGCAGGTGGTGGCC

GTCGCCAGCCACAATGGCGGCAAGCAGGCACTGGAGGCGGTGAGGGCGCAACTGCTGGACCTGCGCGCCGC

ACCCTATGAACTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGG

CAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAGCTCTCTCTCTCTCTACTAGT

CAQ C-terminus fragment (*Nde*I-F2-F3-L2-*Spe*I clone):

CTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGC

ACTGTTGCTGGAACTGCGCGCAGCGCCATATGAGCTTTCGCCTGAGAGAGTCGTCGCTATCGTCTGTATCG
GCGGCAGGTCCGCAGTTGAGGCCATCAGACATGGCCTTCCTGTCAAGAGAATACAGAGAATCCGAAGACGA
AAAGCTCCCGAAGCAACCCCGCCAGCAGGGCCTTTCGGGCCTACACCACAAGAACTCGTGGCCGTCCTTCA
TTTCTTTAGAGCTCATAGACAACCTCGTCATGCATTTGTTGATGCGCTGACAGAGTTCCAAATCACTAGAG
CTGCACTCTTAAGACTGCTGTCAAGCGCTGGTGTGACTGAGATTGAGGCTCTTGGAGGCACGATCCCAGAT
GCTGCTGAAAGATGGCAACGACTTCTCGGCCGACTTGGTACCAGACAGGCTATTGGAGTAGCTGCTCCCTC
ACCAGATTCGTTGCAAGGTTTCGCTCAGTCTCTTGAAAGATCTATGCTTTCGCCGGGGATAGCAGAACAAT
CAGCATCTCCACCCCCAAGAGAGAGGCCAGCTGAAACCGCTATTGCACCCAGAAGTAATACCGGTAGGCCA
TCTGAGCCATGGTTGGACCAACCTGCTTGGTCTCAACATCGTAAGAGAACTGTGAGGTCTCGAATGAGAGC
TGATGCAGTGGCAAACATGCCTGCCAATCCTCATCGTGACGCCAGAGCTCAGTTTACGCCTGGATGCCTCC
CTCCTGCTCCTGGGCCCATGCCTGCCCACATCTCTCCAGCCAGTGCTGGTTCTAGTTCTCACGTCGCTTCT
GGTAGTGTTCTACCTGACCCTGGTACCCCTACTGGAAGCGACCTTGCAATGCTCGAGGCGGAGTCATTTGG
AGCCGCTGCTCTTGCCTTTGATTTTGATAGCTTTTTACAGATGCTCGATGTTTGAACCCAGCTTTCTTGTA
CAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATA
AAATCATTATTTGCTCTCTCTCTCTCTCTACTAGT

Supplementary Sequence 3. *Ralstonia* di-repeats sequences for ordered assembly.

F1

<HN-

HN GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCG
TTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGT
GGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCG
CAGCGCCCTACGAGCTCAGGAGACG

<HN-ND

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTG
GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC
CATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG
CGCCCTACGAGCTCAGGAGACG

<HN-NP

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<HN-NT

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<HN-SH

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<HN-SI

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<HN-SN

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

>SN-SN

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCAGCAACGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<SI-SI

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<ND-ND

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<NP-NP

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<NT-NT

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<IS-IS

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCATCAGCGGCGGCAAGCCGGCGTTG

GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC

CATCGCCAGCATCAGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG

CGCCCTACGAGCTCAGGAGACG

<IG-IG

GGGCCCCACTAAACCTGTCGACCGAGCAGGTAGTGGCCATCGCCAGCATCGGCGGCGGCAAGCCGGCGTTG
GAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGC
CATCGCCAGCATCGGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAG
CGCCCTACGAGCTCAGGAGACG

F2

<ND-HN

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAG
TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCCACAATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA
CGAGCTCAGTACGGAGACG

<ND-ND

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAG
TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA
CGAGCTCAGTACGGAGACG

<ND-NP

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAG
TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA
CGAGCTCAGTACGGAGACG

<ND-NT

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAG
TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA
CGAGCTCAGTACGGAGACG

<ND-SH

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAG
TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA
CGAGCTCAGTACGGAGACG

<ND-SI

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAG
TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA
CGAGCTCAGTACGGAGACG

<ND-SN

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAG
TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA
CGAGCTCAGTACGGAGACG

<SN-SN

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAG
TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA
CGAGCTCAGTACGGAGACG

<SI-SI

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAG
TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC
AGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA
CGAGCTCAGTACGGAGACG

<HN-HN

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC

AGCCACAATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA

CGAGCTCAGTACGGAGACG

<NP-NP

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC

AGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA

CGAGCTCAGTACGGAGACG

<NT-NT

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC

AGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA

CGAGCTCAGTACGGAGACG

<IS-IS

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATCAGCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC

AGCATCAGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA

CGAGCTCAGTACGGAGACG

<IG-IG

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATCGGCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCC

AGCATCGGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTA

CGAGCTCAGTACGGAGACG

F3

<NP-HN

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA

GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCC

ACAATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG

CTCAGTACCGAGGGAGACG

<NP-ND

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA

GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA

ATGATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG

CTCAGTACCGAGGGAGACG

<NP-NP

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA

GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA

ATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG

CTCAGTACCGAGGGAGACG

<NP-NT

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA

GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA

ATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG

CTCAGTACCGAGGGAGACG

<NP-SH

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA

GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA

GCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG

CTCAGTACCGAGGGAGACG

<NP-SI

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA

GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA

GCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG

CTCAGTACCGAGGGAGACG

<NP-SN

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA
GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA
GCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG
CTCAGTACCGAGGGAGACG

<SN-SN

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAA
GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA
GCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG
CTCAGTACCGAGGGAGACG

<SI-SI

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA
GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAG
CAGGTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGA
ACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGGGAGACG

<HN-HN

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTGGAGGCAGTCAA
GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCC
ACAATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG
CTCAGTACCGAGGGAGACG

<ND-ND

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAGTCAA
GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA
ATGATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG
CTCAGTACCGAGGGAGACG

<NT-NT

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA

GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA

ATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG

CTCAGTACCGAGGGAGACG

<IS-IS

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCATCAGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA

GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA

TCAGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG

CTCAGTACCGAGGGAGACG

<IG-IG

CGTCTCCGTACCGAGCAGGTAGTGGCCATCGCCAGCATCGGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAA

GGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCA

TCGGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAG

CTCAGTACCGAGGGAGACG

F4

<NT-HN

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCACAA

TGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<NT-ND

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGA

TGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<NT-NP

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCC

CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<NT-NT

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATAC

CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<NT-SH

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCCA

CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<NT-SI

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAT

CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<NT-SN

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAA

CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<SN-SN

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAA

CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<SI-SI

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAT

CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<HN-HN

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCACAA

TGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<ND-ND

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGA

TGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<NP-NP

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCC

CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<IS-IS

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCATCAGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA

CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATCAG

CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA

GTACCGAGCAGGGGAGACG

<IG-IG

CGTCTCCCGAGCAGGTAGTGGCCATCGCCAGCATCGGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCA
CTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATCGG
CGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCA
GTACCGAGCAGGGGAGACG

F5

<SH-HN

CGTCTCCCAGGTAGTGGCCATCGCCAGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT
TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGC
GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC
CGAGCAGGTAGTGGAGACG

<SH-ND

CGTCTCCCAGGTAGTGGCCATCGCCAGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT
TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGC
GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC
CGAGCAGGTAGTGGAGACG

<SH-NP

CGTCTCCCAGGTAGTGGCCATCGCCAGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT
TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGC
GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC
CGAGCAGGTAGTGGAGACG

<SH-NT

CGTCTCCCAGGTAGTGGCCATCGCCAGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT
TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGC
GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC
CGAGCAGGTAGTGGAGACG

<SH-SH

CGTCTCCCAGGTAGTGGCCATCGCCAGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCCACGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<SH-SI

CGTCTCCCAGGTAGTGGCCATCGCCAGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCATCGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<SH-SN

CGTCTCCCAGGTAGTGGCCATCGCCAGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAACGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<SN-SN

CGTCTCCCAGGTAGTGGCCATCGCCAGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAACGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<SI-SI

CGTCTCCCAGGTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCATCGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<HN-HN

CGTCTCCCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<ND-ND

CGTCTCCCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<NP-NP

CGTCTCCCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<NT-NT

CGTCTCCCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<IS-IS

CGTCTCCCAGGTAGTGGCCATCGCCAGCATCAGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATCAGCGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

<IG-IG

CGTCTCCCAGGTAGTGGCCATCGCCAGCATCGGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGT

TGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATCGGCGGC

GGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCAGTAC

CGAGCAGGTAGTGGAGACG

F6

<SI-HN

CGTCTCCTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<SI-ND

CGTCTCCTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<SI-NP

CGTCTCCTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<SI-NT

CGTCTCCTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<SI-SH

CGTCTCCTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCCACGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<SI-SI

CGTCTCCTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCATCGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<SI-SN

CGTCTCCTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAACGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<SN-SN

CGTCTCCTAGTGGCCATCGCCAGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAACGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<HN-HN

CGTCTCCTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<ND-ND

CGTCTCCTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<NP-NP

CGTCTCCTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<NT-NT

CGTCTCCTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCC

ATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGC

GCCATATGAG

<IS-IS

CGTCTCCTAGTGGCCATCGCCAGCATCAGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATCAGCGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

<IG-IG

CGTCTCCTAGTGGCCATCGCCAGCATCGGCGGCGGCAAGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCT

GGAACTGCGCGCAGCGCCCTACGAGCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATCGGCGGCGGCA

AGCCGGCGTTGGAGGCAGTCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCATATGAG

Supplementary Sequence 4. DNA fragments used for the generation and cloning of the RTL-repeat concatemers in the *dHax3* backbone.

>HN

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCACAATGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>ND

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGATGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>NP

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATCCCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>NT

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATACCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>SH

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCCACGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>SI

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCATCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>SN

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAGCAACGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>NN

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATAATGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>NG

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATGGCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>NK

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATAAGGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>NI

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCAATATAGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>HD

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCATGATGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>GN

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCGGCAATGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>RS

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCGTTCTGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>HG

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCCATGGCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>IG

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATAGGCGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>IS

CGTCTCCCTCAGTACCGAGCAGGTAGTGGCCATCGCCAGCATATCTGGCGGCAAGCCGGCGTTGGAGGCAG

TCAAGGCACTGTTGCTGGAACTGCGCGCAGCGCCCTACGAGCTCACGAGACG

>NH-SI

CGTCTCCCTCACACCGCAGCAGGTGGTGGCGATCGCCAGCCACAATGGCGGAAAACAGGCGTTGGAGGCGG

TCACTGTGCAATTGCGCGTGCTGCGTGGGGCCCGCTATGGGCTGAGCACCGAGCAGGTGGTGGCCGTTGCC

AGCAGCATCGGCGGCAAACAAGCCCTGGCGGCGGTCGAGGCACAACTGCTGAGACTGCGCGCCGCACCCTA

TGAACTCACGAGACG

>NT-ND

CGTCTCCCTCACACCGCAGCAGGTGGTGGCGATCGCCAGCAATACCGGCGGAAAACAGGCGTTGGAGGCGG

TCACTGTGCAATTGCGCGTGCTGCGTGGGGCCCGCTATGGGCTGAGCACCGAGCAGGTGGTGGCCGTTGCC

AGCAATGACGGCGGCAAACAAGCCCTGGCGGCGGTCGAGGCACAACTGCTGAGACTGCGCGCCGCACCCTA

TGAACTCACGAGACG

>SH-NP

CGTCTCCCTCACACCGCAGCAGGTGGTGGCGATCGCCAGCAGCCACGGCGGAAAACAGGCGTTGGAGGCGG

TCACTGTGCAATTGCGCGTGCTGCGTGGGGCCCGCTATGGGCTGAGCACCGAGCAGGTGGTGGCCGTTGCC

AGCAATCCCGGCGGCAAACAAGCCCTGGCGGCGGTCGAGGCACAACTGCTGAGACTGCGCGCCGCACCCTA

TGAACTCACGAGACG

>SN-HD

CGTCTCCCTCACACCGCAGCAGGTGGTGGCGATCGCCAGCAGCAATGGCGGAAAACAGGCGTTGGAGGCGG

TCACTGTGCAATTGCGCGTGCTGCGTGGGGCCCGCTATGGGCTGAGCACCGAGCAGGTGGTGGCCGTTGCC

AGCCACGACGGCGGCAAACAAGCCCTGGCGGCGGTCGAGGCACAACTGCTGAGACTGCGCGCCGCACCCTA

TGAACTCACGAGACG

>SS-RN

CGTCTCCCTCACACCGCAGCAGGTGGTGGCGATCGCCAGCTCTTCTGGCGGAAAACAGGCGTTGGAGGCGG
TCACTGTGCAATTGCGCGTGCTGCGTGGGGCCCGCTATGGGCTGAGCACCGAGCAGGTGGTGGCCGTTGCC
AGCCGCAATGGCGGCAAACAAGCCCTGGCGGCGGTCGAGGCACAACTGCTGAGACTGCGCGCCGCACCCTA
TGAACTCACGAGACG

>*PpuM*I-ND

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG
AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG
ATCGCCAGCAATGATGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC
CCGCTATGGCCTCACGAGACG

>*PpuM*I-HN

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG
AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG
ATCGCCAGCCACAATGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC
CCGCTATGGCCTCACGAGACG

>*PpuM*I-NP

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG
AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG
ATCGCCAGCAATCCCGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC
CCGCTATGGCCTCACGAGACG

>*PpuM*I-NT

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG
AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG
ATCGCCAGCAATACCGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC
CCGCTATGGCCTCACGAGACG

>*PpuM*I-SH

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG
AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG
ATCGCCAGCAATACCGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC

CCGCTATGGCCTCACGAGACG

>*PpuM*I-SI

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG

AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG

ATCGCCAGCAGCATCGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC

CCGCTATGGCCTCACGAGACG

>*PpuM*I-SN

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG

AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG

ATCGCCAGCAGCAACGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC

CCGCTATGGCCTCACGAGACG

>*PpuM*I-NN

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG

AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG

ATCGCCAGCAATAATGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGgGC

CCGCTATGGCCTCACGAGACG

>*PpuM*I-NG

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG

AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG

ATCGCCAGCAATGGCGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC

CCGCTATGGCCTCACGAGACG

>*PpuM*I-HD

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG

AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG

ATCGCCAGCCATGATGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC

CCGCTATGGCCTCACGAGACG

>*PpuM*I-NI

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG

AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG

ATCGCCAGCAATATCGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC

CCGCTATGGCCTCACGAGACG

>*PpuM*I-NS

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG

AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTCACACCGCAGCAGGTGGTGGCG

ATCGCCAGCAATTCTGGCGGAAAACAGGCGTTGGAGGCGGTCACTGTGCAATTGCGCGTGCTGCGTGGGGC

CCGCTATGGCCTCACGAGACG

>0.5-*Sac*I-NI

CGTCTCCCTCACACCACAACAAGTCGTCGCAATCGCTAGTAATATAGGAGGTAGACCTGCATTGGAGTCGA

TAGTCGCACAACTATCACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAGTCGCACTTGCA

TGTTTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCATTGATTAAAAG

AACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATGCTCAAGTCGTAAGAGTTTTGG

GTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTTGACGATGCAATGACTCAATTTGGAATGAGTAGA

CATGGACTCCTGCAATTATTTCGAAGGGTCGGAGTTACAGAGCTC

>0.5-*Sac*I-NN

CGTCTCCCTCACACCACAACAAGTCGTCGCAATCGCTAGTAATAATGGAGGTAGACCTGCATTGGAGTCGA

TAGTCGCACAACTATCACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAGTCGCACTTGCA

TGTTTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCATTGATTAAAAG

AACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATGCTCAAGTCGTAAGAGTTTTGG

GTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTTGACGATGCAATGACTCAATTTGGAATGAGTAGA

CATGGACTCCTGCAATTATTTCGAAGGGTCGGAGTTACAGAGCTC

>0.5-*Sac*I-HD

CGTCTCCCTCACACCACAACAAGTCGTCGCAATCGCTAGTCATGATGGAGGTAGACCTGCATTGGAGTCGA

TAGTCGCACAACTATCACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAGTCGCACTTGCA

TGTTTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCATTGATTAAAAG

AACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATGCTCAAGTCGTAAGAGTTTTGG

GTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTTGACGATGCAATGACTCAATTTGGAATGAGTAGA

CATGGACTCCTGCAATTATTTCGAAGGGTCGGAGTTACAGAGCTC

>0.5-*Sac*I-ND

CGTCTCCCTCACACCACAACAAGTCGTCGCAATCGCTAGTAATGATGGAGGTAGACCTGCATTGGAGTCGA

TAGTCGCACAACTATCACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAGTCGCACTTGCA

TGTTTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCATTGATTAAAAG

AACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATGCTCAAGTCGTAAGAGTTTTGG

GTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTTGACGATGCAATGACTCAATTTGGAATGAGTAGA

CATGGACTCCTGCAATTATTTCGAAGGGTCGGAGTTACAGAGCTC

>0.5-*Sac*I-HN

CGTCTCCCTCACACCACAACAAGTCGTCGCAATCGCTAGTCATAATGGAGGTAGACCTGCATTGGAGTCGA

TAGTCGCACAACTATCACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAGTCGCACTTGCA

TGTTTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCATTGATTAAAAG

AACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATGCTCAAGTCGTAAGAGTTTTGG

GTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTTGACGATGCAATGACTCAATTTGGAATGAGTAGA

CATGGACTCCTGCAATTATTTCGAAGGGTCGGAGTTACAGAGCTC

>0.5-*Sac*I-SI

CGTCTCCCTCACACCACAACAAGTCGTCGCAATCGCTAGTTCTATAGGAGGTAGACCTGCATTGGAGTCGA

TAGTCGCACAACTATCACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAGTCGCACTTGCA

TGTTTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCATTGATTAAAAG

AACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATGCTCAAGTCGTAAGAGTTTTGG

GTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTTGACGATGCAATGACTCAATTTGGAATGAGTAG2

ACATGGACTCCTGCAATTATTTCGAAGGGTCGGAGTTACAGAGCTC

Supplementary Sequence 5. *Minimal Bs3* and *Bs4* promoter sequences.

