AN EXPANDED TOOLKIT FOR GENOME EDITING

by

JACOB WILLIAM HOYLE

(Under the Direction of Wayne Allen Parrott)

ABSTRACT

CRISPR utilizes endonuclease proteins like Cas9 guided by engineered RNA to create breaks in target DNA. Each Cas9 ortholog only cuts targets near a particular sequence motif, such as *Streptococcus pyogenes* Cas9's -NGG requirement. These requirements mean CRISPR lacks the flexibility of the older technologies. To increase available targets, four alternative nucleases were utilized. These are the Cas9s from *S. pyogenes*, *Staphylococcus aureus*, and *Streptococcus thermophilus*, and Cpf1 from *Acidaminococcus spp*. Nucleases were cloned into an expression cassette of native legume sequences and fitted with a novel nuclear localization signal from *Glycine max*. CRISPR nucleases from *S. pyogenes*, *S. aureus*, and *S. thermophilus* were confirmed to edit hairy root cultures. Hairy root editing efficiency is low with the plant-derived expression cassette, but is sufficient for editing in somatic embryos, as *S. pyogenes* has produced T0 plants with bi-allelic mutations.

INDEX WORDS: Cisgenics, CRISPR, Genome Editing, *In Vitro* Biology, Plant Molecular Biology, RGEN, Soybean, Transgenic Breeding

AN EXPANDED TOOLKIT FOR GENOME EDITING

by

JACOB WILLIAM HOYLE

B.S., Appalachian State University, 2013

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2017

AN EXPANDED TOOLKIT FOR GENOME EDITING

by

JACOB WILLIAM HOYLE

Major Professor:	Wayne Allen Parrott
Committee:	Robert J. Schmitz
	Chung-Jui Tsai

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

ACKNOWLEDGEMENTS

Gary Orr, Sabrena Rutledge, Virgille Sonon, and Alex Vincell contributed labor towards the completion of this project. Thank you to Dr. Edward McAssey and Dr. Peter LaFayette for reviewing this document. To Dr. Wayne Parrott, thank you for allowing me to believe that the ideas contained in this volume are my own.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

CRISPR Genome Editing

In a post-genomic era, gene discovery is a cornerstone of plant research. Ongoing projects in mutagenesis and QTL discovery need efficient and reproducible ways to fully knockout genes (Bolon et al., 2011; Cooper et al., 2008; Cui et al., 2013; Hancock et al., 2011). Virus Induced Gene Silencing (VIGS) (Burch-Smith et al., 2004) offers a solution by forcing a plant to repress one or more of its own genes in response to an engineered viral vector (Liu et al., 2002; Ratcliff et al., 2001; Ruiz et al., 1998). Another option is RNA interference, which induces gene silencing by recruiting Argonaute proteins to cleave a transcript of interest (Esvelt et al., 2013; Jacobs et al., 2016; Tricoli et al., 1995; Zetsche et al., 2015). These methods are efficient at knocking-down some transcripts, but are not necessarily reproducible due to partial silencing and heritability issues.

Another strategy is genome editing, and CRISPR is the newest of these methods yet available. The CRISPR/Cas9 genome editing system uses an RNA-guided endonuclease (RGEN) to cut DNA at genomic positions homologous to a "spacer" or "target" region of an engineerable RNA guide (gRNA). This gRNA complexes with an endonuclease and creates a double stranded break (DSB) at the site of binding (Jinek et al., 2012)(figure 1.1). In plants, DSBs are most frequently repaired by Non-Homologous End Joining (NHEJ) which often creates insertions or deletions (indels) near the site of

the break (Puchta, 2005). Indels created in specific genes can be of interest to both plant breeding and genetics research. However, not every genomic region can be targeted by the CRISPR system. To create a DSB at a sequence homologous to the gRNA, the endonuclease must recognize a protospacer-adjacent motif (PAM). PAM sequences are specific to the endonuclease, and different for each bacterial species that uses CRISPR (Jinek et al., 2012; Ran et al., 2013; Shah et al., 2013).

CRISPR systems used for genome editing are repurposed from the CRISPR immune systems built into prokaryotic genomes for defense (Garneau et al., 2010; Horvath and Barrangou, 2010). In these native systems, multiple proteins are required to work together to perform the many functions that initiate DSB in the phages attacking CRISPR hosts. CRISPR been optimized so that the fewest number of components are sufficient. For example, CRISPR model *Streptococcus pyogenes* natively uses RNaseIII and Csn1 to produce a CRISPR RNA (crRNA) (Deltcheva et al., 2011), but it is possible to generate single-stranded gRNA (sgRNA) for use in genome editing without either of these components, thus simplifying the system (Jinek et al., 2012).

CRISPR prokaryotic defenses are divided into a series of classes, types, and subtypes to reflect their modes of action, evolutionary history, and complexity. Class 1 CRISPR-Cas systems can use up to 11 proteins to perform CRISPR-RNA (crRNA) binding, target binding, dsDNA unwinding, and DNA cleavage in large multiprotein complexes. With over 11 separate parts, Class 1 CRISPR is too complex to use for genome editing. Despite the intricacy of the Class 1 complexes, more than 70% of sequenced bacterial genomes containing a CRISPR immune system utilize these defenses (Makarova et al., 2015). The simpler CRISPR system performs crRNA and target

binding, as well as target cleavage, all in one large protein called Cas9 or Cpf1. Other proteins are involved in crRNA processing and collection, but target recognition and nuclease activity occurs in one RNA Guided Endonuclease (RGEN). Therefore, Class 2 systems are best suited for utilization in genome-editing work because the only components required are endonuclease and an RNA guide transcribed by RNA pol III.

The 2 classes of CRISPR systems have been subdivided into types, of which types II and V are the most attractive for use in genome editing. Type II systems perform target recognition and cleavage using the protein Cas9 in the presence of a triple-hairpin and trans-activating CRISPR-RNA (tracrRNA) with RNase III (Deltcheva et al., 2011). Unlike Type II systems, Type V systems use Cpf1 proteins for target recognition, binding, and cleavage, with shorter and simpler crRNAs and no trans elements (Zetsche et al., 2015).

RNA Guided Endonucleases

Cas9 and Cpf1 proteins from different bacteria bring different advantages to plant genome editing. The classical Cas9 RGEN was isolated from *Streptococcus pyogenes* (SpyCas9) and adopted across biological fields due to its presumed wide-targeting capabilities imbued by the relatively frequent PAM sequence -NGG (Jinek et al., 2012). This requirement, combined with the selected RNA pol III's 5' G requirement for transcription initiation with U6 promoters, means that SpyCas9 targets are defined as GN(19)NGG.

GN(19)NGG targets, while common, are not always present in genomic regions that will most likely lead to gene knockouts, such as regions preceding important

functional domains. When potential targets are available, they may be repeated elsewhere in the target organism's genome. If an experiment requires that paralogs remain unaffected by gene editing, a target must be chosen that is highly specific to the chosen locus. This consideration is especially pertinent to plant genomes, many of which are characterized by multiple polyploidy events (Soltis et al., 2009). Access to alternative RGENs with varied PAMs will give researchers more potential targets, increasing specificity to target loci.

The four RGENS selected for study are the Cas9 from *S. pyogenes* (SpyCas9), the Cas9 from *Staphylococcus aureus* (SaCas9), the first Cas9 from *Staphylococcus thermophilus* (St1Cas9), and the Cpf1 from *Acidaminococcus sp.* (AsCpf1). These proteins were selected based on their diverse PAM recognition capabilities, sizes, and specific advantages (Table 1.1).

SpyCas9, like all the Cas9 proteins selected, is a Class 2 RGEN of type II (Makarova et al., 2015). The coding sequence of SpyCas9 is 4,104 bp, and is the longest studied, most utilized, and best understood RGEN in Class 2 CRISPR systems (Cradick et al., 2013; Nishimasu et al., 2014; Semenova et al., 2011; Sternberg et al., 2014; Xie and Yang, 2013). Soybean (*Glycine max* (L.) Merrill) embryos have been edited with SpyCas9 following biolistic transformation to integrate Cas9 and gRNA sequences, the two components required for CRISPR edits (Jacobs et al., 2015; Li et al., 2015).

S. thermophilus a facultative anaerobe widely used in the dairy industry where, unlike many of its *Streptococci* relatives, it has no use for pathogenicity. Following adaptation to the dairy environment, its virulence loci have pseudogenized and there is no record of the bacteria causing disease (Bolotin et al., 2004). As such, engineering with the

St1Cas9 gene does not create transgenic plants subject to regulation under the Federal Plant Pest Act. Under the Plant Protection Act of 2000 and the Code of Federal Regulations, plants "modified to contain… genetic material from animal or human pathogens" are not eligible for BRS shipping notifications, but instead require complete BRS Permits to be approved before inter-state transfer (§340.3(b)(6)). The human codon-optimized St1Cas9's recognition motif is N(20)RGAA and the coding sequence is 3,362 bp without a nuclear localization signal (NLS) (Esvelt et al., 2013; Leenay et al., 2016).

S. aureus is the opportunistic pathogen that codes for the shortest Cas9 gene yet discovered at the outset of this study. The human codon-optimized version of SaCas9 comes in at only 3,159 bp before addition of nuclear localization signal. Small size is attractive when considered in conjunction with a viral transient expression system (Friedland et al., 2015). Viruses have been used to induce post-transcriptional silencing in plants (Liu et al., 2002; Pflieger et al., 2014), and a virally delivered CRISPR construct could one day produce heritable knockouts in the same period. Gemini viruses have limited capacity due to coat protein restrictions (Gilbertson et al., 2003), so future CRISPR work with viral systems will benefit from the smallest RGENs available. Smaller sized plasmids also increase the ease of cloning and transformation. SaCas9 has been shown to edit various sequences of GN(20)NNGR (Leenay et al., 2016), but the more conservative GN(20)NNGRRT PAM was used for this study (Ran et al., 2015).

AsCpf1 was isolated from genus *Acidaminococcus.* Many members of this genus form part of the human gut microbiome and are opportunistic pathogens (Chatterjee and Chakraborti, 1995). Cpf1 is a Class 2 RGEN similar to Cas9, but has been categorized as Type V due to several important differences (Makarova et al., 2015). AsCpf1 recognizes

crRNA directly, and does not need the trans-acting elements that Cas9 requires to recognize its guides (Zetsche et al., 2015). The implication for genome editing is that gRNA lengths can be much shorter at only 42 nt long, verses 103 nt for SpyCas9 (fig. 3). That advantage is magnified in multiplex gene editing, where multiple guides are required. Cpf1 has the ability to process a single crRNA array into individual crRNAs for gene editing all under one promoter (Zetsche et al., 2017), which represents a 1,728 bp decrease for a 5-target multiplex construct, less than a quarter of the length that would be used for a traditional construct (fig. 2).

The recognition sequence of AsCpf1, N(20)NTT, is as short as SpyCas9's N(20)NGG (Leenay et al., 2016). Simple recognition motifs like these should be plentiful in genomes, but AsCpf1's target recognition in AT-rich regions could have implications for targeting gene promoters. cDNA in plants typically has higher GC content than intergenic space (Schwarzacher et al., 1997), so access to the AT-targeting AsCpf1 and the GCp-targeting SpyCas9 will allow for better genome coverage. The human-codon-optimized version of AsCpf1 is 3,921 bp without the NLS (Zetsche et al., 2015). Another advantage of AsCpf1 is its offset, or staggered cut-sites. The protein creates a double stranded break with a 4-, sometimes 5-, bp overhang (Zetsche et al., 2015). These overhangs prevent deletions < 5 bp from occurring (Hu et al., 2016). In contrast, the use of Cas9 primarily results in small deletions (Shan et al., 2013). , and has been shown to edit plant cells (Tang et al., 2017).

These 4 RGENs recognize different PAM signals, but they also complex with crRNA/tracrRNA in different configurations in their native systems. As SpyCas9's tracrRNA has been shortened and optimized as a single-molecule gRNA (Jinek et al.,

2012), so have the SaCas9 and St1Cas9 gRNAs (Esvelt et al., 2013; Ran et al., 2015)(fig. 3).

The selected RGENs have been shown to create knockouts in plants (table 1.1) but only SpyCas9 has been verified to have activity in soybean to date.

Pathogen-Free Vectors

Plants produced through genetic engineering are designated regulated articles by the Animal and Plant Health Inspection Service (APHIS) whenever the plants contain sequences cloned from plant or animal pests and pathogens. Sequences derived from all viruses are regulated as plant pests, as well as sequences from *Agrobacterium.* An alternative is that edits can be made with CRISPR, followed by the removal of vector sequences through progeny segregation after crosses.

To decrease the paperwork burden and enhance ability to distribute RGEN-containing plants, transformation vectors need to be free of pathogen-derived sequences. In previous CRISPR research in soybean, 3 out of the 4 sequences pertinent to expression and nuclear targeting of Cas9 were derived from pathogens. These were CaMV 35S promoter, SV40 NLS, *Agrobacterium's* Nos terminator. The plant-derived *Medicago* U6 promotor (MTU6) was utilized as well, and will be kept (Jacobs et al., 2015; Kim and Nam, 2013). Development and demonstration of a pathogen-free vector system will relax the permitting process for all plants developed from the system. The plasmids generated will contain cloning sites between major sequences, so it will also benefit the future deployment of future RGENs derived from non-pathogens, as the modular vector system makes RGEN swapping quick and uncomplicated.

The RGEN promoter selected for the pathogen-free or plant-derived cassette is GmUbi-3 P, which has high expression levels in embryo (Hernandez-Garcia et al., 2010; Hernandez-Garcia et al., 2009). The selected terminator is the *Pisum sativum* ribulose-1,4-bisphosphate carboxylase's 3'UTR (RbcsT) (Coruzzi et al., 1984) as utilized in pKYLX7 (Schardl et al., 1987). RGENs require an NLS for protein targeting to the nucleus where they interact with genomic DNA, but plant derived NLSs are generally not used in CRISPR-edited transgenic plants. A soybean putative NLS from the E1 flowering time locus was selected for further study. The backbone of the transformation vectors is based on pUGA, which uses the pBR322 origin of replication (Thomson et al., 2002; Watson, 1988). The pBR322 sequences, derived from *Escherichia coli* strain K-12 derivatives, are not regulated as plant pests (7 CFR 340).

An objective is to move 4 different Cas9 or Cpf1 encoding genes into a toolkit with enough PAM diversity that any genomic region can be edited.

The CRISPR constructs were evaluated in hairy roots, which requires the use of *Agrobacterium rhizogenes.* A vector series without *Agrobacterium* sequences can still be used with particle bombardment. A CRISPR construct completely free from *Agrobacterium* sequences needs to be evaluated in the context of biolistic transformation. To this end, target genes with visual knockout phenotypes are desirable for CRISPR genome editing.

Another objective is to verify the activity of CRISPR in plant-derived cassettes within the somatic embryo and biolistic transformation protocol. The target gene for biolistic CRISPR editing was selected from a short trichome mutant in a soybean fast neutron population developed at the University of Minnesota from variety 'MN1302'

(Bolon et al., 2011; Orf and Denny, 2004). The mutation was mapped to a region on chromosome 6 that covers a homolog for *Arabidopsis Cpr5* (Dr. Robert Stupar, pers. comm., unpublished). *Arabidopsis* mutants of Cpr5 have a deformed trichome phenotype (Bowling et al., 1997). The gene could be agronomically important as it has been implicated as a modulator of the growth-defense balance, such as in the ethylene pathway (Wang et al., 2017), effector-triggered immunity (Wang et al., 2014), and the unfolded protein response (Meng et al., 2017).

Table 1.1: RGENs selected for the study and details

.RGEN	Donating Bacteria	Donating Publication	Codon Opt.	Length (no NLS)	PAM	Use in Plants	Reported Efficiency
SpyCas9	*Staphylococcus pyogenes*	Mali et al 2013	human	4104 bp	-NGG	Shan et al. (2013)	9.40%
						Zhang et al. (2014)	21.1-66.7%
SaCas9	*Staphylococcus aureus*	Ran et al 2015	human	3159 bp	-NNGRRT	Kaya et al. (2016)	87.5-93.7%
						Steinert et al. (2015)	68.8-98.5%
St1Cas9	*Streptococcus thermophilus*	Esvelt et al 2013	human	3362 bp	-NNRGAAW*	Steinert et al. (2015)	6.1-77.7%
AsCpf1	*Acidaminococcus* sp.	Zetsche et al 2015	human	3921 bp	TTTN-	Tang et al. (2017)	4.3-68.2%

* not -NNGGAAG or NNGGAAC (Leenay et al.)

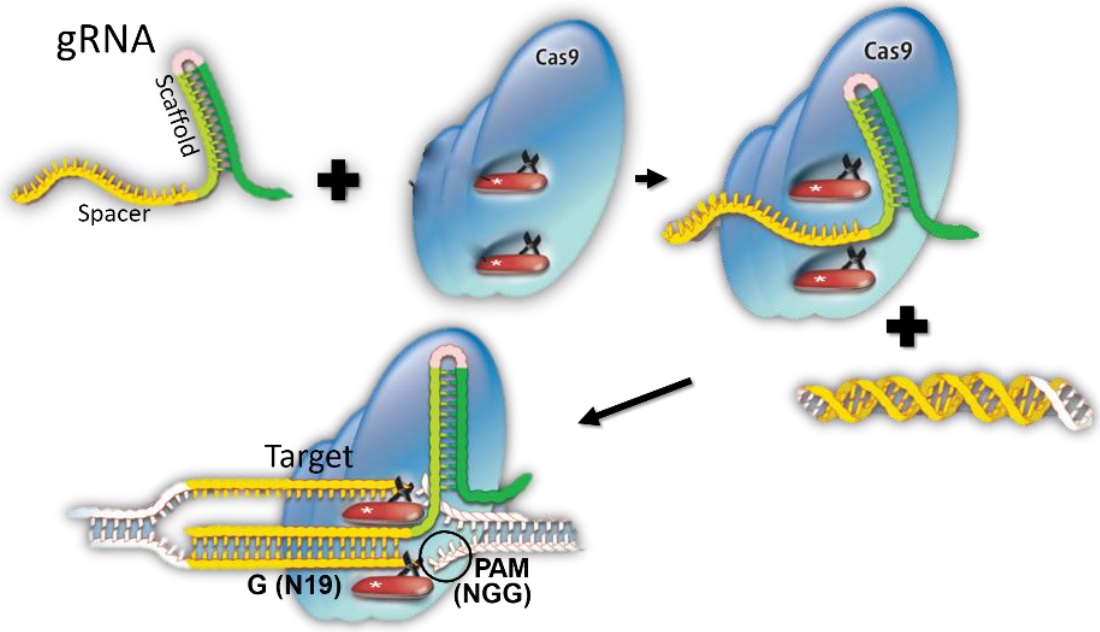

Figure 1.1: Overview of CRISPR Genome Editing. The Cas9 RGEN and gRNA complex and together create DSBs. Adapted from K. Sutliff/Science.

Figure 1.2: Size comparison of multiplexed gRNA expression regions in SpyCas9 and AsCpf1 CRISPR constructs. RNA production for four targets requires 1,820 bp for SpyCas9, and only AsCpf1 requires only 508.