Minimal Bs3 promoters:

m*Bs3* polyA

GCCTGACCAATTTTATAAAAAAAAAAAAAAACCTCACAACTTCAAGTTATCATCCCCTTTCTCTTTTCTCC

TCTTGTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCCTAGTTGCACATATATTTC

m*Bs3* polyC

GCCTGACCAATTTTATCCCCCCCCCCCCCCCTCACAACTTCAAGTTATCATCCCCTTTCTCTTTTCTCCTC

TTGTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCCTAGTTGCACATATATTTC

m*Bs3* polyG

GCCTGACCAATTTTATGGGGGGGGGGGGGGCCTCACAACTTCAAGTTATCATCCCCTTTCTCTTTTCTCC

TCTTGTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCCTAGTTGCACATATATTTC

m*Bs3* polyT

GCCTGACCAATTTTATTTTTTTTTTTTTTTCCTCACAACTTCAAGTTATCATCCCCTTTCTCTTTTCTCC

TCTTGTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCCTAGTTGCACATATATTTC

m*Bs3 dHax3*

GCCTGACCAATTTTATCCCTTTATCTCTCCTCACAACTTCAAGTTATCATCCCCTTTCTCTTTTCTCCTCT

TGTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCCTAGTTGCACATATATTTC

Minimal Bs4 promoters:

Bs4 polyC:

CCGCGGCCGCCCCCTTCACCTCCCCCCCCCCCCCCTTCTTTCTTGTATATAACTTTGTCCAAAATATCATC

AATTGATCTCATCCATACAATTTATTTTTAATCGAATCTTCTAGACCCAAGGGTGGGCGCGCCG

Bs4 polyC.dT:

CCGCGGCCGCCCCCTTCACCCCCCCCCCCCCCCCTTCTTTCTTGTATATAACTTTGTCCAAAATATCATCA

ATTGATCTCATCCATACAATTTATTTTTAATCGAATCTTCTAGACCCAAGGGTGGGCGCGCCG

Bs4 polyG:

CCGCGGCCGCCCCCTTCACCTGGGGGGGGGGGGGGTTCTTTCTTGTATATAACTTTGTCCAAAATATCATC

AATTGATCTCATCCATACAATTTATTTTTAATCGAATCTTCTAGACCCAAGGGTGGGCGCGCCG

Bs4 polyG.dT:

CCGCGGCCGCCCCCTTCACCGGGGGGGGGGGGGGTTCTTTCTTGTATATAACTTTGTCCAAAATATCATCA

ATTGATCTCATCCATACAATTTATTTTTAATCGAATCTTCTAGACCCAAGGGTGGGCGCGCCG

Bs4 polyA:

CCGCGGCCGCCCCCTTCACCTAAAAAAAAAAAAAAATTCTTTCTTGTATATAACTTTGTCCAAAATATCATC

AATTGATCTCATCCATACAATTTATTTTTAATCGAATCTTCTAGACCCAAGGGTGGGCGCGCCG

Bs4 polyA.dT:

CCGCGGCCGCCCCCTTCACCAAAAAAAAAAAAAAATTCTTTCTTGTATATAACTTTGTCCAAAATATCATCA

ATTGATCTCATCCATACAATTTATTTTTAATCGAATCTTCTAGACCCAAGGGTGGGCGCGCCG

Bs4 polyT:

CCGCGGCCGCCCCCTTCACCTTTTTTTTTTTTTTTTTCTTTCTTGTATATAACTTTGTCCAAAATATCATC

AATTGATCTCATCCATACAATTTATTTTTAATCGAATCTTCTAGACCCAAGGGTGGGCGCGCCG

Bs4 dHax3:

CCGCGGCCGCCCCCTTCACCTCCCTTTATCTCTTTCTTTCTTGTATATAACTTTGTCCAAAATATCATCAA

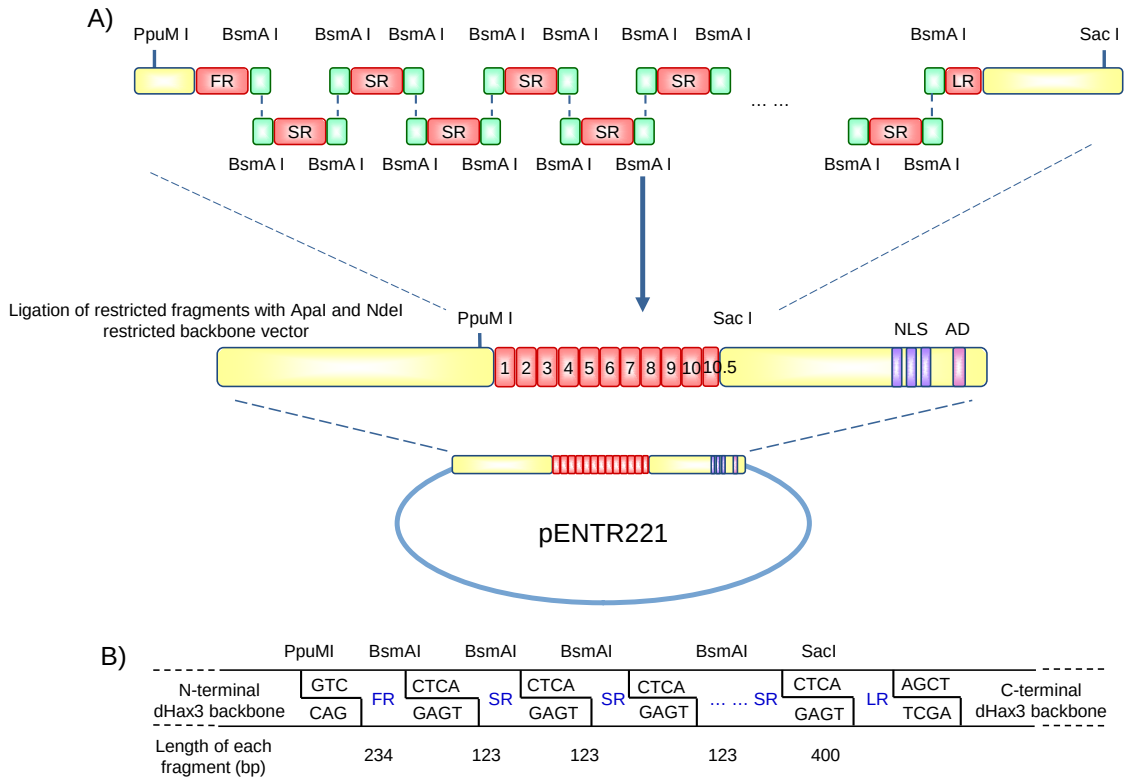
TTGATCTCATCCATACAATTTATTTTTAATCGAATCTTCTAGACCCAAGGGTGGGCGCGCCG

Bs4 dEBE:

CCGCGGCCGCCCCCTTCACCTTCTTTCTTGTATATAACTTTGTCCAAAATATCATCAATTGATCTCATCCA

TACAATTTATTTTTAATCGAATCTTCTAGACCCAAGGGTGGGCGCGCCG

Supplementary Figure 1. Assembly and cloning of *Ralstonia* RVDs to *Xamthomonas dHax3* backbone.A) Assembly of *dHax3*.RR.RVD tandem repeat domain. (B) Schematic representation of the *dHax3*.RR.RVD repeat region showing the restriction fragments and the overhang sequence of each fragment produced by a restriction digestion. FR, first repeat type; SR, single repeat; LR, last repeat type.



Supplementary Table 1. Primers for cloning of RTL repeat fragments into RTL backbones.

Primer name	Primer sequence in 5' to 3' orientation	Primer purpose
YP.REPA.*Apa*I F	CTCACGGGGGCCCCACTAAACCTGAGCACCGAGCAGGTGG	YP Repeat Assembly
YP.REPA.*Nde*I R	TGCGACGACTTGTTGTGGTGTGAGCTCATATGGCGCTGCGCGCA GTTC	YP Repeat Assembly
CAQ.REPA.*Sal*I F	GCACCACTAGGCCTGTCGACCGAGCAGGTGGTGGCC	CAQ Repeat Assembly
YP.REPA.*Nde*I R	TGCGACGACTTGTTGTGGTGTGAGCTCATATGGCGCTGCGCGCA GTTC	CAQ Repeat Assembly
NP.REPA.*Sal*I F	CTCACGGGCGCACCACTACACCTGTCGACCGAGCAGGTGGTGGC	NP Repeat Assembly

	C	
YP.REPA.*Nde*I R	TGCGACGACTTGTTGTGGTGTGAGCTCATATGGCGCTGCGCGCA GTTC	NP Repeat Assembly

6.3 Supplementary Information for Chapter 3

Name	Description
Supplementary Sequence 1	Sequences of individual fragments used for the assembly.
Supplementary Sequence 2	*In silico* assembled protein *dHax3* with 13 repeats.
Supplementary Sequence 3	*Bs3* promoter sequences.
Supplementary Sequence 4	*PDS*3 cDNA sequence.
Supplementary Table 1	TALE-TF/TALEN, RVD sequences and their corresponding EBE sequences.
Supplementary Table 2	Primer for sequence confirmation of assembled fragments.

Supplementary Sequence 1. Sequences of individual fragments used for the assembly. Changes in the underlined positions will produce variants of each fragment. The variants of fragments F1 to F7 are underlined and indicated in tables.

Fragment 1:

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG
AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTTACTCCTGAACAGGTTGTCGCA
ATAGCTTCAAATAAAGGCGGAAAACAAGCTCTTGAAACAGTGCAACGTCTCCTTCCCGTCCTCTGTCAG

The F1 variant sequence translates to NK amino acids, additional variants include NG (AATGGG), HD (CATGAT) and NI (AATATT) as indicated in the table below.

NK	AATAAA
NG	AATGGG
HD	CATGAT

NI	AATATT

Fragment 2:

CAACGTCTCCTTCCCGTCCTCTGTCAGGCTCACGGATTGACTCCTCAGCAGGTCGTCGCAATTGCATCAAA

CAAGGGAGGCAAACAAGCTTTAGAAACAGTACAAAGACTATTGCCCGTTCTTTGCCAAGCGCATGGGTTAA

CTCCCGAACAAGTCGTTGCCATTGCAAGTAATAAGGGAGGTAAACAAGCTCTCGAAACGGTTCAAGCACTT

TTACCCGTTCTCTGTCAAGCACATGGTCTCACACCTGAACAA

The F2 variant sequence translates to NK–NK amino acids, in the two underlined positions, additional variants are indicated in the table below.

NK–NK	AACAAG - AATAAG
NG-NG	AATGGG -AATGGG
HD-HD	CATGAT - CATGAT
NI-NI	AATATT - AATATT
NK-HD	AACAAG - CATGAT
NK-NI	AACAAG - AATATT
NK-NG	AACAAG - AATGGG
HD-NG	CATGAT - AATGGG
HD-NI	CATGAT - AATATT
HD-NK	CATGAT - AATAAG
NG-HD	AATGGG - CATGAT
NG-NI	AATGGG - AATATT
NG-NK	AATGGG - AATAAG
NI-HD	AATATT - CATGAT
NI-NG	AATATT - AATGGG
NI-NK	AATATT - AATAAG

Fragment 3:

CAAGCACATGGTCTCACACCTGAACAAGTAGTTGCTATCGCATCGAATATTGGTGGAAAACAAGCACTGGA

AACTGTACAAAGACTTTTGCCAGTTTTATGTCAAGCGCACGGTCTTACTCCTCAACAAGTTGTCGCCATTG

CCTCTCATGATGGTGGAAAACAAGCTCTTGAAACTGTCCAGAGACTTCTG

The F3 variant sequence translates to NI–HD amino acids, in the two underlined positions, additional variants are indicated in the table below.

NK–NK	AACAAG - AATAAG
NG-NG	AATGGG - AATGGG
HD-HD	CATGAT - CATGAT
NI-NI	AATATT - AATATT
NK-HD	AACAAG - CATGAT
NK-NI	AACAAG - AATATT
NK-NG	AACAAG - AATGGG
HD-NG	CATGAT - AATGGG
HD-NI	CATGAT - AATATT
HD-NK	CATGAT - AATAAG
NG-HD	AATGGG - CATGAT
NG-NI	AATGGG - AATATT
NG-NK	AATGGG - AATAAG
NI-HD	AATATT - CATGAT
NI-NG	AATATT - AATGGG
NI-NK	AATATT - AATAAG

Fragment 4:

CTTGAAACTGTCCAGAGACTTCTGCCCGTTCTATGTCAGGCTCATGGGCTAACCCCTCAACAGGTTGTTGC

AATCGCATCTCATGACGGAGGAAAACAAGCTTTAGAAACTGTCCAACGACTACTGCCCGTTCTCTGCCAAG

CACACGGACTTACCCCACAACAAGTTGTGGCAATAGCTTCTCACGATGGTGGTAAACAAGCCCTTGAGACT
GTTCAA

The F4 variant sequence translates to HD–HD amino acids, in the two underlined positions, additional variants are indicated in the table below.

NK–NK	AACAAG - AATAAG
NG-NG	AATGGG - AATGGG
HD-HD	CATGAT - CATGAT
NI-NI	AATATT - AATATT
NK-HD	AACAAG - CATGAT
NK-NI	AACAAG - AATATT
NK-NG	AACAAG - AATGGG
HD-NG	CATGAT - AATGGG
HD-NI	CATGAT - AATATT
HD-NK	CATGAT - AATAAG
NG-HD	AATGGG - CATGAT
NG-NI	AATGGG - AATATT
NG-NK	AATGGG - AATAAG
NI-HD	AATATT - CATGAT
NI-NG	AATATT - AATGGG
NI-NK	AATATT - AATAAG

Fragment 5:

GGTAAACAAGCCCTTGAGACtGTTCAAAGACTTCTACCAGTTCTTTGTCAGGCACATGGATTGACCCCACA
ACAGGTCGTAGCAATCGCATCTAATATAGGTGGTAAGCAAGCTCTAGAAACGGTACAAAGATTACTTCCCG
TGCTTTGTCAAGCTCATGGACTCACTCCTCAACAAGTGGTCGCTATTGCAAGTCATGATGGTGGAAAGCAA
GCACTAGAgACgGTCCAA

The F5 variant sequence translates to NI–HD amino acids, in the two underlined positions, additional variants are indicated in the table below.

NK–NK	AACAAG - AATAAG
NG-NG	AATGGG - AATGGG
HD-HD	CATGAT - CATGAT
NI-NI	AATATT - AATATT
NK-HD	AACAAG - CATGAT
NK-NI	AACAAG - AATATT
NK-NG	AACAAG - AATGGG
HD-NG	CATGAT - AATGGG
HD-NI	CATGAT - AATATT
HD-NK	CATGAT - AATAAG
NG-HD	AATGGG - CATGAT
NG-NI	AATGGG - AATATT
NG-NK	AATGGG - AATAAG
NI-HD	AATATT - CATGAT
NI-NG	AATATT - AATGGG
NI-NK	AATATT - AATAAG

Fragment 6:

GGAAAGCAAGCACTAGAGACGGTCCAACGACTCCTTCCTGTTCTCTGTCAAGCACATGGTCTTACGCCCGA

ACAAGTTGTTGCTATAGCTTCGAATAAGGGTGGAAAACAAGCTCTCGAaACCGTCCAAAGGCTCCTCCCAG

TACTTTGCCAAGCACATGGATTAACCCCTGAGCAAGTAGTTGCAATTGCCTCGAATAAGGGAGGAAAGCAA

GCTCTCGAGACCGTCCAA

The F6 variant sequence translates to NK–NG amino acids, in the two underlined positions, additional variants are indicated in the table below.

NK–NK	AACAAG - AATAAG
NG-NG	AATGGG - AATGGG
HD-HD	CATGAT - CATGAT
NI-NI	AATATT - AATATT
NK-HD	AACAAG - CATGAT
NK-NI	AACAAG - AATATT
NK-NG	AACAAG - AATGGG
HD-NG	CATGAT - AATGGG
HD-NI	CATGAT - AATATT
HD-NK	CATGAT - AATAAG
NG-HD	AATGGG - CATGAT
NG-NI	AATGGG - AATATT
NG-NK	AATGGG - AATAAG
NI-HD	AATATT - CATGAT
NI-NG	AATATT - AATGGG
NI-NK	AATATT - AATAAG

Fragment 7:

GGAAAACAAGCTCTCGAgACCGTCCAAAGGCTCCTCCCAGTACTTTGCCAAGCACATGGATTAACCCCTGA

GCAAGTAGTTGCAATTGCCTCGAACGGAGGAGGAAAGCAAGCATTAGAAACTGTTCAGAGACTTTTGCCTG

TCCTGTGTCAAGCCCACGGTCTAACACCACAACAAGTCGTCGCAATCGCTAGTAACAAAGGAGGTAGACCT

GCATTGGAGTCGATAGTCGCACAACTATCACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTT

AGTCGCACTTGCATGTTTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTG

CATTGATTAAAAGAACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATGCTCAAGTC

GTAAGAGTTTTGGGTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTTGACGATGCAATGACTCAATT

TGGAATGAGTAGACATGGACTCCTGCAATTATTTCGAAGGGTCGGAGTTACAGAGCTC

The F6 variant sequence translates to NG–NK amino acids, in the two underlined positions, additional variants are indicated in the table below.

NK–NK	AACAAG - AATAAG
NG-NG	AATGGG - AATGGG
HD-HD	CATGAT - CATGAT
NI-NI	AATATT - AATATT
NK-HD	AACAAG - CATGAT
NK-NI	AACAAG - AATATT
NK-NG	AACAAG - AATGGG
HD-NG	CATGAT - AATGGG
HD-NI	CATGAT - AATATT
HD-NK	CATGAT - AATAAG
NG-HD	AATGGG - CATGAT
NG-NI	AATGGG - AATATT
NG-NK	AATGGG - AATAAG
NI-HD	AATATT - CATGAT
NI-NG	AATATT - AATGGG
NI-NK	AATATT - AATAAG

Supplementary Sequence 2. *In silico* assembled protein *dHax3* with 13 repeats.

AGGTCCTCCACTCCAATTGGATACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCG

AAGCCGTTCATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTTACTCCTGAACAGGTTGTCGCA

ATAGCTTCAAATAAAGGCGGAAAACAAGCTCTTGAAACAGTGCAACGTCTCCTTCCCGTCCTCTGTCAGGC

TCACGGATTGACTCCTCAGCAGGTCGTCGCAATTGCATCAAACAAGGGAGGCAAACAAGCTTTAGAAACAG

TACAAAGACTATTGCCCGTTCTTTGCCAAGCGCATGGGTTAACTCCCGAACAAGTCGTTGCCATTGCAAGT

AATAAGGGAGGTAAACAAGCTCTCGAAACGGTTCAAGCACTTTTACCCGTTCTCTGTCAAGCACATGGtCT

CACACCTGAACAAGTAGTTGCTATCGCATCGAATATTGGTGGAAAACAAGCACTGGAAACTGTACAAAGAC

TTTTGCCAGTTTTATGTCAAGCGCACGGTCTTACTCCTCAACAAGTTGTCGCCATTGCCTCTCATGATGGT

GGAAAACAAGCTCTTGAAACTGTCCAGAGACTTCTGCCCGTTCTATGTCAGGCTCATGGGCTAACCCCTCA

ACAGGTTGTTGCAATCGCATCTCATGACGGAGGAAAACAAGCTTTAGAAACTGTCCAACGACTACTGCCCG

TTCTCTGCCAAGCACACGGACTTACCCCACAACAAGTTGTGGCAATAGCTTCTCACGATGGTGGTAAACAA

GCCCTTGAGACtGTTCAAAGACTTCTACCAGTTCTTTGTCAGGCACATGGATTGACCCCACAACAGGTCGT

AGCAATCGCATCTAATATAGGTGGTAAGCAAGCTCTAGAAACGGTACAAAGATTACTTCCCGTGCTTTGTC

AAGCTCATGGACTCACTCCTCAACAAGTGGTCGCTATTGCAAGTCATGATGGTGGAAAGCAAGCACTAGAg

ACgGTCCAACGACTCCTTCCTGTTCTCTGTCAAGCACATGGTCTTACGCCCGAACAAGTTGTTGCTATAGC

TTCGAATAAGGGTGGAAAACAAGCTCTCGAAACCGTCCAAAGGCTCCTCCCAGTACTTTGCCAAGCACATG

GATTAACCCCTGAGCAAGTAGTTGCAATTGCCTCGAACGGAGGAGGAAAGCAAGCTCTCGAgACCGTCCAA

AGGCTCCTCCCAGTACTTTGCCAAGCACATGGATTAACCCCTGAGCAAGTAGTTGCAATTGCCTCGAACGG

AGGAGGAAAGCAAGCATTAGAAACTGTTCAGAGACTTTTGCCTGTCCTGTGTCAAGCCCACGGTCTAACAC

CACAACAAGTCGTCGCAATCGCTAGTAACAAAGGAGGTAGACCTGCATTGGAGTCGATAGTCGCACAACTA

TCACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAGTCGCACTTGCATGTTTAGGAGGACG

ACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCATTGATTAAAAGAACAAACAGACGAA

TCCCGGAGAGAACTTCACATCGTGTAGCCGATCATGCTCAAGTCGTAAGAGTTTTGGGTTTCTTCCAATGT

CATTCCCACCCAGCTCAAGCTTTTGACGATGCAATGACTCAATTTGGAATGAGTAGACATGGACTCCTGCA

ATTATTTCGAAGGGTCGGAGTTACAGAGCTC

Supplementary Sequence 3. *Bs3* promoter sequences.

Bs3 promoter

TCATAGTCAAGCTAACGAAACTTATGCAAGGGAAATATGAAATTAGTATGCAAGTAAACTCAAAGAACTAA

TCATTGAACTGAAAGATCAATATATCAAAAAAAAAAAAAAAACAATAAAACCGTTTAACCGATAGATTAACC

ATTTCTGGTTCAGTTTATGGGTTAAACCACAATTTGCACACCCTGGTTAAACAATGAACACGTTTGCCTGA

CCAATTTTATTATATAAACCTAACCATCCTCACAACTTCAAGTTATCATCCCCTTTCTCTTTTCTCCTCTT

GTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCCTAGTTGCACATATATTTC

Bs3△*dHax3* EBE promoter

TCATAGTCAAGCTAACGAAACTTATGCAAGGGAAATATGAAATTAGTATGCAAGTAAACTCAAAGAACTAA
TCATTGAACTGAAAGATCAATATATCAAAAAAAAAAAAAAAACAATAAAACCGTTTAACCGATAGATTAACC
ATTTCTGGTTCAGTTTATGGGTTAAACCACAATTTGCACACCCTGGTTAAACAATGAACACGTTTGCCTGA
CCAATTTTATCCCTTTATCTCTAACCATCCTCACAACTTCAAGTTATCATCCCCTTTCTCTTTTCTCCTCT
TGTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCCTAGTTGCACATATATTTC

Bs3△*PDS*3.2F EBE promoter

TCATAGTCAAGCTAACGAAACTTATGCAAGGGAAATATGAAATTAGTATGCAAGTAAACTCAAAGAACTAA
TCATTGAACTGAAAGATCAATATATCAAAAAAAAAAAAAAAACAATAAAACCGTTTAACCGATAGATTAACC
ATTTCTGGTTCAGTTTATGGGTTAAACCACAATTTGCACACCCTGGTTAAACAATGAACACGTTTGCCTGA
CCAATTTTATCCTTTTACCAGAAACCATCCTCACAACTTCAAGTTATCATCCCCTTTCTCTTTTCTCCTCT
TGTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCCTAGTTGCACATATATTTC

Bs3△*PDS*3.2R EBE promoter

TCATAGTCAAGCTAACGAAACTTATGCAAGGGAAATATGAAATTAGTATGCAAGTAAACTCAAAGAACTAA
TCATTGAACTGAAAGATCAATATATCAAAAAAAAAAAAAAAACAATAAAACCGTTTAACCGATAGATTAACC
ATTTCTGGTTCAGTTTATGGGTTAAACCACAATTTGCACACCCTGGTTAAACAATGAACACGTTTGCCTGA
CCAATTTTATCCAACTAATTTAAACCATCCTCACAACTTCAAGTTATCATCCCCTTTCTCTTTTCTCCTCT
TGTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCCTAGTTGCACATATATTTC

Underlined sequences represent target sequence of dTALE-TFs.

Supplementary Sequence 4. *PDS*3 cDNA sequence.

AAATTACGTTGAGATGCATGGTCTCTCTCTACTCAATTAACCAAATAAGGAAAAGAATCATATGGTCATCA
ATTCGTAAATCAAAATTTTAATTTGTGTGGTATTTAATCCATCTACATGTTTCGTAAGCAACAAAAGAGCT
TGGTCTGAAAACCAAACAAGACCATATGGGCACTCGAATACTCCATTTTGTTATCGGCTACTTCCACTAGC
CTCCTCCTTCGCTGCGTCTCCTGTTTCTCTACTTCACGATTACTCGCTAGATTCATTGAAGCAGTTGTGAG

TTAAGTTGGAGAAAATGGTTGTGTTTGGGAATGTTTCTGCGGCGAATTTGCCTTATCAAAACGGGTTTTTG
GAGGCACTTTCATCTGGAGGTTGTGAACTAATGGGACATAGCTTTAGGGTTCCCACTTCTCAAGCGCTTAA
GACAAGAACAAGGAGGAGGAGTACTGCTGGTCCTTTGCAGGTAGTTTGTGTGGATATTCCAAGGCCAGAGC
TAGAGAACACTGTCAATTTCTTGGAAGCTGCTAGTTTATCTGCATCCTTCCGTAGTGCTCCTCGTCCTGCT
AAGCCTTTGAAAGTTGTAATTGCTGGTGCTGGATTGGCTGGATTGTCAACTGCAAAGTACCTGGCTGATGC
AGGCCACAAACCTCTGTTGCTTGAAGCAAGAGATGTTCTTGGTGGAAAGATAGCTGCATGGAAGGATGAAG
ATGGGGACTGGTATGAGACTGGTTTACATATTTTCTTCGGTGCTTATCCGAATGTGCAGAATTTATTTGGA
GAACTTGGGATCAATGATCGGTTGCAGTGGAAGGAACACTCCATGATTTTTGCTATGCCAAGTAAACCTGG
AGAATTTAGTAGATTTGACTTCCCAGATGTCCTACCAGCACCCTTAAATGGTATTTGGGCTATTTTGCGGA
ACAACGAGATGCTGACATGGCCAGAGAAAATAAAGTTTGCTATTGGACTTTTGCCAGCCATGGTCGGCGGT
CAGGCTTATGTTGAGGCCCAAGATGGTTTATCAGTCAAAGAATGGATGGAAAAGCAGGGAGTACCTGAGCG
CGTGACCGACGAGGTGTTTATTGCCATGTCAAAGGCGCTAAACTTTATAAACCCTGATGAACTGTCAATGC
AATGCATTTTGATAGCTTTGAACCGGTTTCTTCAGGAAAAACATGGTTCCAAGATGGCATTCTTGGATGGT
AATCCTCCGGAAAGGCTTTGTATGCCAGTAGTGGATCATATTCGATCACTAGGTGGGGAAGTGCAACTTAA
TTCTAGGATAAAGAAAATTGAGCTCAATGACGATGGCACGGTTAAGAGTTTCTTACTCACTAATGGAAGCA
CTGTCGAAGGAGACGCTTATGTGTTTGCCGCTCCAGTCGATATCCTGAAGCTCCTTTTACCAGATCCCTGG
AAAGAAATACCGTACTTCAAGAAATTGGATAAATTAGTTGGAGTACCAGTTATTAATGTTCATATATGGTT
TGATCGAAAACTGAAGAACACATATGATCACCTACTCTTTAGCAGAAGTAACCTTCTGAGCGTGTATGCCG
ACATGTCCTTAACTTGTAAGGAATATTACGATCCTAACCGGTCAATGCTGGAGCTAGTATTTGCACCAGCA
GAGGAATGGATATCACGGACTGATTCTGACATCATAGATGCAACAATGAAAGAACTTGAGAAACTCTTCCC
TGATGAAATCTCAGCTGACCAAAGCAAAGCTAAAATTCTGAAGTACCATGTCGTTAAGACTCCAAGATCTG
TGTACAAGACCATCCCAAACTGTGAACCATGTCGTCCTCTACAAAGATCACCTATTGAAGGATTCTACTTA
GCTGGAGATTACACAAAACAGAAGTACTTAGCTTCCATGGAAGGCGCTGTCCTCTCTGGCAAATTCTGCTC
TCAGTCTATTGTTCAGGATTACGAGCTACTGGCTGCGTCTGGACCAAGAAAGTTGTCGGAGGCAACAGTAT
CATCATCATGAGAAGAGGACAAAACTTAAAGATGATTTGCTTGTAAGCATTATTATTTGTGTATAAATCTC
ATTGCAATCCAAACTTAACCTTACTCTCTTCAGTAAATGAATCTCACAGATTTGACATCTCACGTTTCTGT
CAATTTTATAATTTTTAAAAAGTAATTACTGTCGACCTTTTGTAATCATAGTGATTTATCATTATGTCTCT
CTTTTTAAAACC

First underlined sequence indicates target site for dTALEN.*PDS*3.1F and homodimer; second underlined sequence indicates target site for dTALEN.*PDS*3.1R; third underlined sequence indicates target for dTALEN.*PDS*3.2F; fourth underlined sequence indicates target for dTALEN.*PDS*3.2R.

Supplementary Table 2. TALE-TF/TALEN, RVD sequences and their corresponding EBE sequences.

TALE-TF/TALEN Name	RVD Sequence	EBE Sequence
dTALEN.*PDS*3.1F	HD HD HD NI HD NG NG HD NG HD NI NI	TCCCACTTCTCAA
dTALEN.*PDS*3.1R	NI HD NG HD HD NG HD HD NG NG NG NG	TACTCCTCCTCCT
dTALEN.*PDS*3.2F	HD HD NG NG NG NG NI HD HD NI NN NI	TCCTTTTACCAGA
dTALEN.*PDS*3.2R	HD HD NI NI HD NG NI NI NG NG NG NI	TCCAACTAATTTA
dTALEN.*PDS*3.homodimer	HD HD NS NS HD NG NS HD NG NS NN NS	TCCCACTTCTCAA
dAvr*Bs3*	NI NG NI NG NI NI NI HD HD NG NI NI	TATATAAACCTAA
dHax3	HD HD HD NG NG NG NI NG HD NG HD NG	TCCCTTTATCTCT
dTALE.*PDS*3.2F	NG NG NI NG HD HD NG NG NG NG NI HD	TTTATCCTTTTAC
dTALE.*PDS*3.2R	HD HD NI NI HD NG NI NI NG NG NG NI	TCCAACTAATTTA
dTALE.TYLCV.2F	HD HD NK NG NI NG NI NI NG NI NG NG	TCCGTATAATATT

Supplementary Table 3. Primer for sequence confirmation of assembled fragments.

Primer name	Primer sequence in 5' to 3' orientation	Primer purpose
Rep.Seqconf.F1	GGCGCAAGAGCATTAGAGGCGCTTCTCACCGTA	sequence confirmation of assembled clones
Rep.Seqconf.R1	TTTAGCTCTTTTCATTCCAGAGGCTTGGAGAATTCTATC	sequence confirmation of assembled clones
Rep.Seqconf.R2	TTTCTTAACGGCATCAAGTGCTGGTCGTCCTCCTAAACATGC	sequence confirmation of assembled clones

6.4 Supplementary Information for Chapter 4

Name	Description
Supplementary Sequence 1	DNA sequence of dCas9 construct in 44246 original vector.
Supplementary Sequence 2	DNA sequence of the *Mlu*I-dCas9:EDLL-*Eco*RI construct designed for this study.
Supplementary Sequence 3	DNA sequence of the *Mlu*I-dCas9:TAD-*Xho*I construct designed for this study.
Supplementary Sequence 4	DNA sequence of the *Mlu*I-dCas9:SRDX-*Eco*RI construct designed for this study.
Supplementary Sequence 5	DNA sequence of the *Bs3* promoter construct used in this study.
Supplementary Sequence 6	DNA sequence of the gRNA 1 construct designed for this study.
Supplementary Sequence 7	DNA sequence of the gRNA 2 construct designed for this study.
Supplementary Sequence 8	DNA sequence of the gRNA 3 construct designed for this study.
Supplementary Sequence 9	DNA sequence of the gRNA 4 construct designed for this study.
Supplementary Sequence 10	DNA sequence of the gRNA 5 construct designed for this study.
Supplementary Sequence 11	DNA sequence of the gRNA 6 construct designed for this study.
Supplementary Sequence 12	DNA sequence of phytoene desaturase (*PDS*) gene with part of the promoter sequence in *Nicotiana benthamiana*.
Supplementary Figure 1	Schematic representation of possible applications for the CRISPR/dCas9 transcriptional-modulation platform.
Supplementary Table 1	List of primers used in this study.

Supplementary Sequence 1. DNA sequence of dCas9 construct in 44246 original vector.

TGAAAGACCCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGAAAATACAT

AACTGAGAATAGAGAAGTTCAGATCAAGGTTAGGAACAGAGAGACAGCAGAATATGGGCCAAACAGGATAT

CTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCCGCCCTCAG

CAGTTTCTAGAGAACCATCAGATGTTTCCAGGGTGCCCCAAGGACCTGAAATGACCCTGTGCCTTATTTGA

ACTAACCAATCAGTTCGCTTCTCGCTTCTGTTCGCGCGCTTCTGCTCCCCGAGCTCAATAAAAGAGCCCAC

AACCCCTCACTCGGCGCGCCAGTCCTCCGATAGACTGCGTCGCCCGGGTACCCGTATTCCCAATAAAGCCT

CTTGCTGTTTGCATCCGAATCGTGGACTCGCTGATCCTTGGGAGGGTCTCCTCAGATTGATTGACTGCCCA

CCTCGGGGGTCTTTCATTTGGAGGTTCCACCGAGATTTGGAGACCCCTGCCCAGGGACCACCGACCCCCCC

GCCGGGAGGTAAGCTGGCCAGCGGTCGTTTCGTGTCTGTCTCTGTCTTTGTGCGTGTTTGTGCCGGCATCT

AATGTTTGCGCCTGCGTCTGTACTAGTTAGCTAACTAGCTCTGTATCTGGCGGACCCGTGGTGGAACTGAC

GAGTTCTGAACACCCGGCCGCAACCCTGGGAGACGTCCCAGGGACTTTGGGGGCCGTTTTTGTGGCCCGAC

CTGAGGAAGGGAGTCGATGTGGAATCCGACCCCGTCAGGATATGTGGTTCTGGTAGGAGACGAGAACCTAA

AACAGTTCCCGCCTCCGTCTGAATTTTTGCTTTCGGTTTGGAACCGAAGCCGCGCGTCTTGTCTGCTGCAG

CGCTGCAGCATCGTTCTGTGTTGTCTCTGTCTGACTGTGTTTCTGTATTTGTCTGAAAATTAGGGCCAGAC

TGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAAAGATGTCGAGCGGATCGCTCACAACCAGTCGG

TAGATGTCAAGAAGAGACGTTGGGTTACCTTCTGCTCTGCAGAATGGCCAACCTTTAACGTCGGATGGCCG

CGAGACGGCACCTTTAACCGAGACCTCATCACCCAGGTTAAGATCAAGGTCTTTTCACCTGGCCCGCATGG

ACACCCAGACCAGGTCCCCTACATCGTGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGC