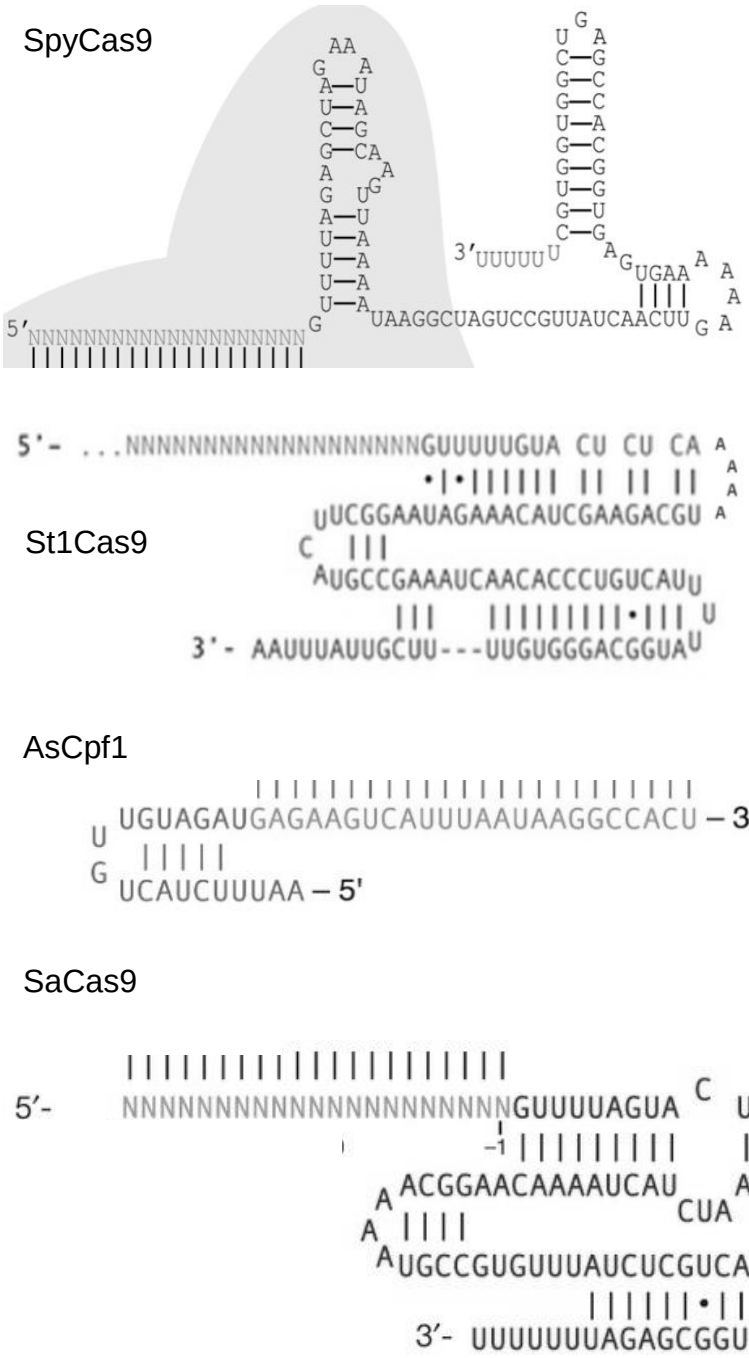

Figure 1.3: Optimized sgRNA sequences and secondary structures for 4 RGENs (DiCarlo et al., 2013; Esvelt et al., 2013; Ran et al., 2015; Zetsche et al., 2015)

CHAPTER 2

Glycine max E1 NUCLEAR LOCALIZATION SIGNAL

Introduction

CRISPR genome editing takes place in the nucleus, but when inserted as transgenes, RGENs are produced outside of the nucleus like all proteins. A popular choice for nuclear localization is a 7 amino acid sequence from the large T antigen of Simian Virus 40 (Kalderon et al., 1984). Codon optimized versions of SV40-NLS have been used in the most influential CRISPR studies in plants (Feng et al., 2013; Shan et al., 2013; Wang et al., 2014; Zhang et al., 2014), including the soybean work on which this study was based (Jacobs et al., 2015). Simian Virus 40 is an oncovirus (Soriano et al., 1974), therefore transgenic plants containing its sequences are regulated articles under 7CFR340. A plant sequence would be a preferable alternative for nuclear localization.

NLSs are peptide motifs in proteins that bind to and facilitate transport through importin proteins. NLSs are categorized into 6 classes, the sixth of which is bipartite, which are characterized by two Lysine and Arginine rich motifs spaced 10-20 residues apart (Kosugi et al., 2009). Multiple tandem SV40-NLSs have been shown to increase nuclear localization rates, and a bipartite NLS could potentially offer a similar rate of localization (Girard et al., 2004).

E1 is a maturity locus in soybean that was recently mapped to Glyma06g23040 in Glycine max Williams 82 v.1 (Bernard, 1971; Xia et al., 2012). The link between this

gene and flowering time is tenuous; the Williams 82 *E1* gene represented by Glyma06g23040's is the recessive form, and was removed from the second genome assembly, replaced by the much longer coding sequence Glyma.06g207800. The genetic story of *E1* and chromosome 6 is still being written, but important to this study is how it was shown to localize to the nucleus. To determine *E1*'s potential as a transcription factor, a cellular localization study was performed on *E1* and its two alternative allelic forms. The three forms of the gene localized to the nucleus at varying degrees, with the *E1* allele producing the strongest nuclear signal (Xia et al., 2012). *E1* contains a lysine-arginine rich sequence near the N-terminus (figure 2.1), but it remains to be seen if the soybean putative NLS alone is sufficient for nuclear localization.

Materials and Methods

Vector Construction

GmE1-NLS was synthesized as a 63-nucleotide single-stranded oligo with 3' homology to the 3' end of sGFP, so that the oligo could be used as a reverse primer for GFP amplification (figure 2.1). The 5' end of 63 NT oligo called sGFP-e1L-R contained a *Pst* I restriction site for cloning into a transformation vector. Similarly, the forward primer, sGFP-F adds a 5' *Nco* I restriction site. The transformation vector, p201sGFP, is a modified pZP-P201BK (Covert et al., 2001) where kanamycin resistance is replaced with sGFP under the CsVMV promoter and Nos terminator (Li et al., 2001; Mitsuhara et al., 1996; Pedelacq et al., 2006).

The p201 backbone series utilizes pVS1 *Sta* and *Rep* genes (vanderBij et al., 1996), as well as pBR322 *Bom* and *Ori* (Finnegan and Sherratt, 1982). The sGFP primer

set (figure 2.1) was used to amplify the marker gene from p201sGFP, then both plasmid and amplicon were digested with *Pst* I and *Nco* I in a double digestion. Then, p201G's 5' phosphate group was removed with TSAP Thermosensitive Alkaline Phosphatase. Vector and insert were assembled with T4 Ligase. Plasmid 201sGFP and p201sGFP:e1NLS are identical aside from the NLS fused to the reporter (figure 2.2, 2.3).

Onion Dermal Assay

Biolistic transformation was performed with the PDS-1000/He Particle Delivery System and microcarrier preparation executed similarly to the PDS-1000 protocol described in Trick et al. (1997) with modifications. Four hundred fifty ng of plasmid DNA were precipitated on 10 mg of 0.6-um diameter Au particles. Four days before shooting, onion dermal tissue was removed from one white onion obtained from Kroger (The Kroger Company, 1883) in a sterile hood using scalpel and forceps, and dissected to lay flat on MSO-lite medium (Trick et al., 1997). The tissue was left uncovered for 20 minutes before shooting. Pressure and microcarriers were released at 1100 psi. Cells were imaged 20 hours after shooting in a Zeiss Axio Imager 2 using AxioVision v4.8 software.

Results and Discussion

Eight GFP and 18 GFP:E1NLS cells were initially imaged, and the experiment was repeated with a similar number of images. Cells transformed with p201sGFP clearly demonstrate a fluorescent signal in the nucleus, cytoplasm, and cell membrane, but not in the vacuole. This is consistent with non-localized fluorescence. Meanwhile, in cells transformed with p201sGFP:E1NLS, fluorescence is found only in the nucleus (figure

2.4). This result suggests the 63-residue sequence at the N-terminus of Glyma.04g156400 is a bipartite nuclear localization signal, and is sufficient for localizing transgenes to the nucleus in plants.

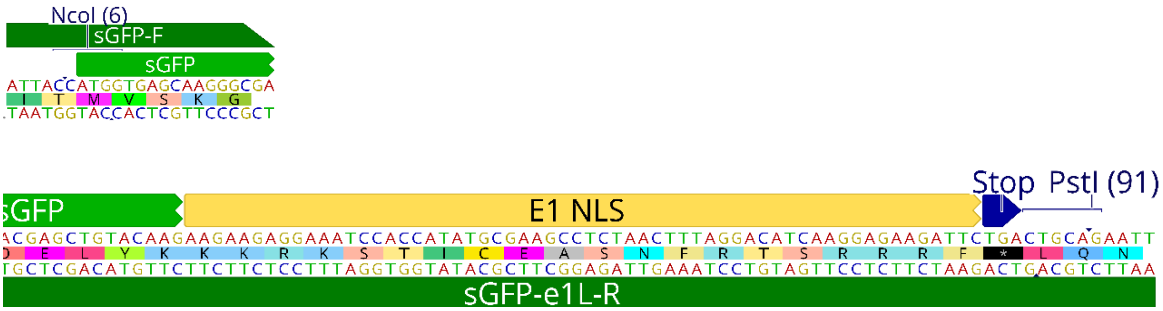

Figure 2.1: Primers used for cloning. Putative bipartite NLS is indicated in yellow

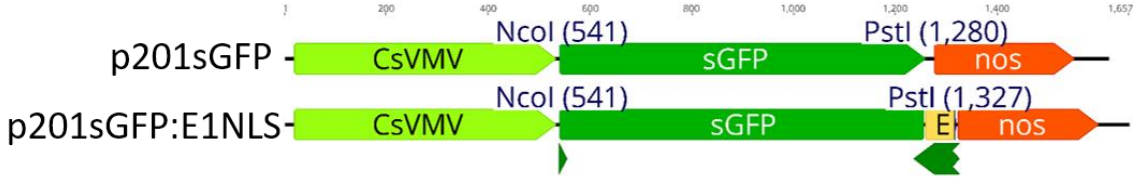

Figure 2.2: Detail of sGFP cassettes and NLS fusion

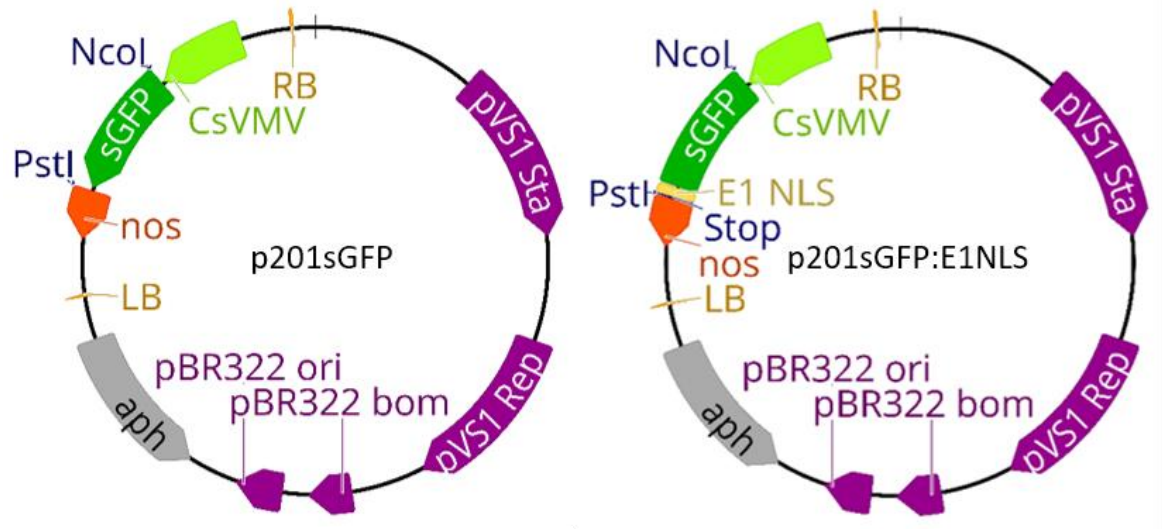

Figure 2.3: Plasmids used for localization experiment

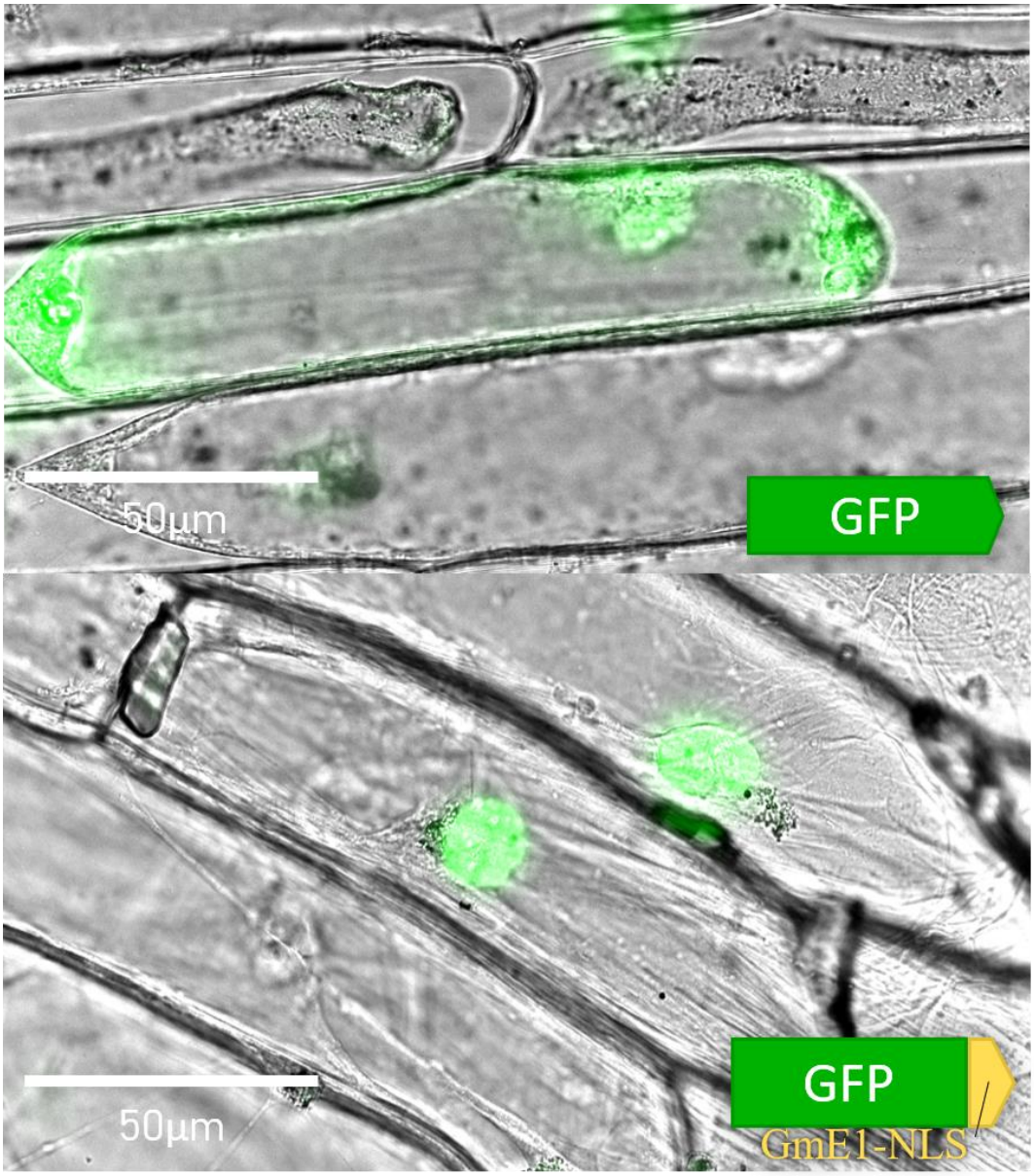

Figure 2.4: Subcellular localization of sGFP and sGFP:E1NLS. Images are merged bright-field and GFP fluorescence.

CHAPTER 3

VECTORS

Introduction

An objective of the project was to build a toolkit, with the intention of providing modular, ready-to-use plasmids through the Addgene service as a resource for plant geneticists. Three types of vectors were used: Biolistic transformation plasmids, *Agrobacterium* transformation plasmids, and shuttle vectors. Biolistic transformation vectors contain the selectable marker for hygromycin, and *E. coli* replication regions. *Agrobacterium* transformation vectors contain border repeats, an antibiotic selectable marker, and both *E. coli* and *Agrobacterium* replication regions. Shuttle vectors are small plasmids suited for replication in *E. coli.* The shuttle vector used was pSce with an ampicillin resistance marker (Thompson and Parrott, 1998).

Biolistic transformation vector pStu is an adaptation of the pMECA (Thompson and Parrott, 1998) vector after the addition of *hph* for hygromycin resistance. *Hph* is under the control of the *Solanum tuberosum* Ubiquitin 3 (Stubi-3) promoter and terminator (Garbarino and Belknap, 1994; Thompson and Parrott, 1998).

The binary vectors used for *Agrobacterium* transformation were built on the p201 vector series, which utilizes pVS1 *Sta* and *Rep* genes (vanderBij et al., 1996), as well as pBR322 *Bom* and *Ori* (Finnegan and Sherratt, 1982). A p201 plasmid with *nptII* for G418

resistance under Stubi-3 promoter and terminators (Garbarino and Belknap, 1994) called p201N is well suited for hairy root transformations.

Materials and Methods

First, plant derived cassettes were assembled in the plasmid shuttle vector pSce. A soybean ubiquitin-3 promoter (GmUbi-3 P) was amplified from pGmubi (Hernandez-Garcia et al., 2009) with a reverse primer that added a destination vector overhang for In-Fusion© cloning (Clontech) and a forward primer that added *Avr* II and *Nhe* I restriction sites as well as an overhang for the selected terminator. The selected terminator, RbcsT, was amplified from pGmute (Jacobs et al., 2016) with similar In-Fusion© primers. The shuttle vector pSce was linearized with I-*Sce* I. An In-Fusion© cloning reaction followed by ligation with T4 ligase completed the plasmid pSceGmu (figure 3.1).

Biolistic CRISPR Vectors

The Gmubi-3 P RbcsT cassette was moved into the destination vector pStu with traditional cloning. Gmubi-3 P-> *Avr* II -> *Nhe* I -> RbcsT was isolated by digesting pSceGmu with I-*Sce* I and gel purifying with Zymoclean™ Gel DNA Recovery Kit. Destination vector pStu was linearized with I-*Sce* I and the two fragments were joined with T4 Ligase creating pStuGmu (figure 3.1).

RGEN DNA was obtained from Addgene. SpyCas9 is in plasmid #41815, SaCas9 #61592, St1Cas9 #48669, AsCpf1 #69982 (Esvelt et al., 2013; Mali et al., 2013; Ran et al., 2015; Zetsche et al., 2015)(table 1).

RGENs were amplified using the HiFi HotStart ReadyMix by Kapa Biosystems, using forward primers designed with Gmubi-3 P overhangs and reverse primers with RbcsT overhangs. pStuGmubi was linearized with *Avr* II and *Nhe* I, and an NEBuilder (New England BioLabs Inc.) reaction inserted the RGENs into the cassette. These reactions created the pStuRGEN series (figure 3.1). RGEN cassettes were Sanger sequenced to confirm that errors were not introduced during PCR amplification.

Hairy Root Transformation Vectors

A forward primer was designed with homology to 5' of Gmubi-3 promoter with a 15 bp 5' overhang homologous to the destination backbone p201Npt at the I-*Sce*I site. A reverse primer was designed to amplify the 3' of RbcsT with 5' homology to the p201Npt as well. This pair was used to amplify RGEN cassettes from the pStuRGEN series with Kapa HiFi HotStart ReadyMix. The intended backbone p201Npt was linearized with I-*Sce*I, then the amplified RGEN cassettes were inserted with NEBuilder (New England BioLabs Inc.)® HiFi DNA Assembly kit, creating the p201N-RGEN series (figure 3.1). RGEN cassettes were Sanger sequenced to confirm that errors were not introduced during amplification.

Guide RNA cassettes were constructed similarly, but with varying complexity. In each case, MtU6 was amplified from a gRNA shuttle vector (Jacobs et al., 2015) with a forward primer adding 5' homology to p201NRGEN at the *Swa* I cut site. The Spy Cas9 scaffold region was PCR-amplified from a gRNA shuttle (Jacobs et al., 2015) with a reverse primer that adds an overhang to p201NSpyCas9 at the *Spe* I cut site. Optimized sgRNA and crRNA scaffolds for SaCas9 and St1Cas9 (figure 1.3) were synthesized as

single stranded oligonucleotides with 3’ overhang homology to p201N-RGEN at the *Spe* I cut site. St1’s long scaffold length was synthesized in two pieces (figure 3.2). Spy, Sa, and St1 selected gRNAs were synthesized with overhangs for both MtU6 and the respective scaffolds, taking care to complement the other single-stranded oligonucleotides (figure 3.2). AsCpf1’s scaffold is short enough that the target and scaffold was ordered as one single-stranded oligonucleotide (figure 3.2). The p201N-RGEN plasmids were linearized with *Spe* I and *Swa* I. MtU6, gRNA, and scaffolds were incorporated in single NEBuilder (New England BioLabs Inc.) reactions. These final plasmids were named with a convention that specifies their backbone, RGEN, and target, such as p201NSpyCas9xT2.

NLS Swap Vectors

NLS swap plasmids were constructed from p201NSpyCas9x2015 and p201N-35S-SpyCas9x2015 (Jacobs et al., 2015). A single forward primer was designed to amplify the 3’ end of SpyCas9 starting at the *Pml* I restriction site, with an overhang with homology 5’ of the cut site. Two reverse primers were designed at the end of SpyCas9 before the NLS, one to add E1NLS, and one to add SV40, and homology to RbcsT. The Gmubi & E1 plasmid and the 35S & SV40 plasmids were digested with *Pml* I and *Sac* I to remove the 3’ end of the gene along with their original NLSs, and Kapa HiFi amplicons from the alternating plasmids were inserted with NEBuilder (New England BioLabs Inc.).