CCTTTGTACACCCTAAGCCTCCGCCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGAACCTCCTCGT

TCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTCACTCCTTCTCTAGGCGCCGGAATTAGATCTCGCCA

CCATGGACAAGAAGTATTCTATCGGACTGGCCATCGGGACTAATAGCGTCGGGTGGGCCGTGATCACTGAC

GAGTACAAGGTGCCCTCTAAGAAGTTCAAGGTGCTCGGGAACACCGACCGGCATTCCATCAAGAAAAATCT

GATCGGAGCTCTCCTCTTTGATTCAGGGGAGACCGCTGAAGCAACCCGCCTCAAGCGGACTGCTAGACGGC

GGTACACCAGGAGGAAGAACCGGATTTGTTACCTTCAAGAGATATTCTCCAACGAAATGGCAAAGGTCGAC

GACAGCTTCTTCCATAGGCTGGAAGAATCATTCCTCGTGGAAGAGGATAAGAAGCATGAACGGCATCCCAT

CTTCGGTAATATCGTCGACGAGGTGGCCTATCACGAGAAATACCCAACCATCTACCATCTTCGCAAAAAGC

TGGTGGACTCAACCGACAAGGCAGACCTCCGGCTTATCTACCTGGCCCTGGCCCACATGATCAAGTTCAGA

GGCCACTTCCTGATCGAGGGCGACCTCAATCCTGACAATAGCGATGTGGATAAACTGTTCATCCAGCTGGT

GCAGACTTACAACCAGCTCTTTGAAGAGAACCCCATCAATGCAAGCGGAGTCGATGCCAAGGCCATTCTGT

CAGCCCGGCTGTCAAAGAGCCGCAGACTTGAGAATCTTATCGCTCAGCTGCCGGGTGAAAAGAAAAATGGA

CTGTTCGGGAACCTGATTGCTCTTTCACTTGGGCTGACTCCCAATTTCAAGTCTAATTTCGACCTGGCAGA

GGATGCCAAGCTGCAACTGTCCAAGGACACCTATGATGACGATCTCGACAACCTCCTGGCCCAGATCGGTG

ACCAATACGCCGACCTTTTCCTTGCTGCTAAGAATCTTTCTGACGCCATCCTGCTGTCTGACATTCTCCGC

GTGAACACTGAAATCACCAAGGCCCCTCTTTCAGCTTCAATGATTAAGCGGTATGATGAGCACCACCAGGA

CCTGACCCTGCTTAAGGCACTCGTCCGGCAGCAGCTTCCGGAGAAGTACAAGGAAATCTTCTTTGACCAGT

CAAAGAATGGATACGCCGGCTACATCGACGGAGGTGCCTCCCAAGAGGAATTTTATAAGTTTATCAAACCT

ATCCTTGAGAAGATGGACGGCACCGAAGAGCTCCTCGTGAAACTGAATCGGGAGGATCTGCTGCGGAAGCA

GCGCACTTTCGACAATGGGAGCATTCCCCACCAGATCCATCTTGGGGAGCTTCACGCCATCCTTCGGCGCC

AAGAGGACTTCTACCCCTTTCTTAAGGACAACAGGGAGAAGATTGAGAAAATTCTCACTTTCCGCATCCCC

TACTACGTGGGACCCCTCGCCAGAGGAAATAGCCGGTTTGCTTGGATGACCAGAAAGTCAGAAGAAACTAT

CACTCCCTGGAACTTCGAAGAGGTGGTGGACAAGGGAGCCAGCGCTCAGTCATTCATCGAACGGATGACTA

ACTTCGATAAGAACCTCCCCAATGAGAAGGTCCTGCCGAAACATTCCCTGCTCTACGAGTACTTTACCGTG

TACAACGAGCTGACCAAGGTGAAATATGTCACCGAAGGGATGAGGAAGCCCGCATTCCTGTCAGGCGAACA

AAAGAAGGCAATTGTGGACCTTCTGTTCAAGACCAATAGAAAGGTGACCGTGAAGCAGCTGAAGGAGGACT

ATTTCAAGAAAATTGAATGCTTCGACTCTGTGGAGATTAGCGGGGTCGAAGATCGGTTCAACGCAAGCCTG

GGTACCTACCATGATCTGCTTAAGATCATCAAGGACAAGGATTTTCTGGACAATGAGGAGAACGAGGACAT

CCTTGAGGACATTGTCCTGACTCTCACTCTGTTCGAGGACCGGGAAATGATCGAGGAGAGGCTTAAGACCT

ACGCCCATCTGTTCGACGATAAAGTGATGAAGCAACTTAAACGGAGAAGATATACCGGATGGGGACGCCTT

AGCCGCAAACTCATCAACGGAATCCGGGACAAACAGAGCGGAAAGACCATTCTTGATTTCCTTAAGAGCGA

CGGATTCGCTAATCGCAACTTCATGCAACTTATCCATGATGATTCCCTGACCTTTAAGGAGGACATCCAGA

AGGCCCAAGTGTCTGGACAAGGTGACTCACTGCACGAGCATATCGCAAATCTGGCTGGTTCACCCGCTATT

AAGAAGGGTATTCTCCAGACCGTGAAAGTCGTGGACGAGCTGGTCAAGGTGATGGGTCGCCATAAACCAGA

GAACATTGTCATCGAGATGGCCAGGGAAAACCAGACTACCCAGAAGGGACAGAAGAACAGCAGGGAGCGGA

TGAAAAGAATTGAGGAAGGGATTAAGGAGCTCGGGTCACAGATCCTTAAAGAGCACCCGGTGGAAAACACC

CAGCTTCAGAATGAGAAGCTCTATCTGTACTACCTTCAAAATGGACGCGATATGTATGTGGACCAAGAGCT

TGATATCAACAGGCTCTCAGACTACGACGTGGACGCCATCGTCCCTCAGAGCTTCCTCAAAGACGACTCAA

TTGACAATAAGGTGCTGACTCGCTCAGACAAGAACCGGGGAAAGTCAGATAACGTGCCCTCAGAGGAAGTC

GTGAAAAAGATGAAGAACTATTGGCGCCAGCTTCTGAACGCAAAGCTGATCACTCAGCGGAAGTTCGACAA

TCTCACTAAGGCTGAGAGGGGCGGACTGAGCGAACTGGACAAAGCAGGATTCATTAAACGGCAACTTGTGG

AGACTCGGCAGATTACTAAACATGTCGCCCAAATCCTTGACTCACGCATGAATACCAAGTACGACGAAAAC

GACAAACTTATCCGCGAGGTGAAGGTGATTACCCTGAAGTCCAAGCTGGTCAGCGATTTCAGAAAGGACTT

TCAATTCTACAAAGTGCGGGAGATCAATAACTATCATCATGCTCATGACGCATATCTGAATGCCGTGGTGG

GAACCGCCCTGATCAAGAAGTACCCAAAGCTGGAAAGCGAGTTCGTGTACGGAGACTACAAGGTCTACGAC

GTGCGCAAGATGATTGCCAAATCTGAGCAGGAGATCGGAAAGGCCACCGCAAAGTACTTCTTCTACAGCAA

CATCATGAATTTCTTCAAGACCGAAATCACCCTTGCAAACGGTGAGATCCGGAAGAGGCCGCTCATCGAGA

CTAATGGGGAGACTGGCGAAATCGTGTGGGACAAGGGCAGAGATTTCGCTACCGTGCGCAAAGTGCTTTCT

ATGCCTCAAGTGAACATCGTGAAGAAAACCGAGGTGCAAACCGGAGGCTTTTCTAAGGAATCAATCCTCCC

CAAGCGCAACTCCGACAAGCTCATTGCAAGGAAGAAGGATTGGGACCCTAAGAAGTACGGCGGATTCGATT

CACCAACTGTGGCTTATTCTGTCCTGGTCGTGGCTAAGGTGGAAAAAGGAAAGTCTAAGAAGCTCAAGAGC

GTGAAGGAACTGCTGGGTATCACCATTATGGAGCGCAGCTCCTTCGAGAAGAACCCAATTGACTTTCTCGA

AGCCAAAGGTTACAAGGAAGTCAAGAAGGACCTTATCATCAAGCTCCCAAAGTATAGCCTGTTCGAACTGG

AGAATGGGCGGAAGCGGATGCTCGCCTCCGCTGGCGAACTTCAGAAGGGTAATGAGCTGGCTCTCCCCTCC

AAGTACGTGAATTTCCTCTACCTTGCAAGCCATTACGAGAAGCTGAAGGGGAGCCCCGAGGACAACGAGCA

AAAGCAACTGTTTGTGGAGCAGCATAAGCATTATCTGGACGAGATCATTGAGCAGATTTCCGAGTTTTCTA

AACGCGTCATTCTCGCTGATGCCAACCTCGATAAAGTCCTTAGCGCATACAATAAGCACAGAGACAAACCA

ATTCGGGAGCAGGCTGAGAATATCATCCACCTGTTCACCCTCACCAATCTTGGTGCCCCTGCCGCATTCAA

GTACTTCGACACCACCATCGACCGGAAACGCTATACCTCCACCAAAGAAGTGCTGGACGCCACCCTCATCC

ACCAGAGCATCACCGGACTTTACGAAACTCGGATTGACCTCTCACAGCTCGGAGGGGATGAGGGAGCTGAT

CCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTATA

GAATTCTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGG

CACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCACCGGTAGGCGCCAACCGGCTC

CGTTCTTTGGTGGCCCCTTCGCGCCACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAG

CTCGCGTCGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAGTCTCGTGCAGATGGACAGCACCGC

TGAGCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGGCTC

AGAGGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGGGCGGGCGCCCGAAGG

TCCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCTGCCGCGCTGTTCTCCTCTTCCTCATC

TCCGGGCCTTTCGACCTGCAGCCCAAGCTTACCATGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCG

CGACGACGTCCCCAGGGCCGTACGCACCCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCG

TCGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCGGGCTCGAC

ATCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCGGTGGCGGTCTGGACCACGCCGGAGAGCGTCGAAGC

GGGGGCGGTGTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAAC

AGATGGAAGGCCTCCTGGCGCCGCACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGCGTCTCG

CCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGG

GGTGCCCGCCTTCCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCA

CCGCCGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGACGCCCG

CCCCACGACCCGCAGCGCCCGACCGAAAGGAGCGCACGACCCCATGCATCGATAAAATAAAAGATTTTATT

TAGTCTCCAGAAAAAGGGGGGAATGAAAGACCCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGCCAT

TTTGCAAGGCATGGAAAATACATAACTGAGAATAGAGAAGTTCAGATCAAGGTTAGGAACAGAGAGACAGC

AGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTC

CCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGAGAACCATCAGATGTTTCCAGGGTGCCCCAAGGACCTG

AAATGACCCTGTGCCTTATTTGAACTAACCAATCAGTTCGCTTCTCGCTTCTGTTCGCGCGCTTCTGCTCC

CCGAGCTCAATAAAAGAGCCCACAACCCCTCACTCGGCGCGCCAGTCCTCCGATAGACTGCGTCGCCCGGG

TACCCGTGTATCCAATAAACCCTCTTGCAGTTGCATCCGACTTGTGGTCTCGCTGTTCCTTGGGAGGGTCT

CCTCTGAGTGATTGACTACCCGTCAGCGGGGGTCTTTCATGGGTAACAGTTTCTTGAAGTTGGAGAACAAC

ATTCTGAGGGTAGGAGTCGAATATTAAGTAATCCTGACTCAATTAGCCACTGTTTTGAATCCACATACTCC

AATACTCCTGAAATAGTTCATTATGGACAGCGCAGAAGAGCTGGGGAGAATTAATTCGTAATCATGGTCAT

AGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGT

AAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC

GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGC

GCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCAC

TCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCA

GCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGC

ATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCC

CCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCC

TTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCA

AGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG

TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTA

TGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTA

TCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC

GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC

TTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGAT

TATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATAT

GAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCG

TTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCA

GTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGA

AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGC

TAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCAC

GCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG

TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATC

ACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTG

GTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATA

CGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAA

ACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG

CATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATA

AGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTA

TTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTC

CCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATC

ACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGAC

GGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCG

GGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAA

ATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTG

GGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGAT

TAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCGCAAGGAATGGTGCATGCA

AGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATG

AGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGT

GGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGCGATTAGTCCAATTTGTTAAAGACAGGATAT

CAGTGGTCCAGGCTCTAGTTTTGACTCAACAATATCACCAGCTGAAGCCTATAGAGTACGAGCCATAGATA
AAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGGAA

Supplementary Sequence 2. DNA sequence of the *Mlu*I-dCas9:EDLL-*EcoR*I construct designed for this study.

TCTAAACGCGTCATTCTCGCTGATGCCAACCTCGATAAAGTCCTTAGCGCATACAATAAGCACAGAGACAA
ACCAATTCGGGAGCAGGCTGAGAATATCATCCACCTGTTCACCCTCACCAATCTTGGTGCCCCTGCCGCAT
TCAAGTACTTCGACACCACCATCGACCGGAAACGCTATACCTCCACCAAAGAAGTGCTGGACGCCACCCTC
ATCCACCAGAGCATCACCGGACTTTACGAAACTCGGATTGACCTCTCACAGCTCGGAGGGGATGAGGGAGC
TGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGG
TAGGATCCGAAGTTTTTGAATTTGAATATCTTGATGATAAGGTTCTTGAAGAACTTCTTGATTCTGAAGAA
AGAAAGAGATGACTCGAGGAATTC

Supplementary Sequence 3. DNA sequence of the *Mlu*I-dCas9:TAD-*Xho*I construct designed for this study.

AGTTTTCTAAACGCGTCATTCTCGCTGATGCCAACCTCGATAAAGTCCTTAGCGCATACAATAAGCACAGA
GACAAACCAATTCGGGAGCAGGCTGAGAATATCATCCACCTGTTCACCCTCACCAATCTTGGTGCCCCTGC
CGCATTCAAGTACTTCGACACCACCATCGACCGGAAACGCTATACCTCCACCAAAGAAGTGCTGGACGCCA
CCCTCATCCACCAGAGCATCACCGGACTTTACGAAACTCGGATTGACCTCTCACAGCTCGGAGGGGATGAG
GGAGCTGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAG
AAAGGTATCGATAGTCGCACAACTATCACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAG
TCGCACTTGCATGTTTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCA
TTGATTAAAAGAACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATGCTCAAGTCGT
AAGAGTTTTGGGTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTTGACGATGCAATGACTCAATTTG
GAATGAGTAGACATGGACTCCTGCAATTATTTCGAAGGGTCGGAGTTACAGAGCTCGAAGCCAGGTCAGGA
ACGCTGCCCCCCGCATCTCAACGATGGGATAGAATTCTCCAAGCCTCTGGAATGAAAAGAGCTAAACCTTC
ACCAACGTCCACACAAACACCAGACCAAGCTTCTCTCCACGCTTTTGCCGACTCACTAGAGAGAGATCTAG
ATGCACCGTCACCTATGCATGAAGGAGACCAAACAAGAGCCTCTTCAAGAAAACGTTCTCGTTCTGATAGA

GCTGTCACTGGACCTTCCGCCCAACAATCTTTCGAAGTCCGAGTTCCTGAGCAACGAGATGCCCTACACCT
GCCTTTGCTTTCTTGGGGAGTTAAGCGACCACGTACTAGAATTGGTGGACTACTCGATCCAGGTACACCAA
TGGATGCTGATCTCGTTGCTTCCTCTACCGTAGTATGGGAGCAAGACGCAGACCCCTTCGCTGGAACTGCT
GACGATTTCCCAGCCTTTAACGAGGAAGAATTGGCTTGGTTAATGGAACTTCTACCGCAAAAGTAGGGATC
CGGCTCGAGC

Supplementary Sequence 4. DNA sequence of the *Mlu*I-dCas9:SRDX-*Eco*RI construct designed for this study.

TCTAAACGCGTCATTCTCGCTGATGCCAACCTCGATAAAGTCCTTAGCGCATACAATAAGCACAGAGACAA
ACCAATTCGGGAGCAGGCTGAGAATATCATCCACCTGTTCACCCTCACCAATCTTGGTGCCCCTGCCGCAT
TCAAGTACTTCGACACCACCATCGACCGGAAACGCTATACCTCCACCAAAGAAGTGCTGGACGCCACCCTC
ATCCACCAGAGCATCACCGGACTTTACGAAACTCGGATTGACCTCTCACAGCTCGGAGGGGATGAGGGAGC
TGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGG
TAGGATCCCTGGATCTGGATCTGGAACTGCGCCTGGGCTTTGCGTGATGACTCGAGGAATTC

Supplementary Sequence 5. DNA sequence of the *Bs3* promoter construct used in this study.

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTAT
AATGCCAACTTTGTACAAAAAAGCAGGCTTCATAGTCAAGCTAACGAAACTTATGCAAGGGAAATATGAAA
TTAGTATGCAAGTAAACTCAAAGAACTAATCATTGAACTGAAAGATCAATATATCAAAAAAAAAAAAAAAC
AATAAAACCGTTTAACCGATAGATTAACCATTTCTGGTTCAGTTTATGGGTTAAACCACAATTTGCACACC
CTGGTTAAACAATGAACACGTTTGCCTGACCAATTTTATCCCTTTATCTCTAACCATCCTCACAACTTCAA
GTTATCATCCCCTTTCTCTTTTCTCCTCTTGTTCTTGTCACCCGCTAAATCTATCAAAACACAAGTAGTCC
TAGTTGCACATATATTTCACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATT
TGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

Supplementary Sequence 6. DNA sequence of the gRNA 1 construct designed for this study.

GAGCTCGAATTCGGATCCAGAAATCTCAAAATTCCGGCAGAACAATTTTGAATCTCGATCCGTAGAAACGA
GACGGTCATTGTTTTAGTTCCACCACGATTATATTTGAAATTTACGTGAGTGTGAGTGAGACTTGCATAAG
AAAATAAAATCTTTAGTTGGGAAAAAATTCAATAATATAAATGGGCTTGAGAAGGAAGCGAGGGATAGGCC
TTTTTCTAAAATAGGCCCATTTAAGCTATTAACAATCTTCAAAAGTACCACAGCGCTTAGGTAAAGAAAGC
AGCTGAGTTTATATATGGTTAGAGACGAAGTAGTGATTGACCGATAGATTAACCATTTCGTTTTAGAGCTA
GAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT
AGACCCAGCTTTCTTGTACAAAGTTGGCATTATCTAGAAAGCTTGAGCTC

Supplementary Sequence 7. DNA sequence of the gRNA 2 construct designed for this study.

GAGCTCGAATTCGGATCCAGAAATCTCAAAATTCCGGCAGAACAATTTTGAATCTCGATCCGTAGAAACGA
GACGGTCATTGTTTTAGTTCCACCACGATTATATTTGAAATTTACGTGAGTGTGAGTGAGACTTGCATAAG
AAAATAAAATCTTTAGTTGGGAAAAAATTCAATAATATAAATGGGCTTGAGAAGGAAGCGAGGGATAGGCC
TTTTTCTAAAATAGGCCCATTTAAGCTATTAACAATCTTCAAAAGTACCACAGCGCTTAGGTAAAGAAAGC
AGCTGAGTTTATATATGGTTAGAGACGAAGTAGTGATTGAAACCACAATTTGCACACCCGTTTTAGAGCTA
GAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT
AGACCCAGCTTTCTTGTACAAAGTTGGCATTATCTAGAAAGCTTGAGCTC

Supplementary Sequence 8. DNA sequence of the gRNA 3 construct designed for this study.

GAGCTCGAATTCGGATCCAGAAATCTCAAAATTCCGGCAGAACAATTTTGAATCTCGATCCGTAGAAACGA
GACGGTCATTGTTTTAGTTCCACCACGATTATATTTGAAATTTACGTGAGTGTGAGTGAGACTTGCATAAG
AAAATAAAATCTTTAGTTGGGAAAAAATTCAATAATATAAATGGGCTTGAGAAGGAAGCGAGGGATAGGCC
TTTTTCTAAAATAGGCCCATTTAAGCTATTAACAATCTTCAAAAGTACCACAGCGCTTAGGTAAAGAAAGC
AGCTGAGTTTATATATGGTTAGAGACGAAGTAGTGATTGACGTGTTCATTGTTTAACCAGTTTTAGAGCTA
GAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT
AGACCCAGCTTTCTTGTACAAAGTTGGCATTATCTAGAAAGCTTGAGCTC

Supplementary Sequence 9. DNA sequence of the gRNA 4 construct designed for this study.

GAGCTCGAATTCGGATCCAGAAATCTCAAAATTCCGGCAGAACAATTTTGAATCTCGATCCGTAGAAACGA
GACGGTCATTGTTTTAGTTCCACCACGATTATATTTGAAATTTACGTGAGTGTGAGTGAGACTTGCATAAG
AAAATAAAATCTTTAGTTGGGAAAAAATTCAATAATATAAATGGGCTTGAGAAGGAAGCGAGGGATAGGCC
TTTTTCTAAAATAGGCCCATTTAAGCTATTAACAATCTTCAAAAGTACCACAGCGCTTAGGTAAAGAAAGC
AGCTGAGTTTATATATGGTTAGAGACGAAGTAGTGATTGTGGCCAAACCACCAAATTCGTTTTAGAGCTAG
AAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTA
GACCCAGCTTTCTTGTACAAAGTTGGCATTATCTAGAAAGCTTGAGCTC

Supplementary Sequence 10. DNA sequence of the gRNA 5 construct designed for this study.

GAGCTCGAATTCGGATCCAGAAATCTCAAAATTCCGGCAGAACAATTTTGAATCTCGATCCGTAGAAACGA
GACGGTCATTGTTTTAGTTCCACCACGATTATATTTGAAATTTACGTGAGTGTGAGTGAGACTTGCATAAG
AAAATAAAATCTTTAGTTGGGAAAAAATTCAATAATATAAATGGGCTTGAGAAGGAAGCGAGGGATAGGCC
TTTTTCTAAAATAGGCCCATTTAAGCTATTAACAATCTTCAAAAGTACCACAGCGCTTAGGTAAAGAAAGC
AGCTGAGTTTATATATGGTTAGAGACGAAGTAGTGATTGTATAAAAAACTAATATTTGGGTTTTAGAGCTA
GAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT
AGACCCAGCTTTCTTGTACAAAGTTGGCATTATCTAGAAAGCTTGAGCTC

Supplementary Sequence 11. DNA sequence of the gRNA 6 construct designed for this study.

GAGCTCGAATTCGGATCCAGAAATCTCAAAATTCCGGCAGAACAATTTTGAATCTCGATCCGTAGAAACGA
GACGGTCATTGTTTTAGTTCCACCACGATTATATTTGAAATTTACGTGAGTGTGAGTGAGACTTGCATAAG
AAAATAAAATCTTTAGTTGGGAAAAAATTCAATAATATAAATGGGCTTGAGAAGGAAGCGAGGGATAGGCC
TTTTTCTAAAATAGGCCCATTTAAGCTATTAACAATCTTCAAAAGTACCACAGCGCTTAGGTAAAGAAAGC
AGCTGAGTTTATATATGGTTAGAGACGAAGTAGTGATTGGGTATCTTTTTGTGGGTAAGTTTTAGAGCTAG

AAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTA
GACCCAGCTTTCTTGTACAAAGTTGGCATTATCTAGAAAGCTTGAGCTC

Supplementary Sequence 12. DNA sequence of phytoene desaturase (*PDS*) gene with part of the promoter sequence in *Nicotiana benthamiana*.