Cpr5 CRISPR Construct

pStuHSpyCas9 (figure 3.1) was the basis for the CRISPR construct pStuHSpyC9xCpr5. The MtU6 promoter was amplified with a forward primer that added homology to pStuHSpyCas9 at the *Spe* I restriction site. The optimized *S. pyogoenes* sgRNA scaffold (DiCarlo et al., 2013) was synthesized as a single-stranded oligonucleotide with an overhang at its 3' end with homology to pStuHSpyCas9 at the *Spe* I cut site. The selected *GmCpr5* gRNA was synthesized with overhangs for both MtU6 and the scaffold, taking care to compliment the other single-stranded oligonucleotide. pStuHSpyCas9 was linearized with *Spe* I and MtU6, gRNA, and scaffold were incorporated in one NEBuilder (New England BioLabs Inc.) reaction to create pStuHSpyC9xCpr5 (figure 3.3)

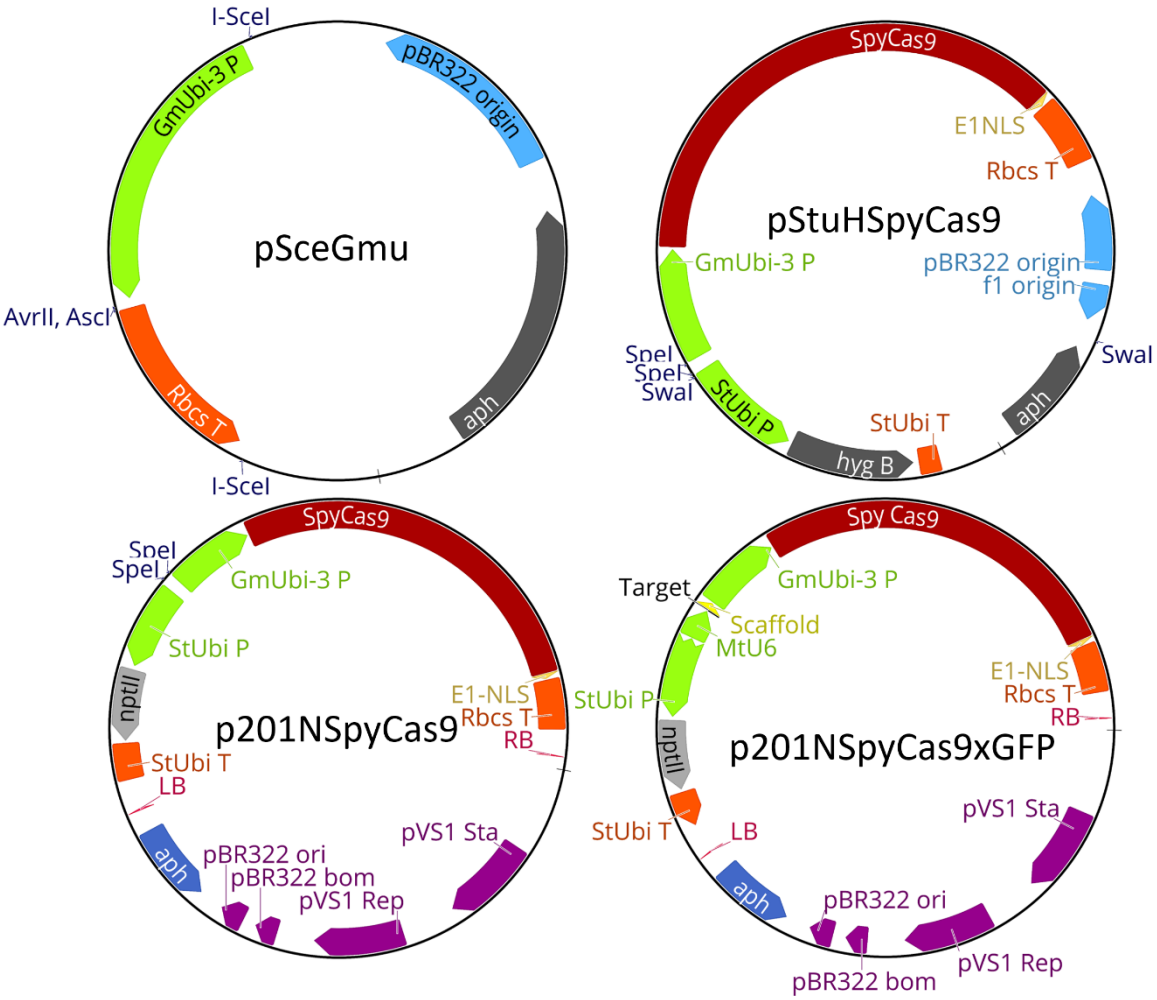

Figure 3.1: Vectors constructed for CRISPR toolkit. Each vector containing SpyCas9 was created with each alternative RGEN as well.

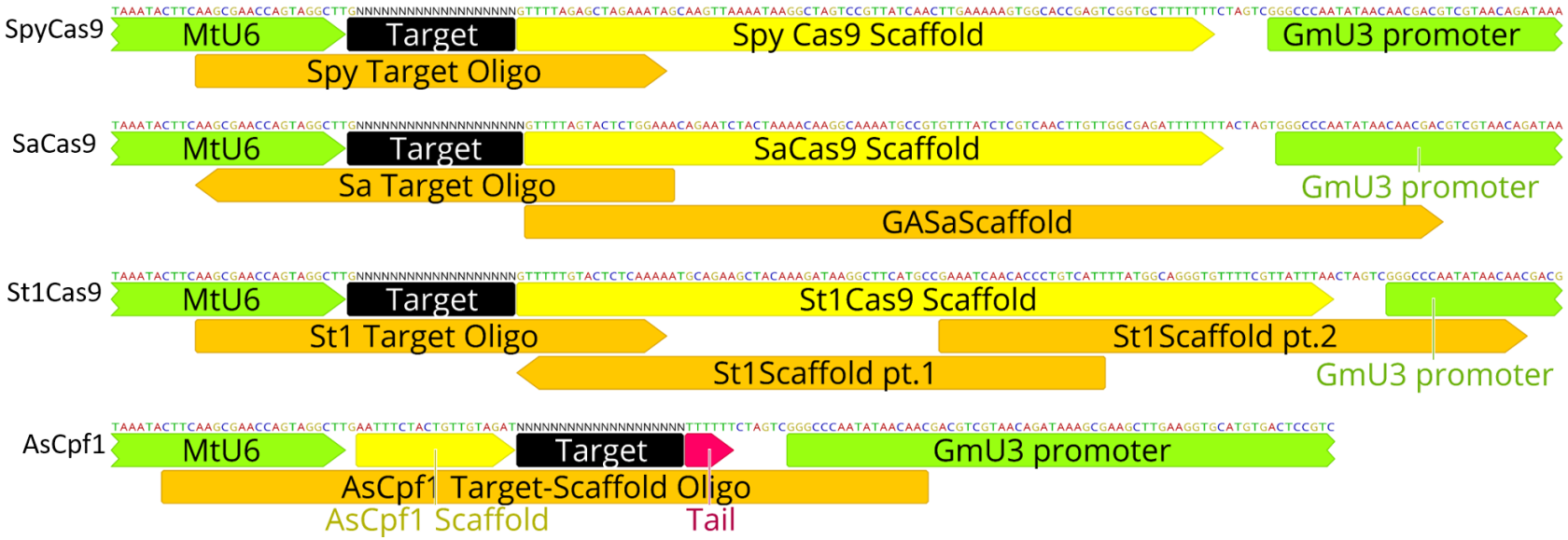

Figure 3.2: Single-stranded oligonucleotides and their assembly in four different guide RNA expression cassettes

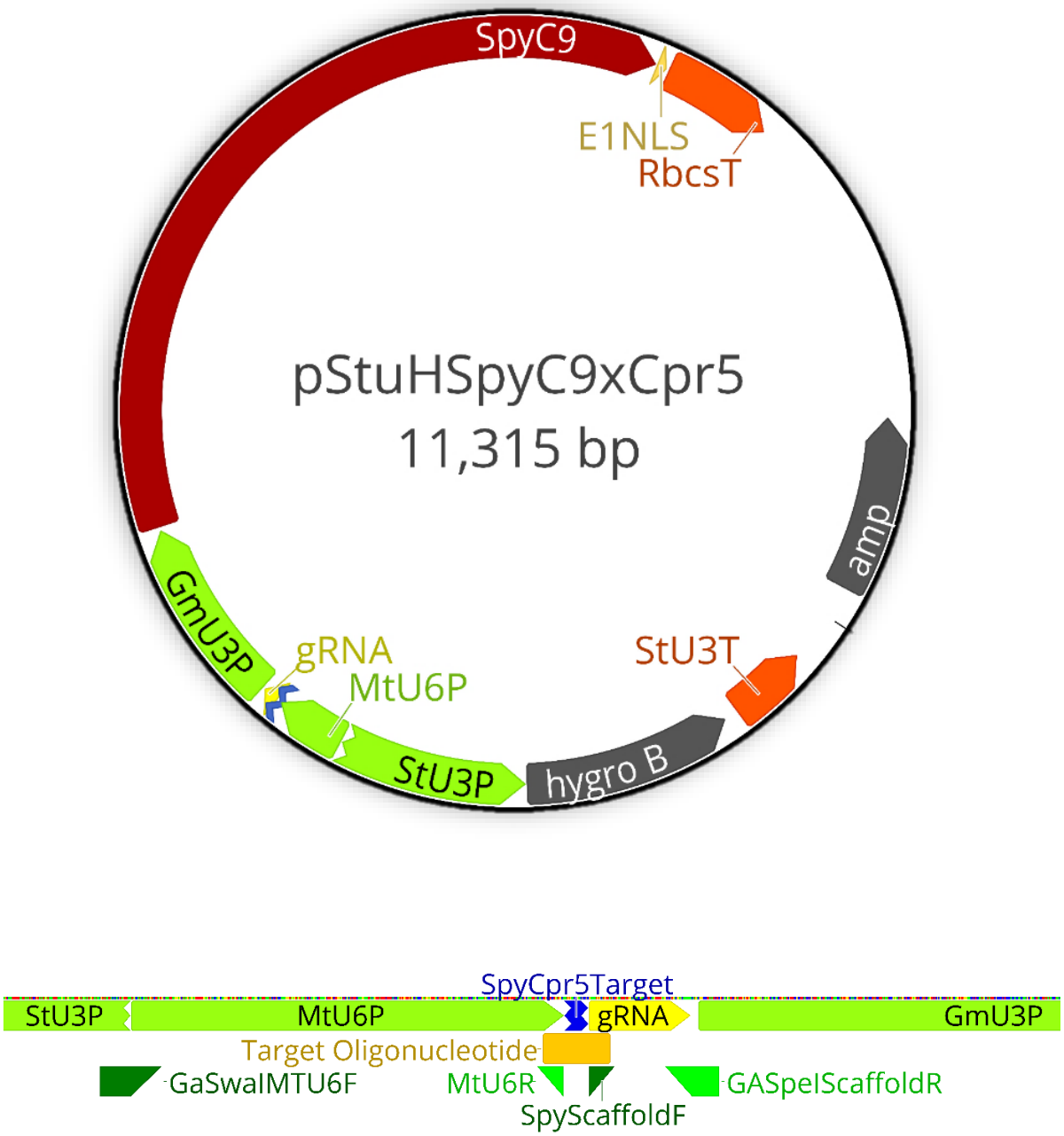

Figure 3.3: *Cpr5* CRISPR knockout construct and detail on the gRNA cassette.

CHAPTER 4

TOOLKIT EVALUATION

Introduction

Soybean transformation is a lengthy process, taking up to 12 months to fully regenerate plants after a transformation event (Trick et al., 1997). A rapid assay is more practical to validate the various CRISPR constructs with new components in plant cells without the lengthy regeneration process. Hairy root transformation with *Agrobacterium rhizogenes* suits the need for an efficient transformation system, where many individual events can be produced from relatively little labor (Cho et al., 2000; Kereszt et al., 2007). This method has previously been utilized to characterize the activity of zinc-finger, TALEN, and CRISPR genome editing in soybean (Curtin et al., 2011; Haun et al., 2014; Jacobs et al., 2015).

Potential targets for CRISPR editing can be evaluated on three criteria: predicted on-target activity, target specificity, and likelihood to cause a functional disruption. The prediction of on-target activity has evolved over recent years. In 2014 a study was conducted on nine mammalian genes with 1,841 gRNAs and the results were used to posit rules governing optimal target sites (Doench et al., 2014). As of version R11, the algorithm described by Doench et al. (2014) is currently the scoring system used by Geneious© software (Biomatters Ltd.). In 2016, a new study was conducted with two massive libraries of gRNAs across the mouse and human genome, generating the most

comprehensive algorithm for scoring editing activity with *S. pyogenes* to date (Doench et al., 2016). In 2017, the group organized the algorithms into a useable webtool called CRISPRko, and have added optimized prediction for *S. aureus* targets as well (unpublished). A multi RGEN activity prediction tool was first published in 2017 and called sgRNA Scorer 2.0 (Chari et al., 2017). This tool used datasets of high and low activity gRNA from *S. pyogenes* Cas9 and *S. thermophilus* Cas9 and produced a model to predict activity on each system. The combined model was confirmed for both *S. pyogenes* and *S. thermophilus* RGENs, and verified to have strong predictive power for several other RGENs as well, including those selected for this study.

Target specificity, or off-target score, is calculated by an algorithm proposed in 2013 that counts mismatches between the proposed gRNA and similar sequences elsewhere in the target genome, weighting them based on position and distribution (Hsu et al., 2013). With the algorithm and access to an annotated genome, such as those implemented in Geneious, researchers can check if these predicted off-target sites lie within annotated CDS regions (Geneious R8-R11, Biomatters Ltd.).

The likelihood of causing a functional disruption in a gene is determined by the candidate target's location within the gene, with targets near the 5' end more likely to disrupt gene function (Jacobs et al., 2015). Another important consideration is the location of functional domains. Frame-shift mutations within important functional domains are certain to cause a loss of function, and even single amino-acid changes at the functional domain can cause a functional knock-out. The functional domain in sGFP is a fluorophore centered around AA 63, so mutations at or before this region are certain to cause functional knockout.

CRISPR in Somatic Embryos

The goal of testing alternative RGENs and new cassette configurations is to assess their viability to recover fully edited plants, not just explants. A CRISPR construct pStuHSpyCas9xCPR5 (figure 3.3) was designed to target a gene and transformed into soybean embryo cultures. The gene targeted is Glyma.06g145800, predicted to code for an *Arabidopsis Cpr5* homolog. *Arabidopsis cpr5* mutants display a short trichome phenotype (Bowling et al., 1997).

Materials and Methods

Target Selection

Targets for sGFP were selected for each RGEN based on their PAM signals (Table 1.1). Initial targets (T1s) were designed at the 5' end of sGFP, taking care to keep the targets as close as possible to minimize the predicted loss of efficiency towards the 3' end of a gene (Jacobs et al., 2015)(figure 4.1). Later, when the importance of predicted on-target efficiency was realized, additional targets (T2s, T3s) were selected to control for the effect of differing predicted CRISPR activity, and the 2015 target (Jacobs et al., 2015) serves as a positive control (figure 4.2). On-target activity was predicted with the algorithm published in Chari et al. (2017) (figure 4.2).

A *Cpr5* target was selected in Exon 1 with a high specificity score of 95.24% (Hsu et al., 2013). Two predicted off-target sites are located in CDS regions, but both have mis-matches proximal to the PAM, making these potential sites very unlikely to

form modifications. The predicted activity of the target was 0.384 on a scale from 0 to 1 according to Doench et al. (2014).

Hairy Root Transformation

The p201 vectors were electroporated into *Agrobacterium rhizogenes* strain K599 (Mankin et al., 2007). Hairy root preparation proceeded as described in Cho et al. (2000) with all adaptations described in Olhoft et al. (2003) and Zhou et al. (2015). Split cotyledons were divided between treatments so that each germinated seed produced two experimental units that did not receive the same treatment. Selection was provided by G418 at 10 mg L^{-1}. Hairy roots were considered transgenic events if they grew between weeks 4 and 6 while on G418 medium.

After six weeks, positive transformation events were imaged for fluorescence. Images were captured with a UVP ChemiDoc-It TS3 Imager with a UVP BioLite Xe MultiSpectral light source with 1.71 second exposure and 4x4 pixel binning. Roots were scored on a binary scale for GFP expression.

Somatic Embryo Transformation

Embryo transformation and regeneration is, at ~10 months, a much longer process than the 6 weeks typical of hairy root assays. To obtain results around the same time as the hairy root experiment, embryos were transformed before the results of the hairy root assay could evaluate the validity of the toolkit.

Soybean somatic embryos were transformed by particle bombardment as described in (Trick et al., 1997) with modifications described in (Hancock et al., 2011).

Forty-two ng of pStuHSpyC9xCpr5 was precipitated on 0.6-µm diameter gold particles. Selection followed in liquid FNL medium (Samoylov et al., 1998) with 20 µg mL^{-1} hygromycin-B. Hygromycin-resistant events were verified via Long-Range PCR with Promega GoTaq® Long PCR Master Mix to confirm the presence of Cas9. RT-PCR was used to confirm expression. Positive events were histodifferentiated in SHaM medium for 5 weeks and desiccated for one, then germinated on MSO medium and hardened off in soil boxes (Parrott et al., 1988; Schmidt et al., 2005).

Experimental Design for hairy root transformation

RGENs were tested over 14 treatments. Each RGEN used three sGFP targets, for a total of 12, plus the positive control of the 2015 target with the Jacobs et al. (2015) plasmid, and a negative control of no-plasmid K599 vector. The Spy T1 was discarded after failing to produce hairy roots in initial testing and was replaced with the 2015 gRNA design, so the list of all treatments is p201NSpyCas9x 2015, p201NSpyCas9xGfpT2, p201NSpyCas9xGfpT3, p201NSaurCas9xGfpT1, p201NSaurCas9xGfpT2, p201NSaurCas9xGfpT3, p201NSt1Cas9xGfpT1, p201NSt1Cas9xGfpT2, p201NSt1Cas9xGfpT3, p201NAsCpf1xGfpT1, p201NAsCpf1xGfpT2, p201NAsCpf1xGfpT3, p201N-35S-Cas9x2015, and K599. The experiment was designed in randomized complete blocks. Each block included each of the 14 treatments were stored in the same box during the inoculation and root growth stages. Ten replicate experiments were performed in total. T1, T2, and T3 targets were bulked during analysis.

Initial results were not encouraging. Since the difference to the Jacobs et al (2015) vector was the GmUbi promoter and the E1-NLS, E1-NLS was swapped with the SV40-

NLS to isolate the effect of NLS sequence from promoter sequence. This created four versions of SpyCas9 constructs with the same '2015' gRNA of sGFP, completing the factorial design.

Amplicon sequencing and data analysis

Genomic DNA was purified from 2 cm lengths of hairy roots, first unopened trifoliolates, or 3.5 mg embryo culture samples using the C-TAB method (Murray and Thompson, 1980). to determine the nature (length, % of cells affected) of the edits. To produce amplicons for analysis in TIDE decomposition analysis software, PCR was performed without a proof-reading polymerase in accordance with the TIDE protocol (Brinkman et al., 2014) on *sGfp*, using the primers Gmubi_F(OH) and NosT-Rev. PCR was performed with Apex 2.0X Taq RED Master Mix for 30 cycles following the polymerase protocol, and used an annealing temperature of 58 degrees. Sanger sequencing was conducted through GENEWIZ® with the closest primer to the respective target, either Gmubi_F(OH) or NosT-Rev. The resulting sequences were analyzed using TIDE decomposition analysis software (Brinkman et al., 2014)

Results

Hairy Root Assay

Initial testing of T1 gRNAs (figure 4.1) verified the activity of p201N-35S-Cas9x2015 but failed to produce a useful number of knockouts using constructs containing plant-derived sequence features (figure 4.3). In fact, K599 harboring *Spy* T1 failed to produce more than one hairy root.

The addition of new gRNAs (figure 4.2) produced similarly low editing efficiency in the plant-derived vector series. However, out of the 17 roots tested with TIDE per treatment, two roots from *S. aureus* and one from *S. thermophilus* Cas9 systems were successfully edited based on Sanger sequencing and TIDE analysis (Figure 4.5, 4.6, 4.7). Both of these roots failed to produce a fluorescent signal. No *S. pyogenes* Cas9 or *Acidaminococcus* Cpf1 transformed roots showed editing in TIDE analysis at 6 weeks.