TAACGAAAACATGGAATATGCGGGTAATAACCAAACAAGGCACAATAATACTAATAAGGTTTGATATATGG
CATATTTAGGTAAAAATCCCTTTTAATAAAAGTATCAGGACGAGAAACAAAAAAAGAAGATTACTACTACT
AGAAAACCACGGGTCAGGTTGGATTACGTGGATCCTCTATGACCCAGTAACCCATGTGGGAGATGGGAGCA
AAGTGGTCAAACTTTAGAAGGAATGAGCAAAGCAAGAAATTAAAAAGAGAGAGAGGTGCTTTATCCATCAA
ATGTGGCTATGGTAGGAAGAGCCAATGGTGGGACATTTTTGGAGTGTAGCCAAAACATAAAGGAAGGTCCA
GTGCGAGTTACTGCAAATTGAGTTGGGAGTGAGGATTAAAGAAGATAGTAACATATTTCTAGCTAAATAGC
AAACAAATGATCCGTTAACAGAAGTGGCCAAACCACCAAATTCAGGCATCTCCACCAAATATTAGTTTTTT
ATACACAAAAGATTCAACACAAACAGTTAAGTACTTCTTTAATCGTTCCTAATTCTTTGTTCAGGGGTATC
TTTTTGTGGGTAACGGCCAAACCACCACAAATTTTCAGTTCCCACTCTTAACTCTTTCAACTTCAACACAA
CAAATTAGTATTTGCTTTTCCTTCTTTGCTTATCTAGTGCATAACGATTTTCTACAACTTTAGCATAGTCC
ACAACGTGAAACACAACTCCTTGGCGGTTTATACCGAGGTAAGAAATGATTTTGGTTTCTTTGGTTACATC
AGCTGAATGCTTTGCTTGAGAAAAGCTCTCTTTTTCCCGTTTAGGATCTTGTTTATTTGCTTTCGTTTTTC
TACTCGTTTGAATTTTAACTTGATTTTGTGGGTGAAGGCTAATTTTTCTCATAGTGTAAGAACAAGTTTCA
TATGTACTGTAAAAGCTAGAATCTTTTTTACTTTTGCATATAAATTTGTGTAATAAATGCTTAAGAACCAG
AATATTTGAAAAAGATAAGGAATTTTGCATAGTATTTAGGTTCACAAGTGGGACAATCTTCTTACACTGAA
ATATCTTTATGTCAGGCTTAATTTACTGCTATCTTGTTCAATAAAATGCCCCAAATTGGACTTGTTTCTGC
CGTTAATTTGAGAGTCCAAGGTAATTCAGCTTATCTTTGGAGCTCGAGGTCTTCGTTGGGAACTGAAAGTC
AAGATGTTTGCTTGCAAAGGAATTTGTTATGTTTTGGTAGTAGCGACTCCATGGGGCATAAGTTAAGGATT
CGTACTCCAAGTGCCACGACCCGAAGATTGACAAAGGACTTTAATCCTTTAAAGGTTTGTTTTGAATGCGA
AAGTGTGATGCTGGATTTATGATCGTGGGCATATATCCTCTAAAATAAGAGATGTATATCTTGCCATTCAG
GTAGTCTGCATTGATTATCCAAGACCAGAGCTAGACAATACAGTTAACTATTTGGAGGCGGCGTTATTATC
ATCATCGTTTCGTACTTCCTCACGCCCAACTAAACCATTGGAGATTGTTATTGCTGGTGCAGGTGATTTTT
TCCAGCCATCTATATTTGTAGTTTTCATTTTTCTTTCTTTGGAAGGAAGATCATTCTATTAGTTATATTAT

CACTAGAATATTTACCTGTACATTCTTTTCTGATTAACTGTTTTGGACCGCAAAATTTTAGGTTCTTACTT

CTTCGCCATTTTGCAACTAATCAGCAATTAGGAGCGGTTTGAAAACTAGTTTGTTTTGAACTATTTTTGCC

GTCACTCTATTTATATACTGTTGAATTGTCCCAAATCGGTGGAATTTGAGGTCCTTGGTCTCATCTCATAA

GCTAGCTTTTGGGGTTGAGTTACCACATCGGTGGGATTTGAGGTCCTTCGTCTCCTTATATGTTCTTGGAC

AAGCTTCACCTCATAAGCTAGCTTTTGGGGTTAGAGTTAGGCCCAAGGTCCATTTATCATATGCTTGTCTA

TTCTCTCTTATCATCTGAGCCATGATAAGCGGGTGAACGTGCTGTCTATTGGGTGGCATGTCCAAAGGATC

ATTCTGAAATATTGGAGGCAAATGAACCAATACCTTGTGCAAGATTGATCTCACTATACCTATAATCAGAG

TACTGAGTTCCAAAAATTTCAAAACCCATTGAAAAGTCAAACGAGTTACATATAGGGGTTGCACTCTTCTA

CGGCTTGCAATATGTGAGAAAAAGATGAGAAGTCGATCTTCATATTTCATCTTTACTAGGCTGGACCATTG

ACTGGTTAGCAGTTTTGAACTTGTTCTTCAACTTGGCTTGCATGGTACTGTGCCGATCATTTCTTTTGTAT

TGTCATCAGCTGGTTGATTATCTGAGTACCTAAAGAAAGAATGTTATATGCATGATATATTCTACTGTACT

ATAAAAGATATAATAAAGAATGCTAGCCGAGGTACTGCATGGCCTTTTCAGATAAATAGAAGCTGTAGCAT

GATTCTAATTCAATTTTTTTGGGAATATCAGGTTTGGGTGGTTTGTCTACAGCAAAATATCTGGCAGATGC

TGGTCACAAACCGATATTGCTGGAGGCAAGAGATGTCCTAGGTGGGAAGGTGAAGAATATCCAATCTTTCC

TTTAATTTTATTCCTTTTTCTTTTGTGTCCTTCCCTATTGATAGTCCCTTTTCAGGAAGGCTTCTGTTTGT

TTTATTTGAAATCATTTTTCATACTCTTTAAGCATTCAGTTGCTCAAACAATTGCAAGGATATTCACTATT

CCTAATTTTGACCGTCTTCTTTTCTCTCAGTTTAGTTTTATTCCCCTCTCTTTTTGAAGGAAATAGATCTG

TCCTAAAAATTTCCAGCTTTACTACTAATAGTGTTAATTGTCGATAAAATAGTACATCATATTAGGTAAAA

GATATGGACTGTATATTATTATCATTCTCTATTATTTTAAACTGAGTCAATTTTAACCGTCCTGTTGGGTG

CATTTCTCATATAAACAGTCTTTTCTGTGAGATGCTATGTGAATTAGCTGATTGTTTTGGTATAGAGCACT

ATGTTAGTCAGTTTTATCTTACTGAAGCAGTCACCAAGAGTCTAGTTGTATAGGCTAGAAGATTGAATTAG

CATTAATCTTTATGTGTTCTGCACCTGAATACTTGTACCTCCCTTTTAGGTAGCTGCATGGAAAGATGATG

ATGGAGATTGGTACGAGACTGGGTTGCACATATTCTGTAAGTTTGACTCCTCAAGAATGCTACTTTAATCT

TCTAATACAGTCATAGCAATTTCTTTCAAGATCTCTTTTATTAATCAGATAGCTATCCCTGTTTGTCTTTT

GTCTTTTGCAAATAGCCAATTTTTGTCAGTCGATCTGTATTCTGCCTTGCCTCTCTTTATTTATCTGCTAA

CTCGTATGGTGACTCATACAAGTTGGTGCATCTCCTTTAAGTTGGGGCTTACCCAAATATGCAGAACCTGT

TTGGAGAACTAGGGATTGATGATCGGTTGCAGTGGAAGGAACATTCAATGATATTTGCGATGCCTAACAAG

CCAGGGGAGTTCAGCCGCTTTGATTTTCCTGAAGCTCTTCCTGCGCCATTAAATGGTAAGTACTTAATCAT

GAGTAAATTTCTCCCTTCAGCGTTGATTATGCAAACTTCCCCAATAAGGTATGAAATTGATTAGTCTTAAT

ACCCTGGCACATTGCTAACATCAAAAGAACATAAAGGTTCATTACGTCTTGATCAGAATTTCTGCATGTAG

CTAAAGTGAATGAGTGTCTGTATAGATTTTTACACATTGCAAGCATAAGCCTGTTATGTTATCTCTTTTTT

TCATTTCTCTACCTGTATCTCTTATTCTCATTTCTCTATCTATGCGTTATTACTTCTACAGGAATTTTGGC

CATACTAAAGAACAACGAAATGCTTACGTGGCCCGAGAAAGTCAAATTTGCTATTGGACTCTTGCCAGCAA

TGCTTGGAGGGCAATCTTATGTTGAAGCTCAAGACGGTTTAAGTGTTAAGGACTGGATGAGAAAGCAAGTA

TGTGATCGTTTTATCTTATTCTTTAAAGTTCATAACCTTGAGGACATAGTTGACTTGCATATTGTTGATTT

AACATGTTCGAATTGTCTACCTGCCTTTCTTTTTCTAACAACATAGATCTTACAATCTCAGCAGCAGCTAT

TTGCTTAATGCTTTTCAGGGTGTGCCTGATAGGGTGACAGATGAGGTGTTCATTGCCATGTCAAAGGCACT

TAACTTCATAAACCCTGACGAGCTTTCGATGCAGTGCATTTTGATTGCTTTGAACAGATTTCTTCAGGTTA

GAATCCTGATCCACCCTCAAAACAAAAAGAGAGAAAGGGATATAATCCTACCAAAGCTGTAAATCATGTTA

GGGACCTGACATATCGGTGCAGGAAACTTATGAGTGAACTTGTCCACTCTGTTTAACTTTTCTGATATATT

TGAATTATTAATCTGCAGGAGAAACATGGTTCAAAAATGGCCTTTTTAGATGGTAACCCTCCTGAGAGACT

TTGCATGCCGATTGTGGAACATATTGAGTCAAAAGGTGGCCAAGTCAGACTAAACTCACGAATAAAAAAGA

TCGAGCTGAATGAGGATGGAAGTGTCAAATGTTTTATACTGAATAATGGCAGTACAATTAAAGGAGATGCT

TTTGTGTTTGCCACTCCAGGTATAATATCCATTATACTAGTATCGATGCTTCCAGTTTTCACATTTTTAGT

ATGAGTACAATTAAAGGAGATGCTTTTGTGTTTGCCACTCCAGGTATAATATCCATTACACTAGTATGACG

CTTCCAGTTTTCACATTTTAATATGAATTTATAGTTTTTTGCTGACTTTTGATTATCCAATTAGTGGATAT

CTTGAAGCTTCTTTTGCCTGAAGACTGGAAAGAGATCCCATATTTCCAAAAGTTGGAGAAGCTAGTGGGAG

TTCCTGTGATAAATGTCCATATATGGTTAGTGATGAAAATTTTGCTTTTCAGTGTTTGGTCTTCCTCTAGC

ATATCTATGTATGTGCATGTTAATGTCTATACGTACATGTTTATGTGGTCCTCCCGTATTGTGTTTACTTC

CCTTGAATGAGGAACTTATGGATGTACGCTTTTCCAACTTTGATTGTACACATTGCAATTGTCTGTTCAAC

TTTGAGGAGCAGAACTTCCATTGTTTAGCTATTAGTGGCTGAGATTCCTGCTGAAAAGATTTGTATAAATT

TAATTTGCAGGTTTGACAGAAAACTGAAGAACACATCTGATAATCTGCTCTTCAGCAGGTTCATTTTTGAT

CAATTTTATTGTTCCAGACCAGTTTCTGCGTGTCCATGACTACATTCTCATATTAGCTCCCCCCCCCCCCC

CCCCCCCCCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCCCG

GTCTCTTTTTTGCCATTTAAATGAGACCTTACAATTTGTTTAGTACTCTACCATAGTTTTTTAATCAATAA

GCCAAAGGGGAAAAACTAATAAAAGTGTATAAAATTTCTTCCTGTATTAGTCCAATTCTTTCGCAACTTAT

ATTGTTAATTATTATTTATCTTTTGGATTGAAATGGATTTTGTATATCTAATAATATAAACAAATATATCT

CTTCCTCTTATAAGATTTTTCACCATAGAAAAATGCTCCCATAAGGTCAGTCATTCTGGCTAAATATCCCA

CACTTCAACCATTGAGATATTTTGTTCTTTGCATCCAGGAATACATTTGGCATCAATAGATAGGAATCAAT

GAAGATATATTATCAATTTCCTGCAAGTTTCTTGGCACTAGAAACATTAGATCCATATCATGTAAATTGCC

TTTGTTAAATTGAAGGTCTATGAAATTTGGGTTGGTTTGAAAACCTTTTGTTTTTCCCCCCCACATCCCTA

ATCGTTTATTTAGTCAAGGTCAGACCTGACATGTTATGATGACCATTTCTCCAAGGCATTTATAATGGACT

GGAGTATCCATGCCACATTTCATCAGCTACATGTCGATTATGTTCCCCTACTTTTAAATGGCACCATTGTT

GGTGGAGCAAGATTATAGATTTTCCTGATACTTGTATGGGTTCCCTTGCTCAATCTCTCTTTTACTTCATG

CAGAAGCCCGTTGCTCAGTGTGTACGCTGACATGTCTGTTACATGTAAGGTATTGACTCGTCTGTACCATT

ATACTGGTCTAATCTGTTGGGTATGAGTTGCTGGTAAATTGCATAATGCTTGTTGGATTTGTGTGTGAGTT

GCTGCTAGATCTATGTCCTGCTATATTTATGTATGAGTTGCTGTTGTTGCAATCTTCATTTCGAATGCATA

ATGATATAGGTTCTGTATGTACGGAATAGTCAGGACAATGCTCCTGTCTGTGCACGGGGGCTCTACAGGAA

GCAACTTTCGGAGGAGAAGTACAGAAAGTGTGATGAATCTTAAGGCGGTTAAAGTAGTTTCTTTTAGCTAA

ATTTTGAAATAATTTGAAGGAGGGGAAAACGCTCTCTCAGTCTGTGGTTGCATTGGTTGTGGGNNNNNNN

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGGGACCTTAATTCAG

TGTTACCTGCATATAAATCAGACTAAAGCCTGGAGATCAGACGTTCTGCATATAAATAGATAATTAATAAT

GATCTCGTAATACTCTAAAGCCTGGAGATCAGACTGTTTTAACTATCCTGAGATGATTACTTTTACTCTCG

GATTAGCTTAGGCGAGCTGCAAGACTACATCGAATCTTTAGAAATGGGAACATAAAAAAGGTGCGAAGTGG

GGAAGTGGCTGAACAATAGGCATATGTGAGTGAGTGGGGAGTAAAATTACTTCCTTTACTTGGGTACAGTC

AAGAATGGATGACAGCTTAGCCCACTATATCTGTTCATGTGTTCTTTAGGGTCCTCTGATATAACTGGTCT

CTCTGCAGGAATATTACAACCCCAATCAGTCTATGTTGGAATTGGTATTTGCACCCGCAGAAGAGTGGATA

AATCGTAGTGACTCAGAAATTATTGATGCTACAATGAAGGAACTAGCGAAGCTTTTCCCTGATGAAATTTC

GGCAGATCAGAGCAAAGCAAAAATATTGAAGTATCATGTTGTCAAAACCCCAAGGTCAGTAATCATTTTGC

TTTCATAGTTGTGTAGTATGCGAGAATTACTGTCCACGTGGAATCTATTCCTGTTATGAATCCTGATTAAT

CTGCTTTTTACTTTCAGGTCTGTTTATAAAACTGTGCCAGGTTGTGAACCCTGTCGGCCCTTGCAAAGATC

CCCTATAGAGGGTTTTTATTTAGCTGGTGACTACACGAAACAGAAGTACTTGGCTTCAATGGAAGGTGCTG

TCTTATCAGGAAAGCTTTGTGCCGAAGCTATTGTACAGGTTAGCTCTCACATTTTTTTCCCTTCCATTGAT

AGTGTATTTGATTATATTTTGTCATCTTTGCTGCGGTAGAGAATTTTAGAAGCATTTCTCAGACATTAGTT

AGCAGAGTTACTCAGGATATCTGCAGTTTTGGAGCTTCAGTAGTAGCATGATAAAATGCAGAGGATTGTGT

TTTTTCATTCTTTATTAAACCTTGTGCCAAAGGTCTTTTGGAAACAACCTCTCTACCCCGAGGTAGGGGTA

AGGTCTGCGTACATATTACCCTCCCCATACCCCATGCGTGGGATTATACTGGGTGGTTGTTGTATAAACCT

ATATCTCTATAATTTGCAGGATTACGAGTTACTTCTTGGCCGGAGCCAGAAGATGTTGGCAGAAGCAAGCG

TAGTTAGCATAGTGAACTAAAATGTTAATTCTGTACACAAAATTTAAGATGAAGGCGGCCACGCTGAATTA

GCGTTGTACACAACTTATACAAGCACAGTACAACATTGAAACCAAATACGAGAAATGTTACACAAATATGT

GCTGCTTTCCCTCCGATTTAGTTCACAAGTTACGGACTAATTATAAGATGGAATTGTATGCAAATTGATTC

ATTCAAAACCAAACTTTTAAGCGTCAGTTATACTAGCAATACCTATAAGACATTAAACCTTCACGTCTCAA

GACATCAAAACTCGCTTTTAGAGTAAGTATCCAACTGAACCATGCAAAACACAAAACTGAAACAACGAACC

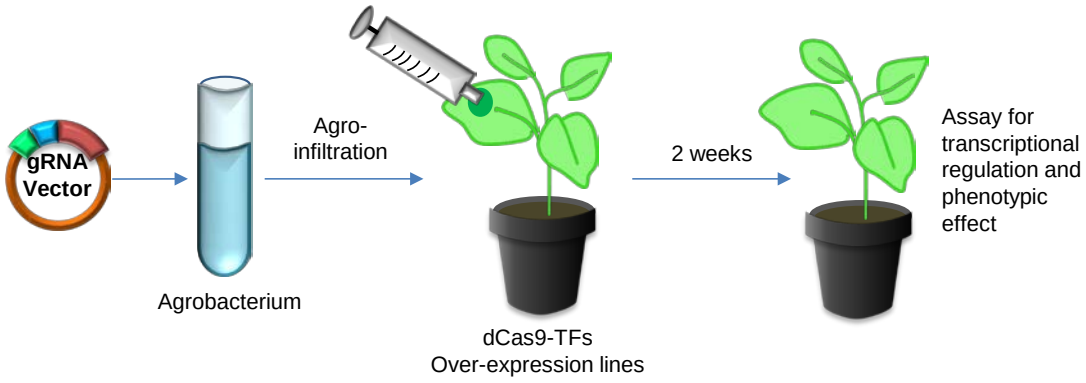
AAATGGAACTACTCAAATTTAAAAAGAAAGAGAATTAAAAATTTTAATTTCGTCAACTATTAAAATGTGCT

TGCAAGTATGGGATTTTTTTTTGAAATCCCCAAAAAACCTGTAGCCGCGACAACCTTAGCCTTTCGAGTGA

GCACTTTGTGCTCATTGGATAAACCTCCCATTGTGTAATAGCTTGCAAACTACACACGGGATATAAATTGT

ACTAGGCAAGCCCTGTGCGACAGGCTCGACCTAGAAGATATTGAGGGAATCAATCCCAGGTTGTCTG

Supplementary Figure 2. Schematic representation of possible applications for the CRISPR/dCas9 transcriptional-modulation platform. A viral vector harboring a gRNA backbone and 20 nucleotide target sequence, is transformed into *Agrobacterium tumefaciens* and co-infiltrated into an over-expression plant line of dCas9 or dCas9-TF. The virus vector is capable of movement from cell to cell, mediating the expression of the gRNA molecules. The systemic infection is rapid and effective, and the plants can be assayed for the modification in the leaves that are systemically infected just after few days after initial infiltration. Depending on the application, plants can be analyzed, for example, for the transcriptional upregulation or downregulation of specific genes or gene families in response to environmental changes. Such viral delivery platform has also potential to function as a trait discovery platform where multiple gRNAs could be delivered to the targeted plant species, overexpressing dCas9:TF, to interrogate genome wide gene functions.



Supplementary Table 4. List of primers used in this study.