NLS Swap

At 4 weeks, antibiotic resistant roots expressing Cas9 in the 35S+SV40, 35S+E1, and Gmubi+SV40 cassettes displayed loss-of-fluorescence in over 35% of roots. However, the combination of Gmubi+E1 only produced one root without GFP signal out of 31 antibiotic resistant roots (figure 4.7). This rate is consistent with spontaneous silencing in null controls (figure 4.3).

Somatic embryo Editing

At 8 weeks post bombardment, 10 events transformed with the CPR5 CRISPR construct were identified as hygromycin resistant, indicating transformation. Five of those events were PCR-confirmed to contain Cas9, and each of those five events was confirmed to express Cas9. At 10 weeks past bombardment, the five verified events were subjected to TIDE analysis at the *Cpr5* target locus. All of the events were confirmed to have some degree of editing, ranging from 48% to 99% of amplicons edited, with event #5 achieving a 99.3% efficiency, showing a 1 bp deletion (figure 4.8). Two events were lost to contamination, and two more failed to produce roots and/or shoots during germination.

Event #3, which at 10 weeks was edited at 62.1%, was regenerated into five different plants. Two of these plants (xCpr5 #3-1 and xCpr5 #3-2) were sampled again at 20 weeks. Both plants showed >90% editing with all indels representing one of two mutations, suggesting bi-allelic mutation. All five plants were sampled again at 30 weeks, and all 5 showed bi-allelic editing (table 4.1). Plants 3 and 5 have an identical editing pattern, which could suggest that the responsible editing event occurred prior to histodifferentiation.

The percentage of amplicons edited increased with culture time in plant 1 (figure 4.9). Editing patterns in the 30-week samples of plants 1 and 2 were consistent with their patterns at 20 weeks.

At the T0 generation, three out of five plants display the short trichome phenotype (figure 4.12). Plants two and three (xCPR5 #3-2, cCPR5 #3-3) exhibit an intermediate phenotype (figure 4.11). Plants with the intermediate phenotype are the same plants which carry one non-frameshift allele.

Discussion and Conclusions

Many of the results of this study rely on the accuracy and reproducibility of TIDE analysis. Because TIDE analysis relies on PCR amplification of genomic DNA, it is important to verify that the composition of amplicons from one DNA isolation is reproducible. This was verified when the xCPR5 #3-1 line was re-analyzed at 10-weeks, producing the same editing pattern. One bias that may be produced from PCR is that sequences with larger deletions may be over-represented in the final output. Of the bi-allelic mutations shown, the larger deletion (smaller sequence) appears as a larger

percentage of the amplicon sequenced. This is likely an artifact of the shorter sequences beginning to dominate the reaction.

Another consideration is whether or not fluorescence in hairy roots correlates with editing confirmed by sequencing. Figures 4.3 and 4.4 demonstrate the inverse relationship between editing efficiencies and GFP signals across tested constructs. Indeed, the two edited SaCas9 and St1Cas9 roots did lose their fluorescence.

In initial testing of sGFP T1 gRNAs, *A. rhizogenes* transformation with the *S. pyogenes* Cas9 construct produced only one root out of over 20 explants. This was surprising, as the target's specificity score was considered high at 86%. Further investigation revealed that the target bears similarity to four other locations in the soybean genome, each of which have a complete PAM signal and are identical in the important seed region within the first 12 bases. It is conceivable that if these off-target regions were edited, the transformants could be lethal.

No editing events were ever observed with *Acidaminococcus* Cpf1. In accordance with the hairy root transformation protocol, roots were grown at 26 degrees Celsius. A recent study showed that *Acidaminococcus* Cpf1 has a very low activity below 30 degrees and peak activity at 34 degrees (Moreno-Mateos et al., 2017). A previous Cpf1 study in plants showed dramatically decreased nuclease activity with AsCpf1 compared to another RGEN LbCpf1 from *Lachnospiraceae bacterium*, but plants were grown at 25 degrees. Future work with *Acidaminococcus* Cpf1 in plants will need to be conducted at this temperature to expect successful editing. The feasibility of such a study is questionable; Soybean plants survive hot Georgia summers when temperatures reach well

over 34 degrees, but no known hairy-root protocol subjects tissues to the same temperature.

The biggest question raised by this study is why did the cassettes that express Cas9 under the Gmubi promoter, RbcsT terminator, and E1-NLS not edit sGFP in hairy roots? The NLS swap experiment suggests that the plant-derived elements can be successful, but they simply do not work together. However, this is contradicted by the low-level activity seen in *S. aureus* and *S. thermophilus* (figures 4.5, 4.6, 4.7), as well as the high level of activity observed in somatic embryos (figure 4.9). It could be explained that the Gmubi and E1-NLS combination slows down the editing process so that it cannot be observed in the 6-week timeframe of the hairy root experiment. This hypothesis could be legitimate because the editing efficiency observed over time in xCPR5 #3 did not achieve complete bi-allelic editing until somewhere between 10 and 20 weeks, far outside of the hairy root assay's timeframe (figure 4.10).

Another potential explanation for the discrepancy in editing between the root assay and embryo assay has to do with the soybean lines edited. Both edited lines have the Jack background, but the hairy root line contains the 2009 insertion of sGFP under the GmUbi-3 promoter (Hernandez-Garcia et al.). It is possible that sGFP's GmUbi-3 promoter interacts with the RGEN's GmUbi-3 promoter and contributes to silencing of the CRISPR system. This hypothesis could be tested by either by changing RGEN promoters, or by performing the assay on a soybean line that expresses GFP with some other plant-derived promoter.

The activity with *S. aureus* and *S. thermophilus* constructs was high enough to confirm editing with TIDE analysis (figures 4.5 4.6, 4.7) and was sufficiently high to

create bi-allelic mutations before maturity in regenerated somatic embryos (figure 4.10). A system with low but sufficient CRISPR activity could be viewed as ideal, as dosage of Cas9 has been shown to correlate with off-target modifications (Hsu et al., 2013). A CRISPR system which delivers only enough nuclease to create on-target mutation may be at a lower risk of developing off-target mutation as well.

The editing event in the plant xCPR5 #1-1 showed the exact same two indels at both 20 weeks and 30 weeks. Samples were taken from the first unopened trifoliolate leaf, which was more than five nodes separated in that 10 weeks of time. Because individual plants have different editing patterns, we can conclude that plants one through four represent four distinct editing events. Because they display the same pattern of indels across time and nodal distance, we can conclude that the edits are germinal, and therefore should be transmitted to progeny.

The five plants produced from xCPR5 event #3 represent four different editing patterns, because #3-4 and #3-5 have the same mutant alleles (table 4.1). The three events with double frame-shift mutations all mimic the fast-neutron mutant's short trichomes (figure 4.11). Two plants that each have one non-frameshifted allele display what appears to be an intermediate phenotype. This suggests that *Cpr5'*s effect on trichome length is additive. It also suggests that the trichome mutant is controlled by the intended mutation, rather than by any mutation at the CRISPR construct's integration site. This can be verified in subsequent generations as CRISPR machinery and edited loci segregate.

Conclusion

In conclusion, *S. pyogenes, S. aureus,* and *S. thermophilus* Cas9 proteins are all capable of editing soybean cells within plant-derived transformation vectors. The toolkit

developed here has successfully been utilized to confirm a gene function predicted from a mutagenesis assay.

Table 4.1: Nucleotide and predicted translation of *Cpr5* CRISPR targets in five regenerated plants at 30 weeks post transformation.

Highlighted sequences have no framing errors

WT Nucleotides	AAACGACGGAACCCTAGAGTTCTTGTTCGCCGCCATAGGGCTAATAATGTG
WT Residues	K R R N P R V L V R R H R A N N V
xCPR5 #3-1	AAACGACGGAACCCTAG-------GTTCGCCGCCATAGGGCTAATAATGTG
	K R R N P R F A A I G L I M
	AAACGACGGAACCCTAG--------TTCGCCGCCATAGGGCTAATAATGTG
	K R R N P S S P P * G * * C
xCPR5 #3-2	AAACGACGGAACCCTAG--TTCTTGTTCGCCGCCATAGGGCTAATAATGTG
	K R R N P S S C S P P * G * * C
	AAACGACGGAACCCTAG-----------CCGCCATAGGGCTAATAATGTG
	K R R N P S R H R A N N V
xCPR5 #3-3	AAACGACGGAACCCTAG--TTCTTGTTCGCCGCCATAGGGCTAATAATGTG
	K R R N P S S C S P P * G * * C
	AAACGACGGAACCCTAG--------------------GGCTAATAATGTG
	K R R N P R A N N V
xCPR5 #3-4	AAACGACGGAACCCTAG--TTCTTGTTCGCCGCCATAGGGCTAATAATGTG
	K R R N P S S C S P P * G * * C
	AAACGACGGAAC---------------GCCGCCATAGGGCTAATAATGTG
	K R R N A A I G L I M
xCPR5 #3-5	AAACGACGGAACCCTAG--TTCTTGTTCGCCGCCATAGGGCTAATAATGTG
	K R R N P S S C S P P * G * * C
	AAACGACGGAAC---------------GCCGCCATAGGGCTAATAATGTG
	K R R N A A I G L I M

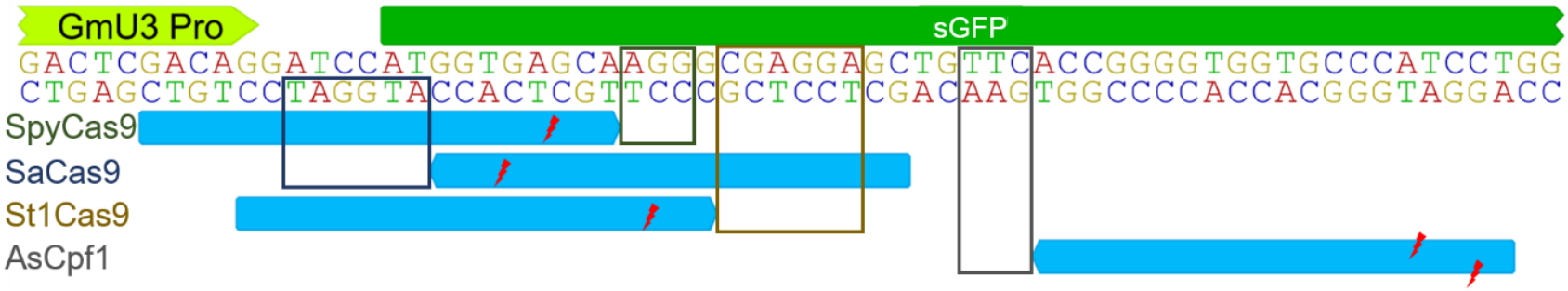

Figure 4.1: First 5' GFP targets selected with 4 RGENS. Squares highlight PAM regions, and red markers denote predicted cut-site.