Primer name	Primer sequence in 5' to 3' orientation	Primer purpose
dCas9-F	CACCATGGACAAGAAGTATTCTATCGGACTGGCCA	dCas9 amplification
dCas9-R	CTATACCTTTCTCTTCTTTTTTGGATCTACCTTTCTC	dCas9 amplification
dCas9-F1	CACCATGGACAAGAAGTATTCTATCGGACTGGCCA	dCas9 sequencing
dCas9-F2	GAGGATCTGCTGCGGAAGCAGCGCACTTTCGA	dCas9 sequencing
dCas9-F3	GAGCTCGGGTCACAGATCCTTAAAGAGCA	dCas9 sequencing
dCas9-F4	GGAAGTCAAGAAGGACCTTATCATCAAGCT	dCas9 sequencing
dCas9-R1	TGGGGAATGCTCCCATTGTCGAAAGTGCGCT	dCas9 sequencing
dCas9-R2	AGCTGGGTGTTTTCCACCGGGTGCTC	dCas9 sequencing

dCas9-R3	TCCAGTTCGAACAGGCTATACTTTGGGAGCTT	dCas9 sequencing
gRNA_F	GAGCTCGAATTCGGATCCAGAAATCTCA	gRNA amplification and sequencing
gRNA_R	GAGCTCAAGCTTTCTAGATAATGCCAAC	gRNA amplification and sequencing
qRT-GUS-F	ATACCGAAAGGTTGGGCAGG	qRT-PCR GUS amplification
qRT-GUS-R	CGGCAATAACATACGGCGTG	qRT-PCR GUS amplification
qRT-NB-*PDS*-F	CGGTTGCAGTGGAAGGAACA	qRT-PCR *PDS* amplification
qRT-NB-*PDS*-R	ATGGCGCAGGAAGAGCTTCAG	qRT-PCR *PDS* amplification
N.B.Actin1-F	TGAAGATCCTCACAGAGCGTGG	*N. benthamiana* actin1 amplification
N.B.Actin1-R	TTGTATGTGGTCTCGTGGATTC	*N. benthamiana* actin1 amplification

6.5 Supplementary Information for Chapter 5

Name	Description
Supplementary Sequence 1	DNA sequence of Cas9 gene under ubiquitin promoter
Supplementary Sequence 2	SR target sequences in polycistronic tRNA-gRNA (PTG) system
Supplementary Table 1	List of primers used in Chapter 5

Supplementary sequence 1. DNA sequence of Cas9 gene under ubiquitin promoter.

TGCATGCCTGCAGGTCCACAAATTCGGGTCAAGGCGGAAGCCAGCGCGCCACCCCACGTCAGCAAATACGG
AGGCGCGGGGTTGACGGCGTCACCCGGTCCTAACGGCGACCAACAAACCAGCCAGAAGAAATTACAGTAAA
AAAAAAGTAAATTGCACTTTGATCCACCTTTTATTACCTAAGTCTCAATTTGGATCACCCTTAAACCTATC
TTTTCAATTTGGGCCGGGTTGTGGTTTGGACTACCATGAACAACTTTTCGTCATGTCTAACTTCCCTTTCA
GCAAACATATGAACCATATATAGAGGAGATCGGCCGTATACTAGAGCTGATGTGTTTAAGGTCGTTGATTG
CACGAGAAAAAAAAATCCAAATCGCAACAATAGCAAATTTATCTGGTTCAAAGTGAAAAGATATGTTTAAA
GGTAGTCCAAAGTAAAACTTATAGATAATAAAATGTGGTCCAAAGCGTAATTCACTCAAAAAAAATCAACG
AGACGTGTACCAAACGGAGACAAACGGCATCTTCTCGAAATTTCCCAACCGCTCGCTCGCCCGCCTCGTCT
TCCCGGAAACCGCGGTGGTTTCAGCGTGGCGGATTCTCCAAGCAGACGGAGACGTCACGGCACGGGACTCC
TCCCACCACCCAACCGCCATAAATACCAGCCCCCTCATCTCCTCTCCTCGCATCAGCTCCACCCCCGAAAA
ATTTCTCCCCAATCTCGCGAGGCTCTCGTCGTCGAATCGAATCCTCTCGCGTCCTCAAGGTACGCTGCTTC
TCCTCTCCTCGCTTCGTTTCGATTCGATTTCGGACGGGTGAGGTTGTTTTGTTGCTAGATCCGATTGGTGG
TTAGGGTTGTCGATGTGATTATCGTGAGATGTTTAGGGGTTGTAGATCTGATGGTTGTGATTTGGGCACGG
TTGGTTCGATAGGTGGAATCGTGGTTAGGTTTTGGGATTGGATGTTGGTTCTGATGATTGGGGGGAATTTT
TACGGTTAGATGAATTGTTGGATGATTCGATTGGGGAAATCGGTGTAGATCTGTTGGGGAATTGTGGAACT
AGTCATGCCTGAGTGATTGGTGCGATTTGTAGCGTGTTCCATCTTGTAGGCCTTGTTGCGAGCATGTTCAG
ATCTACTGTTCCGCTCTTGATTGAGTTATTGGTGCCATGGGTTGGTGCAAACACAGGCTTTAATATGTTAT
ATCTGTTTTGTGTTTGATGTAGATCTGTAGGGTAGTTCTTCTTAGACATGGTTCAATTATGTAGCTTGTGC
GTTTCGATTTGATTTCATATGTTCACAGATTAGATAATGATGAACTCTTTTAATTAATTGTCAATGGTAAA
TAGGAAGTCTTGTCGCTATATCTGTCATAATGATCTCATGTTACTATCTGCCAGTAATTTATGCTAAGAAC

TATATTAGAATATCATGTTACAATCTGTAGTAATATCATGTTACAATCTGTAGTTCATCTATATAATCTAT

TGTGGTAATTTCTTTTTACTATCTGTGTGAAGATTATTGCCACTAGTTCATTCTACTTATTTCTGAAGTTC

AGGATACGTGTGCTGTTACTACCTATCTGAATACATGTGTGATGTGCCTGTTACTATCTTTTTGAATACAT

GTATGTTCTGTTGGAATATGTTTGCTGTTTGATCCGTTGTTGTGTCCTTAATCTTGTGCTAGTTCTTACCC

TATCTGTTTGGTGATTATTTCTTGCAGATAGTTATCAACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAG

ATAGAACCAATTCTCTAAGGAAATACTTAACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCAT

GATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGT

CCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCA

CCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAG

AACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAG

AAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGG

TGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCAC

CCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAA

GAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGT

TCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAG

CTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCAT

CCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGA

ATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTG

GCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGAT

CGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCC

TGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCAC

CAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGA

CCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCA

AGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGG

AAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCG

GCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCA

TCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAA

ACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGAT

GACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA

CCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGC

GAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGA

GGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCT

CCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAG

GACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAA

AACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCA

GGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG

TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACAT

CCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCG

CCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAG

CCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGA

GAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAA

ACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAG

GAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGA

CTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAG

AGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC

GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCT

GGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACG

AGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAG

GATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGT

CGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGT

ACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTAC

AGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGAT

CGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGC

TGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATC

CTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTT

CGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGA

AGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTT

CTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGA

GCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGC
CCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAAT
GAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTT
CTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATA
AGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCC
TTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCT
GATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGC
CGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGTAA

Supplementary Sequence 2. Synthesized SR target sequences in polycistronic tRNA-gRNA (PTG) system.

Underlined fragment indicates SR target.

SR32 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAAAGATCCCTCCGAGGCCA
CCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

SR33a target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAAAAATTCCTCCAAGGCCT
CCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

SR33 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGGAACCTCCCGGGCGACA
TCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

SR40 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACATAAGGATATAGATCTT
GCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

RSZ21a target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACGTGACGGCGCGGGAGCT

CGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

RSZ21 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGGATCCCCGAGTGACTTC

CGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

RSZ23 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGAACCTGGATCCGCGCGT

GAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

SC25 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACCTCGGCCGCTACGGCCC

ACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

SC32 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAATGTCGCGCTTCGGCCGC

TCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

SC34 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACTTCGGGAGGTCGGGGCC
GCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

SCL25 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGAGGACCTTCGTCGGCCA
TTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

SCL26 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAAGTCCTCGTGGGCGCTAT
GGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

SCL28 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACCGGGTACCGGAGCCGGA

GCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

SCL30a/SCL30 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAAGTCCCCCTAGGAGGGGA

TAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

SCL57 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGGAGCCGCAGCCCCAGCA

AGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

RS2Z36/RS2Z38 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCATGTTTTAACTGTGGGATT

GAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

RS2Z37 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCATTCCTCGCGCACCCGGAC

CCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTTTTT

RS2Z39 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACCCGGGATGCAAATGATG
CGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

RS29 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGCTATCCTTTTGGCCCTG
GGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

RS33 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGAGCGCCTCTTCAGCAAA
TAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

SR45-1 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACTCCGCCTCGCGCTCCTC
CTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

SR45-2 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCATCCGTCTCCGTCTCCGGC
GAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTT

SR32, SR33a, SR33 and SR40 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAAAGATCCCTCCGAGGCCA
CCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATT
CCCGGCTGGTGCAAAAATTCCTCCAAGGCCTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGT
ACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGGAACCTCCCGGGCGACATCGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAA
CAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGG
TGCACATAAGGATATAGATCTTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC
AACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTTTT

RSZ21a, RSZ21 and RSZ23 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACGTGACGGCGCGGGAGCT
CGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATT
CCCGGCTGGTGCAGGATCCCCGAGTGACTTCCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGT
ACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGAACCTGGATCCGCGCGTGAGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTT
TTTTTTTT

SC25, SC32 and SC34 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACCTCGGCCGCTACGGCCC
ACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATT
CCCGGCTGGTGCAATGTCGCGCTTCGGCCGCTCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGT
ACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACTTCGGGAGGTCGGGGCCGCGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTT
TTTTTTTT

SCL25, SCL26 and SCL28 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGAGGACCTTCGTCGGCCA
TTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATT
CCCGGCTGGTGCAAGTCCTCGTGGGCGCTATGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGT
ACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACCGGGTACCGGAGCCGGAGCGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTT
TTTTTTTT

SCL30a/SCL30 and SCL57 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAAGTCCCCCTAGGAGGGGA
TAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATT
CCCGGCTGGTGCAGGAGCCGCAGCCCCAGCAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTTTT

RS2Z36/RS2Z38, RS2Z37 and RS2Z39 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA
CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG
GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG
TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT
GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA
CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG
TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCATGTTTTAACTGTGGGATT
GAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATT
CCCGGCTGGTGCATTCCTCGCGCACCCGGACCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGT
ACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACCCGGGATGCAAATGATGCGGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTT
TTTTTTTT

RS29 and RS33 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCAGCTATCCTTTTGGCCCTG

GGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATT

CCCGGCTGGTGCAGAGCGCCTCTTCAGCAAATAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG

TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTTTT

SR45-1 and SR45-2 target

GACCATGATTACGCCAAGCTTAAGGAATCTTTAAACATACGAACAGATCACTTAAAGTTCTTCTGAAGCAA

CTTAAAGTTATCAGGCATGCATGGATCTTGGAGGAATCAGATGTGCAGTCAGGGACCATAGCACAAGACAG

GCGTCTTCTACTGGTGCTACCAGCAAATGCTGGAAGCCGGGAACACTGGGTACGTTGGAAACCACGTGATG

TGAAGAAGTAAGATAAACTGTAGGAGAAAAGCATTTCGTAGTGGGCCATGAAGCCTTTCAGGACATGTATT

GCAGTATGGGCCGGCCCATTACGCAATTGGACGACAACAAAGACTAGTATTAGTACCACCTCGGCTATCCA

CATAGATCAAAGCTGATTTAAAAGAGTTGTGCAGATGATCCGTGGCAAACAAAGCACCAGTGGTCTAGTGG

TAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCACTCCGCCTCGCGCTCCTC

CTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATT

CCCGGCTGGTGCATCCGTCTCCGTCTCCGGCGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG

TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTTTT

Supplementary Table 5. List of primers used in Chapter 5.

Primer name	Primer sequence in 5' to 3' orientation	Primer purpose
pRGEB32-SeqF	CAGGAAACAGCTATGAC	Sequencing of polycistronic tRNA-gRNA
pRGEB32-SeqR	CTAGTCCGTTTTTAGCGCG	Sequencing of polycistronic tRNA-gRNA
SR32-F	GGGAAATAAGTAAGGCATGGC	Sequencing of SR target site
SR32-R	GTAACCAGAACTGTGTGAAGTG	Sequencing of SR target site
SR33a-F	CACTTGGAGAGATTCCACCAG	Sequencing of SR target site
SR33a-R	CACCTCGCTACCTCCTATTC	Sequencing of SR target site
SR33-F	GTGGTCCAGTTTGGGGTTCAG	Sequencing of SR target site
SR33-R	GTAGCTGCTAGAAGCATACAG	Sequencing of SR target site
SR40-F	GTACTAAAGGATGGACCAGTTG	Sequencing of SR target site
SR40-R	GAATAGTTAGGTACCATGCG	Sequencing of SR target site
RSZ21a-F	CTCTCGCCTCCCCCCTCTTAC	Sequencing of SR target site
RSZ21a-R	GCCTGGATGCCATAGCTGGG	Sequencing of SR target site
RSZ21-F	CAGCTTCAGATTTCTGACACC	Sequencing of SR target site
RSZ21-R	CTTATAGTACAAGGGGTACC	Sequencing of SR target site
RSZ23-F	GTACTCTCCAGATCCCCCGC	Sequencing of SR target site
RSZ23-R	CCCAATCCCACCAGTAGAAC	Sequencing of SR target site
SC25-F	GCCGCGTCATCTCTTCTCTTC	Sequencing of SR target site
SC25-R	GAACGCGAAGGCTCTCGAATC	Sequencing of SR target site
SC32-F	GTGGCCACCCTCACCCTAC	Sequencing of SR target site
SC32-R	CAGCCGACCTAATCAGCTCCC	Sequencing of SR target site
SC34-F	CTCCACTGCACACACTCGCTG	Sequencing of SR target site
SC34-R	GAATCCCCTCGAGTCCCCAGTC	Sequencing of SR target site
SCL25-F	CTTAAGCCGGTTATAGCGGCC	Sequencing of SR target site

SCL25-R	CCCTTCTCTTCAACAGACTTG	Sequencing of SR target site
SCL26-F	GCTTGTTTTATTCCCTACCAG	Sequencing of SR target site
SCL26-R	CCTTACAGTTACGAACTTCTG	Sequencing of SR target site
SCL28-F	ACCTCTCCTCGCCCAAGCC	Sequencing of SR target site
SCL28-R	GTAAGGCCTAGATCAAATGGGC	Sequencing of SR target site
SCL30a-F	GTACTTGAAGTTACCTATGGG	Sequencing of SR target site
SCL30a-R	GGAACTCTAAGATCCTCCGC	Sequencing of SR target site
SCL30-F	GTTGGCCCAACATATTTTCTG	Sequencing of SR target site
SCL30-R	CTGCATAACATCAGAACAGC	Sequencing of SR target site
SCL57-F	CTGCCTAGCAAGTAGAATACC	Sequencing of SR target site
SCL57-R	CCAGACATGACAGAAGTTAGAC	Sequencing of SR target site
RS2Z36-F	CCTTGATGGCAGGGATGTTG	Sequencing of SR target site
RS2Z36-R	GACCGGCTGATATAAAGAATG	Sequencing of SR target site
RS2Z37-F	CCCTAGCTTCTCTTGTAGATC	Sequencing of SR target site
RS2Z37-R	CAGCATGCTGCTGGCCTAGAC	Sequencing of SR target site
RS2Z38-F	GGGATGTTGATGGAAGCCGC	Sequencing of SR target site
RS2Z38-R	CTGGACCTAGATCGGCTGATAG	Sequencing of SR target site
RS2Z39-F	CACAGACTAGTGGATGTTCCC	Sequencing of SR target site
RS2Z39-R	CCCTCTCTCCACAGCCATAG	Sequencing of SR target site
RS29-F	GCAGGTGCAGTACTCTTCTG	Sequencing of SR target site
RS29-R	GTGGCCTCTTCCTGTGTTTC	Sequencing of SR target site
RS33-F	CCTAGCATTGGCTATGGTGG	Sequencing of SR target site
RS33-R	CAAATGGAGTACCCACTCGC	Sequencing of SR target site
SR45-1-F	CCCCTCCCTCAGACGACCC	Sequencing of SR target site
SR45-1-R	CGATTGCGCAGAACAAAAATCC	Sequencing of SR target site

REFERENCES

Ali, G.S., Palusa, S.G., Golovkin, M., Prasad, J., Manley, J.L. and Reddy, A.S. (2007) Regulation of plant developmental processes by a novel splicing factor. *Plos one* 2, e471.

Ali, G.S. and Reddy, A.S. (2008) Regulation of alternative splicing of pre-mRNAs by stresses. *Current topics in microbiology and immunology* 326, 257-275.

Ali, Z., Abul-faraj, A., Li, L.X., Ghosh, N., Piatek, M., Mahjoub, A., Aouida, M., Piatek, A., Baltes, N.J., Voytas, D.F., Dinesh-Kumar, S. and Mahfouz, M.M. (2015a) Efficient virus-mediated genome editing in plants using the CRISPR/Cas9 System. *Molecular plant* 8, 1288-1291.

Ali, Z., Abulfaraj, A., Idris, A., Ali, S., Tashkandi, M. and Mahfouz, M.M. (2015b) CRISPR/Cas9-mediated viral interference in plants. *Genome biology* 16, 238.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712.

Barren, B. and Artemyev, N.O. (2007) Mechanisms of dominant negative G-protein alpha subunits. *Journal of neuroscience research* 85, 3505-3514.

Barta, A., Kalyna, M. and Reddy, A.S. (2010) Implementing a rational and consistent nomenclature for serine/arginine-rich protein splicing factors (SR proteins) in plants. *The plant cell* 22, 2926-2929.

Bartsevich, V.V., Miller, J.C., Case, C.C. and Pabo, C.O. (2003) Engineered zinc finger proteins for controlling stem cell fate. *Stem cells* 21, 632-637.

Bedard, K.M., Daijogo, S. and Semler, B.L. (2007) A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation. *Embo Journal* 26, 459-467.

Beerli, R.R., Segal, D.J., Dreier, B. and Barbas, C.F. (1998) Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America* 95, 14628-14633.

Beggs, J.D. (1978) Transformation of yeast by a replicating hybrid plasmid. *Nature* 275, 104-109.

Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363-366.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H. and Luhrmann, R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452, 846-850.

Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* 300, 764-764.

Bikard, D., Jiang, W.Y., Samai, P., Hochschild, A., Zhang, F. and Marraffini, L.A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research* 41, 7429-7437.

Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) *Fok*I dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America* 95, 10570-10575.

Boch, J. and Bonas, U. (2010) *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology* 48, 419-436.

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, 1509-1512.

Bogdanove, A.J., Schornack, S. and Lahaye, T. (2010) TAL effectors: finding plant genes for disease and defense. *Current opinion in plant biology* 13, 394-401.

Bogdanove, A.J. and Voytas, D.F. (2011) TAL effectors: customizable proteins for DNA targeting. *Science* 333, 1843-1846.

Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Paro, R., Perrimon, N. and Consortium, H.F.A. (2004) Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* 303, 832-835.

Briggs, A.W., Rios, X., Chari, R., Yang, L.H., Zhang, F., Mali, P. and Church, G.M. (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic acids research* 40.

Bruggeman, Q., Garmier, M., de Bont, L., Soubigou-Taconnat, L., Mazubert, C., Benhamed, M., Raynaud, C., Bergounioux, C. and Delarue, M. (2014) The polyadenylation factor subunit cleavage and polyadenylation specificity Factor30: a key factor of programmed cell death and a regulator of immunity in *Arabidopsis*. *Plant physiology* 165, 732-746.

Bultmann, S., Morbitzer, R., Schmidt, C.S., Thanisch, K., Spada, F., Elsaesser, J., Lahaye, T. and Leonhardt, H. (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic acids research* 40, 5368-5377.

Cameron, F.H. and Jennings, P.A. (1991) Inhibition of gene-expression by a short sense fragment. *Nucleic acids research* 19, 469-475.

Capecchi, M.R. (1989) Altering the genome by homologous recombination. *Science* 244, 1288-1292.

Carbery, I.D., Ji, D., Harrington, A., Brown, V., Weinstein, E.J., Liaw, L. and Cui, X.X. (2010) Targeted genome modification in mice using zinc-finger nucleases. *Genetics* 186, 451-U437.

Carroll, D. (2011) Genome engineering with zinc-finger nucleases. *Genetics* 188, 773-782.

Carvalho, R.F., Carvalho, S.D. and Duque, P. (2010) The plant-specific SR45 protein negatively regulates glucose and ABA signaling during early seedling development in *Arabidopsis*. *Plant physiology* 154, 772-783.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research* 39, e82.

Chang, N.N., Sun, C.H., Gao, L., Zhu, D., Xu, X.F., Zhu, X.J., Xiong, J.W. and Xi, J.J. (2013) Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell research* 23, 465-472.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S. and Huang, B. (2013a) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479-1491.

Chen, T., Cui, P., Chen, H., Ali, S., Zhang, S.D. and Xiong, L.M. (2013b) A KH-domain RNA-binding protein interacts with FIERY2/CTD phosphatase-like 1 and splicing factors and is important for pre-mRNA splicing in *Arabidopsis*. *Plos genetics* 9, e1003875.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B. and Jaenisch, R. (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell research* 23, 1163-1171.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S. and Kim, J.S. (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research* 24, 132-141.

Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1-8.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. and Voytas, D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757-761.

Christian, M., Qi, Y.P., Zhang, Y. and Voytas, D.F. (2013) Targeted mutagenesis of *Arabidopsis thaliana* using engineered TAL effector nucleases. *G3-Genes genomes genetics* 3, 1697-1705.

Cong, L., Ran, F.A., Cox, D., Lin, S.L., Barretto, R., Habib, N., Hsu, P.D., Wu, X.B., Jiang, W.Y., Marraffini, L.A. and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823.

Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. and Zhang, F. (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nature Communications* 3, 968.

Cronin, C.A., Gluba, W. and Scrable, H. (2001) The lac operator-repressor system is functional in the mouse. *Genes and development* 15, 1506-1517.

Cruz, T.M., Carvalho, R.F., Richardson, D.N. and Duque, P. (2014) Abscisic acid (ABA) regulation of *Arabidopsis* SR protein gene expression. *International journal of molecular sciences* 15, 17541-17564.

Curtin, S.J., Zhang, F., Sander, J.D., Haun, W.J., Starker, C., Baltes, N.J., Reyon, D., Dahlborg, E.J., Goodwin, M.J., Coffman, A.P., Dobbs, D., Joung, J.K., Voytas, D.F. and Stupar, R.M. (2011) Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant physiology* 156, 466-473.

de Lange, O., Schreiber, T., Schandry, N., Radeck, J., Braun, K.H., Koszinowski, J., Heuer, H., Strauss, A. and Lahaye, T. (2013) Breaking the DNA-binding code of *Ralstonia solanacearum* TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. *New phytologist* 199, 773-786.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y. and Yan, N. (2012a) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335, 720-723.

Deng, D., Yin, P., Yan, C.Y., Pan, X.J., Gong, X.Q., Qi, S., Xie, T., Mahfouz, M., Zhu, J.K., Yan, N. and Shi, Y.G. (2012b) Recognition of methylated DNA by TAL effectors. *Cell research* 22, 1502-1504.

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research* 41, 4336-4343.

Duque, P. (2011) A role for SR proteins in plant stress responses. *Plant signaling & behavior* 6, 49-54.

Engler, C., Gruetzner, R., Kandzia, R. and Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *Plos one* 4, e5553.

Feng, Z.Y., Mao, Y.F., Xu, N.F., Zhang, B.T., Wei, P.L., Yang, D.L., Wang, Z., Zhang, Z.J., Zheng, R., Yang, L., Zeng, L., Liu, X.D. and Zhu, J.K. (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 111, 4632-4637.

Feng, Z.Y., Zhang, B.T., Ding, W.N., Liu, X.D., Yang, D.L., Wei, P.L., Cao, F.Q., Zhu, S.H., Zhang, F., Mao, Y.F. and Zhu, J.K. (2013) Efficient genome editing in plants using a CRISPR/Cas system. *Cell research* 23, 1229-1232.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K. and Mockler, T.C. (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome research* 20, 45-58.

Fire, A., Xu, S.Q., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806-811.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* 31, 822-826.

Fu, Y.F., Sander, J.D., Reyon, D., Cascio, V.M. and Joung, J.K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nauret biotechnology* 32, 279-284.

Gaj, T., Mercer, A.C., Sirk, S.J., Smith, H.L. and Barbas, C.F. (2013) A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic acids research* 41, 3937-3946.

Garg, A., Lohmueller, J.J., Silver, P.A. and Armel, T.Z. (2012) Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic acids research* 40, 7584-7595.

Gasiunas, G., Sinkunas, T. and Siksnys, V. (2014) Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular and molecular life sciences* 71, 449-465.

Geissler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.E. and Boch, J. (2011) Transcriptional activators of human genes with programmable DNA-specificity. *Plos one* 6, e19509.

Genin, S. and Denny, T.P. (2012) Pathogenomics of the *Ralstonia solanacearum* species complex. *Annual review of phytopathology* 50, 67-89.

Gersbach, C.A., Gaj, T. and Barbas, C.F., 3rd (2014) Synthetic zinc finger proteins: the advent of targeted gene regulation and genome modification technologies. *Accounts of chemical research* 47, 2309-2318.

Geurts, A.M., Cost, G.J., Freyvert, Y., Zeitler, B., Miller, J.C., Choi, V.M., Jenkins, S.S., Wood, A., Cui, X.X., Meng, X.D., Vincent, A., Lam, S., Michalkiewicz, M., Schilling, R., Foeckler, J., Kalloway, S., Weiler, H., Menoret, S., Anegon, I., Davis, G.D., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D., Jacob, H.J. and Buelow, R. (2009) Knockout rats *via* embryo microinjection of zinc-finger nucleases. *Science* 325, 433-433.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., Qi, L.S., Kampmann, M. and Weissman, J.S. (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159, 647-661.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S. and Qi, L.S. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442-451.

Golovkin, M. and Reddy, A.S. (1999) An SC35-like protein and a novel serine/arginine-rich protein interact with *Arabidopsis* U1-70K protein. *The journal of biological chemistry* 274, 36428-36438.

Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A. and Ruddle, F.H. (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. *Proceedings of the National Academy of Sciences of the United States of America* 77, 7380-7384.

Gossen, M. and Bujard, H. (1992) Tight control of gene-expression in mammalian-cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America* 89, 5547-5551.

Graslund, T., Li, X., Magnenat, L., Popkov, M. and Barbas, C.F., 3rd (2005) Exploring strategies for the design of artificial transcription factors: targeting sites proximal to known regulatory regions for the induction of gamma-globin expression and the treatment of sickle cell disease. *The journal of biological chemistry* 280, 3707-3714.

Grierson, D., Fray, R.G., Hamilton, A.J., Smith, C.J.S. and Watson, C.F. (1991) Does co-suppression of sense genes in transgenic plants involve antisense RNA. *Trends in biotechnology* 9, 122-123.

Guo, S. and Kemphues, K.J. (1995) Par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* 81, 611-620.

Gurushidze, M., Hensel, G., Hiekel, S., Schedel, S., Valkov, V. and Kumlehn, J. (2014) True-breeding targeted gene knock-out in barley using designer TALE nuclease in haploid cells. *Plos one* 9.

Hall, B., Limaye, A. and Kulkarni, A.B. (2009) Overview: generation of gene knockout mice. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]* Chapter 19, Unit 19 12, 11-17.

Hamilton, A.J. and Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286, 950-952.

Hanson, K.D. and Sedivy, J.M. (1995) Analysis of biological selections for high-efficiency gene targeting. *Molecular and cellular biology* 15, 45-51.

Hayashi, H., Czaja, I., Lubenow, H., Schell, J. and Walden, R. (1992) Activation of a plant gene by T-DNA tagging: auxin-independent growth in vitro. *Science* 258, 1350-1353.

Heuer, H., Yin, Y.N., Xue, Q.Y., Smalla, K. and Guo, J.H. (2007) Repeat domain diversity of avrBs3-like genes in *Ralstonia solanacearum* strains and association with host preferences in the field. *Applied and environmental microbiology* 73, 4379-4384.

Heyl, A., Ramireddy, E., Brenner, W.G., Riefler, M., Allemeersch, J. and Schmulling, T. (2008) The transcriptional repressor ARR1-SRDX suppresses pleiotropic cytokinin activities in *Arabidopsis*. *Plant physiology* 147, 1380-1395.

Hiei, Y. and Komari, T. (2008) *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. *Nature protocols* 3, 824-834.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E. and Gersbach, C.A. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* 33, 510-517.

Hiratsu, K., Matsui, K., Koyama, T. and Ohme-Takagi, M. (2003) Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *Plant journal* 34, 733-739.

Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167-170.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y.Q., Fine, E.J., Wu, X.B., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G. and Zhang, F. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnolology* 31, 827-832.

Hu, J.B., Lei, Y., Wong, W.K., Liu, S.Q., Lee, K.C., He, X.J., You, W.X., Zhou, R., Guo, J.T., Chen, X.F., Peng, X.L., Sun, H., Huang, H., Zhao, H. and Feng, B. (2014) Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic acids research* 42, 4375-4390.

Huang, P., Xiao, A., Zhou, M.G., Zhu, Z.Y., Lin, S. and Zhang, B. (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nature biotechnolology* 29, 699-700.

Huang, Y.Q., Gattoni, R., Stevenin, J. and Steitz, J.A. (2003) SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Molecular cell* 11, 837-843.

Hwang, W.Y., Fu, Y.F., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R.J. and Joung, J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* 31, 227-229.

Ichikawa, T., Nakazawa, M., Kawashima, M., Iizumi, H., Kuroda, H., Kondou, Y., Tsuhara, Y., Suzuki, K., Ishikawa, A., Seki, M., Fujita, M., Motohashi, R., Nagata, N., Takagi, T., Shinozaki, K. and Matsui, M. (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. *Plant journal* 48, 974-985.

Iida, K. and Go, M. (2006) Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Molecular biology and evolution* 23, 1085-1094.

International Brachypodium, I. (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763-768.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) Nucleotide-sequence of the IAP gene, responsible for alkaline-phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology* 169, 5429-5433.

Isshiki, M., Tsumoto, A. and Shimamoto, K. (2006) The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. *The plant cell* 18, 146-158.

Jacobs, T.B., LaFayette, P.R., Schmitz, R.J. and Parrott, W.A. (2015) Targeted genome modifications in soybean with CRISPR/Cas9. *BMC biotechnology* 15, 16.

Jagtap, U.B., Gurav, R.G. and Bapat, V.A. (2011) Role of RNA interference in plant improvement. *Naturwissenschaften* 98, 473-492.

Jansen, R., van Embden, J.D.A., Gaastra, W. and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* 43, 1565-1575.

Jiang, W.Y., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013a) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* 31, 233-239.

Jiang, W.Z., Yang, B. and Weeks, D.P. (2014) Efficient CRISPR/Cas9-mediated gene editing in *Arabidopsis thaliana* and inheritance of modified genes in the T2 and T3 generations. *Plos one* 9, e99225.

Jiang, W.Z., Zhou, H.B., Bi, H.H., Fromm, M., Yang, B. and Weeks, D.P. (2013b) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic acids research* 41, e188.

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

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A.T., Charpentier, E., Nogales, E. and Doudna, J.A. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343, 1247997.

Jorda, J. and Kajava, A.V. (2009) T-REKS: identification of tandem repeats in sequences with a K-means based algorithm. *Bioinformatics* 25, 2632-2638.

Kagale, S. and Rozwadowski, K. (2011) EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. *Epigenetics* 6, 141-146.

Kakimoto, T. (1996) CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. *Science* 274, 982-985.

Kalyna, M., Lopato, S. and Barta, A. (2003) Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Molecular biology of the cell* 14, 3565-3577.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P. and Ahringer, J. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421, 231-237.

Karginov, F.V. and Hannon, G.J. (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Molecullar cell* 37, 7-19.

Kataoka, N., Bachorik, J.L. and Dreyfuss, G. (1999) Transportin-SR, a nuclear import receptor for SR proteins. *The Journal of cell biology* 145, 1145-1152.

Kay, S., Hahn, S., Marois, E., Hause, G. and Bonas, U. (2007) A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318, 648-651.

Kearns, N.A., Genga, R.M., Enuameh, M.S., Garber, M., Wolfe, S.A. and Maehr, R. (2014) Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* 141, 219-223.

Kempin, S.A., Liljegren, S.J., Block, L.M., Rounsley, S.D., Yanofsky, M.F. and Lam, E. (1997) Targeted disruption in *Arabidopsis*. *Nature* 389, 802-803.

Kiani, S., Beal, J., Ebrahimkhani, M.R., Huh, J., Hall, R.N., Xie, Z., Li, Y. and Weiss, R. (2014) CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nature methods* 11, 723-726.

Kleinstiver, B.P., Wang, L., Wolfs, J.M., Kolaczyk, T., McDowell, B., Wang, X., Schild-Poulter, C., Bogdanove, A.J. and Edgell, D.R. (2014) The *I-Tev*I nuclease and linker domains contribute to the specificity of monomeric TALENs. *G3-Genes genomes genetics* 4, 1155-1165.

Kleinstiver, B.P., Wolfs, J.M., Kolaczyk, T., Roberts, A.K., Hu, S.X. and Edgell, D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proceedings of the National Academy of Sciences of the United States of America* 109, 8061-8066.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., Nureki, O. and Zhang, F. (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583-588.

Kong, J., Gong, J.M., Zhang, Z.G., Zhang, J.S. and Chen, S.Y. (2003) A new AOX homologous gene OsIM1 from rice (*Oryza sativa L.*) with an alternative splicing mechanism under salt stress. *Theoretical and applied genetics* 107, 326-331.

Kunin, V., Sorek, R. and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome biology* 8.

Larkin, P.D. and Park, W.D. (1999) Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch synthase are temperature-sensitive and controlled by a single-nucleotide polymorphism. *Plant molecular biology* 40, 719-727.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S. and Qi, L.S. (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature protocols* 8, 2180-2196.

Lassus, P., Opitz-Araya, X. and Lazebnik, Y. (2002) Requirement for caspase-2 in stress-induced apoptosis before mitochondrial permeabilization. *Science* 297, 1352-1354.

Laufer, B.I. and Singh, S.M. (2015) Strategies for precision modulation of gene expression by epigenome editing: an overview. *Epigenetics and chromatin* 8, 34.

Lazar, G., Schaal, T., Maniatis, T. and Goodman, H.M. (1995) Identification of a plant Serine-Arginine-rich protein similar to the mammalian splicing factor Sf2/Asf. *Proceedings of the National Academy of Sciences of the United States of America* 92, 7672-7676.

Le Hir, H., Izaurralde, E., Maquat, L.E. and Moore, M.J. (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *The EMBO journal* 19, 6860-6869.

Lee, K.Y., Lund, P., Lowe, K. and Dunsmuir, P. (1990) Homologous recombination in plant cells after *Agrobacterium*-mediated transformation. *The plant cell* 2, 415-425.

LeProust, E.M., Peck, B.J., Spirin, K., McCuen, H.B., Moore, B., Namsaraev, E. and Caruthers, M.H. (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic acids research* 38, 2522-2540.

Li, J.F., Norville, J.E., Aach, J., McCormack, M., Zhang, D.D., Bush, J., Church, G.M. and Sheen, J. (2013a) Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature biotechnology* 31, 688-691.

Li, L., Atef, A., Piatek, A., Ali, Z., Piatek, M., Aouida, M., Sharakuu, A., Mahjoub, A., Wang, G., Khan, S., Fedoroff, N.V., Zhu, J.K. and Mahfouz, M.M. (2013b) Characterization and DNA-binding specificities of *Ralstonia* TAL-like effectors. *Molecular plant* 6, 1318-1330.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P. and Yang, B. (2011a) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and *Fok*I DNA-cleavage domain. *Nucleic acids research* 39, 359-372.

Li, T., Huang, S., Zhao, X.F., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P. and Yang, B. (2011b) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic acids research* 39, 6315-6325.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P. and Yang, B. (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nature biotechnology* 30, 390-392.

Li, X.L. and Manley, J.L. (2005) Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122, 365-378.

Liang, Z., Zhang, K., Chen, K.L. and Gao, C.X. (2014) Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *Journal of genetics and genomics* 41, 63-68.

Lieber, M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* 79, 181-211.

Lin, J.F., Chen, H., Luo, L., Lai, Y.R., Xie, W. and Kee, K. (2015) Creating a monomeric endonuclease TALE-*I-Sce*I with high specificity and low genotoxicity in human cells. *Nucleic acids research* 43, 1112-1122.

Lin, S.R., Coutinho-Mansfield, G., Wang, D., Pandit, S. and Fu, X.D. (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nature structural and molecular biology* 15, 819-826.

Lindbo, J.A. (2012) A historical overview of RNAi in plants. *Methods in molecular biology* 894, 1-16.

Liu, X., Wang, Y.S., Guo, W.J., Chang, B.H., Liu, J., Guo, Z.K., Quan, F.S. and Zhang, Y. (2013) Zinc-finger nickase-mediated insertion of the lysostaphin gene into the beta-casein locus in cloned cows. *Nature communications* 4.

Liu, Y.X., Kern, J.T., Walker, J.R., Johnson, J.A., Schultz, P.G. and Luesch, H. (2007) A genomic screen for activators of the antioxidant response element. *Proceedings of*

the National Academy of Sciences of the United States of America 104, 5205-5210.

Longman, D., Johnstone, I.L. and Caceres, J.F. (2000) Functional characterization of SR and SR-related genes in *Caenorhabditis elegans*. *The EMBO journal* 19, 1625-1637.

Lopato, S., Kalyna, M., Dorner, S., Kobayashi, R., Krainer, A.R. and Barta, A. (1999) atSRp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. *Genes and development* 13, 987-1001.

Lopato, S., Mayeda, A., Krainer, A.R. and Barta, A. (1996a) Pre-mRNA splicing in plants: Characterization of Ser Arg splicing factors. *Proceedings of the National Academy of Sciences of the United States of America* 93, 3074-3079.

Lopato, S., Waigmann, E. and Barta, A. (1996b) Characterization of a novel arginine/serine-rich splicing factor in *Arabidopsis*. *The plant cell* 8, 2255-2264.

Lowder, L.G., Zhang, D., Baltes, N.J., Paul, J.W., 3rd, Tang, X., Zheng, X., Voytas, D.F., Hsieh, T.F., Zhang, Y. and Qi, Y. (2015) A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant physiology* 169, 971-985.

Lykke-Andersen, J., Shu, M.D. and Steitz, J.A. (2001) Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* 293, 1836-1839.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H. and Joung, J.K. (2013a) CRISPR RNA-guided activation of endogenous human genes. *Nature methods* 10, 977-979.

Maeder, M.L., Linder, S.J., Reyon, D., Angstman, J.F., Fu, Y.F., Sander, J.D. and Joung, J.K. (2013b) Robust, synergistic regulation of human gene expression using TALE activators. *Nature methods* 10, 243-245.

Mahfouz, M.M. and Li, L. (2011) TALE nucleases and next generation GM crops. *GM crops* 2, 99-103.

Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X. and Zhu, J.K. (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America* 108, 2623-2628.

Mahfouz, M.M., Li, L.X., Piatek, M., Fang, X.Y., Mansour, H., Bangarusamy, D.K. and Zhu, J.K. (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant molecular biology* 78, 311-321.

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J. and Stoddard, B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335, 716-719.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., van der Oost, J. and Koonin, E.V. (2011) Evolution and classification of the CRISPR-Cas systems. *Nature reviews microbiology* 9, 467-477.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R. and Koonin, E.V. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews microbiology* 13, 722-736.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M. (2013a) Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* 31, 833-838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013b) RNA-guided human genome engineering *via* Cas9. *Science* 339, 823-826.

Mani, M., Smith, J., Kandavelou, K., Berg, J.M. and Chandrasegaran, S. (2005) Binding of two zinc finger nuclease monomers to two specific sites is required for effective double-strand DNA cleavage. *Biochemical and biophysical research communications* 334, 1191-1197.

Manley, J.L. and Krainer, A.R. (2010) A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes and development* 24, 1073-1074.

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F. and Zhu, J.K. (2013) Application of the CRISPR-Cas system for efficient genome engineering in plants. *Molecular plant* 6, 2008-2011.

Matzke, M.A. and Matzke, A.J.M. (1991) Differential inactivation and methylation of a transgene in plants by 2 suppressor loci containing homologous sequences. *Plant molecular biology* 16, 821-830.

McCreath, K.J., Howcroft, J., Campbell, K.H., Colman, A., Schnieke, A.E. and Kind, A.J. (2000) Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature* 405, 1066-1069.

McGinnis, K., Murphy, N., Carlson, A.R., Akula, A., Akula, C., Basinger, H., Carlson, M., Hermanson, P., Kovacevic, N., McGill, M.A., Seshadri, V., Yoyokie, J., Cone, K., Kaeppler, H.F., Kaeppler, S.M. and Springer, N.M. (2007) Assessing the efficiency of RNA interference for maize functional genomics. *Plant physiology* 143, 1441-1451.

McMahon, M.A., Rahdar, M. and Porteus, M. (2012) Gene editing: not just for translation anymore. *Nature methods* 9, 28-31.

Meng, X.D., Noyes, M.B., Zhu, L.H.J., Lawson, N.D. and Wolfe, S.A. (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature biotechnology* 26, 695-701.

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H. and Qu, L.J. (2013) Targeted mutagenesis in rice using CRISPR-Cas system. *Cell research* 23, 1233-1236.

Miao, Z.H. and Lam, E. (1995) Targeted disruption of the TGA3 locus in *Arabidopsis thaliana*. *The Plant journal : for cell and molecular biology* 7, 359-365.

Miller, J.C., Holmes, M.C., Wang, J.B., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., Gregory, P.D., Pabo, C.O. and Rebar, E.J. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature biotechnology* 25, 778-785.

Morbitzer, R., Elsaesser, J., Hausner, J. and Lahaye, T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic acids research* 39, 5790-5799.

Morbitzer, R., Romer, P., Boch, J. and Lahaye, T. (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* 107, 21617-21622.

Moritoh, S., Miki, D., Akiyama, M., Kawahara, M., Izawa, T., Maki, H. and Shimamoto, K. (2005) RNAi-mediated silencing of OsGEN-L (OsGEN-like), a new member of the RAD2/XPG nuclease family, causes male sterility by defect of microspore development in rice. *Plant cell physiology* 46, 699-715.

Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501.

Napoli, C., Lemieux, C. and Jorgensen, R. (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *The plant cell* 2, 279-289.

Nasmyth, K.A. and Reed, S.I. (1980) Isolation of genes by complementation in yeast: molecular cloning of a cell-cycle gene. *Proceedings of the National Academy of Sciences of the United States of America* 77, 2119-2123.

Neet, K.E. and Koshland, D.E., Jr. (1966) The conversion of serine at the active site of subtilisin to cysteine: a "chemical mutation". *Proceedings of the National Academy of Sciences of the United States of America* 56, 1606-1611.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D.G. and Kamoun, S. (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature biotechnology* 31, 691-693.

Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R. and Fluhr, R. (2004) Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. *Plant journal* 39, 877-885.

Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P. and Lu, T.K. (2014) Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular cell* 54, 698-710.

Ohta, M., Matsui, K., Hiratsu, K., Shinshi, H. and Ohme-Takagi, M. (2001) Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *The plant cell* 13, 1959-1968.

Osakabe, K., Osakabe, Y. and Toki, S. (2010) Site-directed mutagenesis in *Arabidopsis* using custom-designed zinc finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America* 107, 12034-12039.

Palusa, S.G., Ali, G.S. and Reddy, A.S. (2007) Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. *The Plant journal : for cell and molecular biology* 49, 1091-1107.

Palusa, S.G. and Reddy, A.S. (2010) Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. *The new phytologist* 185, 83-89.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* 40, 1413-1415.

Paszkowski, J., Baur, M., Bogucki, A. and Potrykus, I. (1988) Gene targeting in plants. *The EMBO journal* 7, 4021-4026.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology* 31, 839-843.

Peng, J.Y., Zhou, Y.X., Zhu, S.Y. and Wei, W.S. (2015) High-throughput screens in mammalian cells using the CRISPR-Cas9 system. *Febs Journal* 282, 2089-2096.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., Guilak, F., Crawford, G.E., Reddy, T.E. and Gersbach, C.A. (2013a) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* 10, 973-976.

Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J. and Gersbach, C.A. (2013b) Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nature methods* 10, 239-242.

Piatek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., Aouida, M. and Mahfouz, M.M. (2015) RNA-guided transcriptional regulation *in planta via* synthetic dCas9-based transcription factors. *Plant biotechnol journal* 13, 578-589.

Polstein, L.R. and Gersbach, C.A. (2015) A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature chemical biology* 11, 198-200.

Porteus, M.H. and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science* 300, 763-763.

Porteus, M.H. and Carroll, D. (2005) Gene targeting using zinc finger nucleases. *Nature biotechnology* 23, 967-973.

Poueymiro, M. and Genin, S. (2009) Secreted proteins from *Ralstonia solanacearum*: a hundred tricks to kill a plant. *Current opinion in microbiology* 12, 44-52.

Prelich, G. (2012) Gene Overexpression: Uses, Mechanisms, and Interpretation. *Genetics* 190, 841-854.

Pruett-Miller, S.M., Reading, D.W., Porter, S.N. and Porteus, M.H. (2009) Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. *Plos genetics* 5, e1000376.

Puchta, H. (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *Journal of experimental botany* 56, 1-14.

Puchta, H., Dujon, B. and Hohn, B. (1993) Homologous recombination in plant-cells is enhanced by *in vivo* induction of double-strand breaks into dna by a site-specific endonuclease. *Nucleic acids research* 21, 5034-5040.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173-1183.

Ramirez, C.L., Certo, M.T., Mussolino, C., Goodwin, M.J., Cradick, T.J., McCaffrey, A.P., Cathomen, T., Scharenberg, A.M. and Joung, J.K. (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic acids research* 40, 5560-5568.

Ramirez, C.L., Foley, J.E., Wright, D.A., Muller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J.A., Cathomen, T., Voytas, D.F. and Joung, J.K. (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature methods* 5, 374-375.

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

Rausin, G., Tillemans, V., Stankovic, N., Hanikenne, M. and Motte, P. (2010) Dynamic nucleocytoplasmic shuttling of an *Arabidopsis* SR splicing factor: role of the RNA-binding domains. *Plant physiology* 153, 273-284.

Rebar, E.J., Huang, Y., Hickey, R., Nath, A.K., Meoli, D., Nath, S., Chen, B.L., Xu, L., Liang, Y.X., Jamieson, A.C., Zhang, L., Spratt, S.K., Case, C.C., Wolffe, A. and Giordano, F.J. (2002) Induction of angiogenesis in a mouse model using engineered transcription factors. *Nature medicine* 8, 1427-1432.

Reddy, A.S. and Shad Ali, G. (2011) Plant serine/arginine-rich proteins: roles in precursor messenger RNA splicing, plant development, and stress responses. *Wiley interdisciplinary reviews. RNA* 2, 875-889.

Reddy, A.S.N. (2004) Plant serine/arginine-rich proteins and their role in pre-mRNA splicing. *Trends in plant science* 9, 541-547.

Remenant, B., Coupat-Goutaland, B., Guidot, A., Cellier, G., Wicker, E., Allen, C., Fegan, M., Pruvost, O., Elbaz, M., Calteau, A., Salvignol, G., Mornico, D., Mangenot, S., Barbe, V., Medigue, C. and Prior, P. (2010) Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *Bmc Genomics* 11.

Remigi, P., Anisimova, M., Guidot, A., Genin, S. and Peeters, N. (2011) Functional diversification of the GALA type III effector family contributes to *Ralstonia*

solanacearum adaptation on different plant hosts. *The new phytologist* 192, 976-987.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology* 30, 460-465.

Romano, N. and Macino, G. (1992) Quelling - transient inactivation of gene-expression in *Neurospora crassa* by transformation with homologous sequences. *Molecular microbiology* 6, 3343-3353.

Romer, 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, 645-648.

Romer, P., Strauss, T., Hahn, S., Scholze, H., Morbitzer, R., Grau, J., Bonas, U. and Lahaye, T. (2009) Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper *Bs3* alleles. *Plant physiology* 150, 1697-1712.

Rong, Y.S. and Golic, K.G. (2000) Gene targeting by homologous recombination in *Drosophila*. *Science* 288, 2013-2018.

Russell, W.L., Russell, L.B. and Kelly, E.M. (1958) Radiation dose rate and mutation frequency. *Science* 128, 1546-1550.

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J. and Boucher, C.A. (2002) Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* 415, 497-502.

Sanchez, J.P., Ullman, C., Moore, M., Choo, Y. and Chua, N.H. (2002) Regulation of gene expression in *Arabidopsis thaliana* by artificial zinc finger chimeras. *Plant cell physiology* 43, 1465-1472.

Sanchez, J.P., Ullman, C., Moore, M., Choo, Y. and Chua, N.H. (2006) Regulation of *Arabidopsis thaliana* 4-coumarate : coenzyme-A ligase-1 expression by artificial zinc finger chimeras. *Plant biotechnology journal* 4, 103-114.

Sanford, J.R., Gray, N.K., Beckmann, K. and Caceres, J.F. (2004) A novel role for shuttling SR proteins in mRNA translation. *Genes and development* 18, 755-768.

Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G. and Zhang, F. (2012) A transcription activator-like effector toolbox for genome engineering. *Nature protocols* 7, 171-192.

Schaefer, D.G. and Zryd, J.P. (1997) Efficient gene targeting in the moss *Physcomitrella patens*. *The plant journal : for cell and molecular biology* 11, 1195-1206.

Schiml, S., Fauser, F. and Puchta, H. (2014) The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. *Plant journal* 80, 1139-1150.

Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K. and Hornung, V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nature biotechnology* 31, 76-+.

Segal, D.J., Goncalves, J., Eberhardy, S., Swan, C.H., Torbett, B.E., Li, X. and Barbas, C.F., 3rd (2004) Attenuation of HIV-1 replication in primary human cells with a designed zinc finger transcription factor. *The Journal of biological chemistry* 279, 14509-14519.

Shalem, O., Sanjana, N.E. and Zhang, F. (2015) High-throughput functional genomics using CRISPR-Cas9. *Nature reviews genetics* 16, 299-311.

Shan, Q.W., Wang, Y.P., Chen, K.L., Liang, Z., Li, J., Zhang, Y., Zhang, K., Liu, J.X., Voytas, D.F., Zheng, X.L., Zhang, Y. and Gao, C.X. (2013a) Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Molecular plant* 6, 1365-1368.

Shan, Q.W., Wang, Y.P., Li, J., Zhang, Y., Chen, K.L., Liang, Z., Zhang, K., Liu, J.X., Xi, J.J., Qiu, J.L. and Gao, C.X. (2013b) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature biotechnology* 31, 686-688.

Shen, B., Zhang, J., Wu, H.Y., Wang, J.Y., Ma, K., Li, Z., Zhang, X.G., Zhang, P.M. and Huang, X.X. (2013) Generation of gene-modified mice *via* Cas9/RNA-mediated gene targeting. *Cell research* 23, 720-723.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., Mitchell, J.C., Arnold, N.L., Gopalan, S., Meng, X.D., Choi, V.M., Rock, J.M., Wu, Y.Y., Katibah, G.E., Zhifang, G., McCaskill, D., Simpson, M.A., Blakeslee, B., Greenwalt, S.A., Butler, H.J., Hinkley, S.J., Zhang, L., Rebar, E.J., Gregory, P.D. and Urnov, F.D. (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459, 437-U156.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *Embo Journal* 30, 1335-1342.

Smith, H.O., Hutchison, C.A., 3rd, Pfannkoch, C. and Venter, J.C. (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* 100, 15440-15445.

Smith, W.C. and Harland, R.M. (1991) Injected Xwnt-8 RNA acts early in *Xenopus* embryos to promote formation of a vegetal dorsalizing center. *Cell* 67, 753-765.

Smyth, D.R. (1997) Gene silencing: Cosuppression at a distance. *Current biology* 7, 793-795.

Staudt, N., Molitor, A., Somogyi, K., Mata, J., Curado, S., Eulenberg, K., Meise, M., Siegmund, T., Hader, T., Hilfiker, A., Bronner, G., Ephrussi, A., Rorth, P., Cohen, S.M., Fellert, S., Chung, H.R., Piepenburg, O., Schafer, U., Jackle, H. and Vorbruggen, G. (2005) Gain-of-function screen for genes that affect *Drosophila* muscle pattern formation. *Plos genetics* 1, 499-506.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62-67.

Streubel, J., Blucher, C., Landgraf, A. and Boch, J. (2012) TAL effector RVD specificities and efficiencies. *Nature biotechnology* 30, 593-595.

Sugio, A., Dreos, R., Aparicio, F. and Maule, A.J. (2009) The cytosolic protein response as a subcomponent of the wider heat shock response in *Arabidopsis*. *The plant cell* 21, 642-654.

Takada, S. (2013) Post-embryonic induction of ATML1-SRDX alters the morphology of seedlings. *Plos one* 8, e79312.

Tanabe, N., Yoshimura, K., Kimura, A., Yabuta, Y. and Shigeoka, S. (2007) Differential expression of alternatively spliced mRNAs of *Arabidopsis* SR protein homologs, atSR30 and atSR45a, in response to environmental stress. *Plant cell physiology* 48, 1036-1049.

Tebas, P., Stein, D., Tang, W.W., Frank, I., Wang, S.Q., Lee, G., Spratt, S.K., Surosky, R.T., Giedlin, M.A., Nichol, G., Holmes, M.C., Gregory, P.D., Ando, D.G., Kalos, M., Collman, R.G., Binder-Scholl, G., Plesa, G., Hwang, W.T., Levine, B.L. and June, C.H. (2014) Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *The New England journal of medicine* 370, 901-910.

Tenzen, T., Zembowicz, F. and Cowan, C.A. (2010) Genome modification in human embryonic stem cells. *Journal of cellular physiology* 222, 278-281.

Tiwari, S.B., Belachew, A., Ma, S.F., Young, M., Ade, J., Shen, Y., Marion, C.M., Holtan, H.E., Bailey, A., Stone, J.K., Edwards, L., Wallace, A.D., Canales, R.D., Adam, L., Ratcliffe, O.J. and Repetti, P.P. (2012) The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *The plant journal : for cell and molecular biology* 70, 855-865.

Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F.L., Maeder, M.L., Joung, J.K. and Voytas, D.F. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459, 442-U161.

Travella, S., Klimm, T.E. and Keller, B. (2006) RNA interference-based gene silencing as an efficient tool for functional genomics in hexaploid bread wheat. *Plant physiology* 142, 6-20.

Upadhyay, S.K., Kumar, J., Alok, A. and Tuli, R. (2013) RNA-guided genome editing for target gene mutations in wheat. *G3-Genes genomes genetics* 3, 2233-2238.

Urnov, F.D. and Rebar, E.J. (2002) Designed transcription factors as tools for therapeutics and functional genomics. *Biochemical pharmacology* 64, 919-923.

Venken, K.J. and Bellen, H.J. (2005) Emerging technologies for gene manipulation in *Drosophila melanogaster*. *Nature reviews. Genetics* 6, 167-178.

Voellmy, R. (2005) Dominant-positive and dominant-negative heat shock factors. *Methods* 35, 199-207.

Waldrip, Z.J., Byrum, S.D., Storey, A.J., Gao, J., Byrd, A.K., Mackintosh, S.G., Wahls, W.P., Taverna, S.D., Raney, K.D. and Tackett, A.J. (2014) A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics* 9, 1207-1211.

Wang, H.Y., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F. and Jaenisch, R. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910-918.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. and Qiu, J.L. (2014a) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature biotechnology* 32, 947-951.

Wang, Y.P., Cheng, X., Shan, Q.W., Zhang, Y., Liu, J.X., Gao, C.X. and Qiu, J.L. (2014b) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature biotechnology* 32, 947-951.

Wang, Z., Li, J., Huang, H., Wang, G.C., Jiang, M.J., Yin, S.Y., Sun, C.H., Zhang, H.S., Zhuang, F.F. and Xi, J.J. (2012) An integrated chip for the high-throughput synthesis of transcription activator-like effectors. *Angewandte Chemie International Edition* 51, 8505-8508.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189-1191.

Weber, E., Gruetzner, R., Werner, S., Engler, C. and Marillonnet, S. (2011) Assembly of designer tal effectors by golden gate cloning. *Plos one* 6, e19722.

Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z.Y., Xia, Y.J., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F. and Chory, J. (2000) Activation tagging in *Arabidopsis*. *Plant physiology* 122, 1003-1013.

Wendt, T., Holm, P.B., Starker, C.G., Christian, M., Voytas, D.F., Brinch-Pedersen, H. and Holme, I.B. (2013) TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. *Plant molecular biology* 83, 279-285.

Wright, D.A., Townsend, J.A., Winfrey, R.J., Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall, B.D., Jondle, M.D. and Voytas, D.F. (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant journal* 44, 693-705.

Wu, Y., Gao, T.L., Wang, X.L., Hu, Y.J., Hu, X.Y., Hu, Z.Q., Pang, J.L., Li, Z., Xue, J.F., Feng, M., Wu, L.Q. and Liang, D.S. (2014) TALE nickase mediates high efficient targeted transgene integration at the human multi-copy ribosomal DNA locus. *Biochemical and biophisical research communications* 446, 261-266.

Xie, K., Minkenberg, B. and Yang, Y. (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences of the United States of America* 112, 3570-3575.

Xiong, A.S., Peng, R.H., Zhuang, J., Liu, J.G., Gao, F., Chen, J.M., Cheng, Z.M. and Yao, Q.H. (2008) Non-polymerase-cycling-assembly-based chemical gene synthesis: strategies, methods, and progress. *Biotechnology advances* 26, 121-134.

Xu, M., Zhu, L., Shou, H.X. and Wu, P. (2005) A PIN1 family gene, OsPIN1, involved in auxin-dependent adventitious root emergence and tillering in rice. *Plant cell physiology* 46, 1674-1681.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S. and Lim, W.A. (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339-350.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M. and Arlotta, P. (2011a) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology* 29, 149-U190.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X.H., Pierick, C.J., Dobbs, D., Peterson, T., Joung, J.K. and Voytas, D.F. (2010a) High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America* 107, 12028-12033.

Zhang, G.J., Guo, G.W., Hu, X.D., Zhang, Y., Li, Q.Y., Li, R.Q., Zhuang, R.H., Lu, Z.K., He, Z.Q., Fang, X.D., Chen, L., Tian, W., Tao, Y., Kristiansen, K., Zhang, X.Q., Li, S.G., Yang, H.M., Wang, J. and Wang, J. (2010b) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome research* 20, 646-654.

Zhang, Y., Kiba, A., Hikichi, Y. and Ohnishi, K. (2011b) prhKLM genes of *Ralstonia solanacearum* encode novel activators of hrp regulon and are required for pathogenesis in tomato. *Fems microbiology letters* 317, 75-82.

Zhang, Z. and Krainer, A.R. (2004) Involvement of SR proteins in mRNA surveillance. *Molecular cell* 16, 597-607.

Zhong, X.Y., Wang, P.P., Han, J.H., Rosenfeld, M.G. and Fu, X.D. (2009) SR proteins in vertical integration of gene expression from transcription to rna processing to translation. *Molecular cell* 35, 1-10.