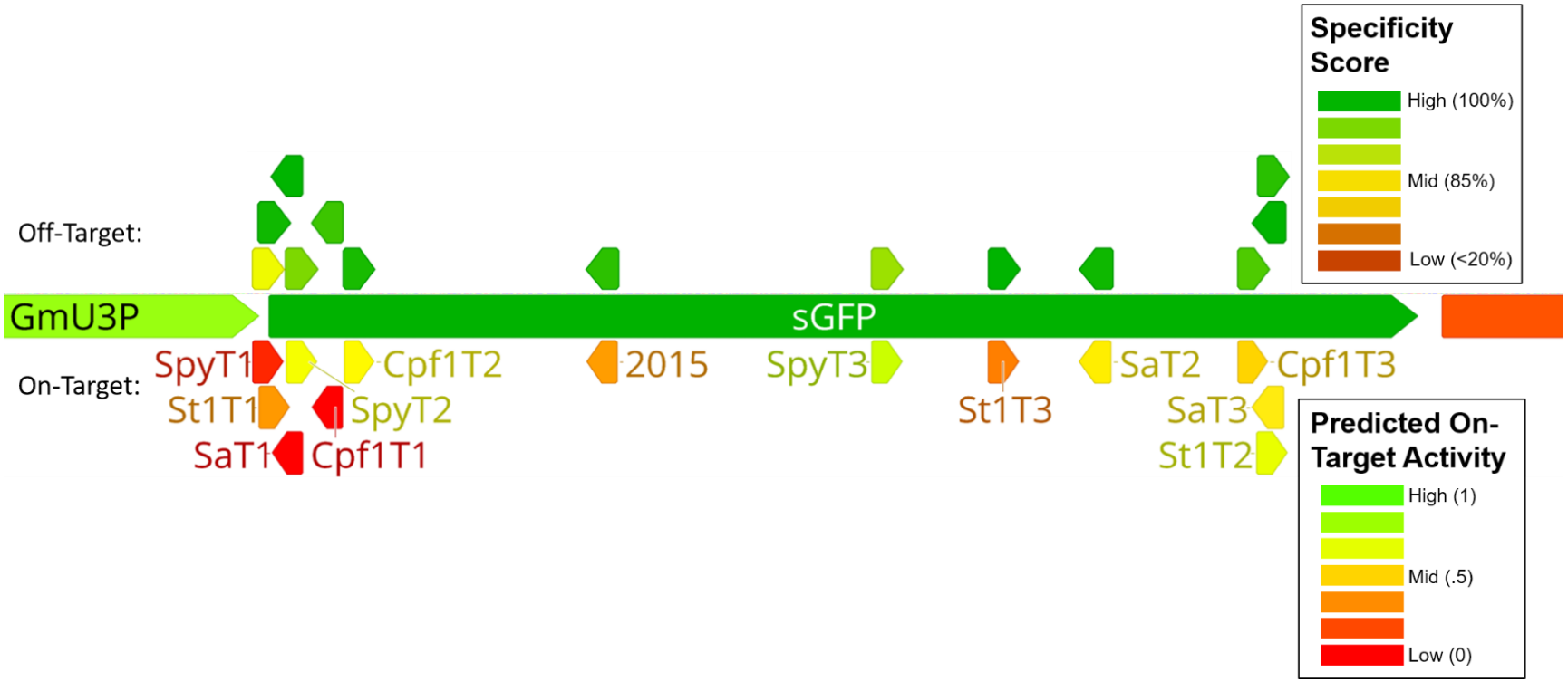

Figure 4.2: Targets and scores for RGEN comparison. Targets scored for predicted on-target activity with sgRNA scorer 2.0 (Chari et al., 2017) and specificity was scored with (Hsu et al., 2013)

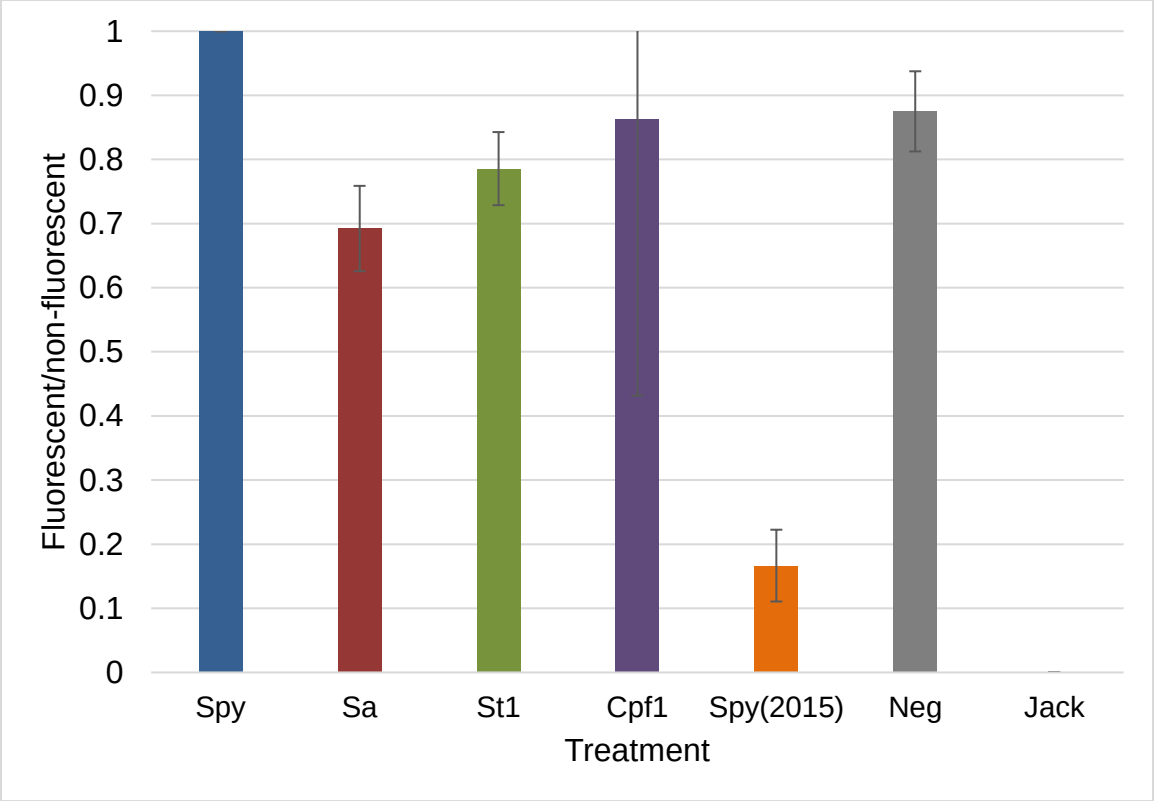

Figure 4.3: Average ratio of fluorescent signal with individual standard error. Initial trial with T1 targets. Scores of 1 represent GFP signal, and 0 represents lack of signal

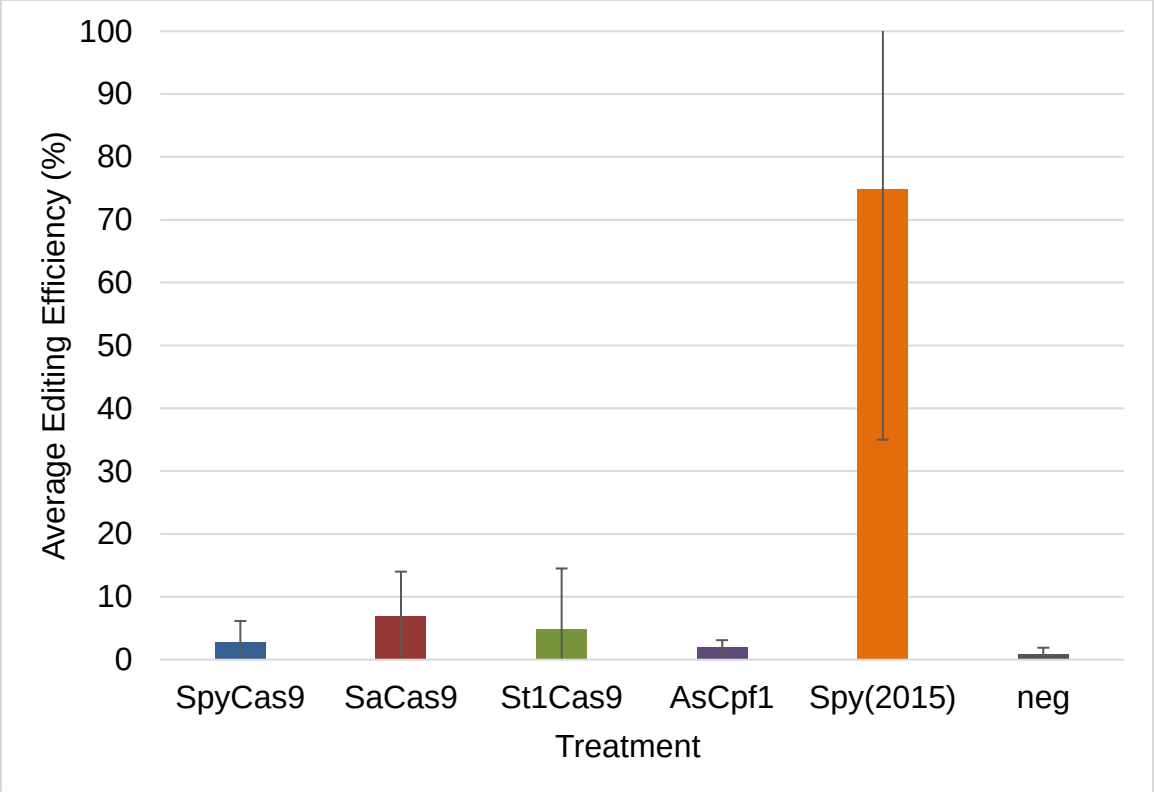

Figure 4.4: Average editing efficiencies and individual standard error for all gRNAs (T1, T2, T3) for each RGEN at 6 weeks

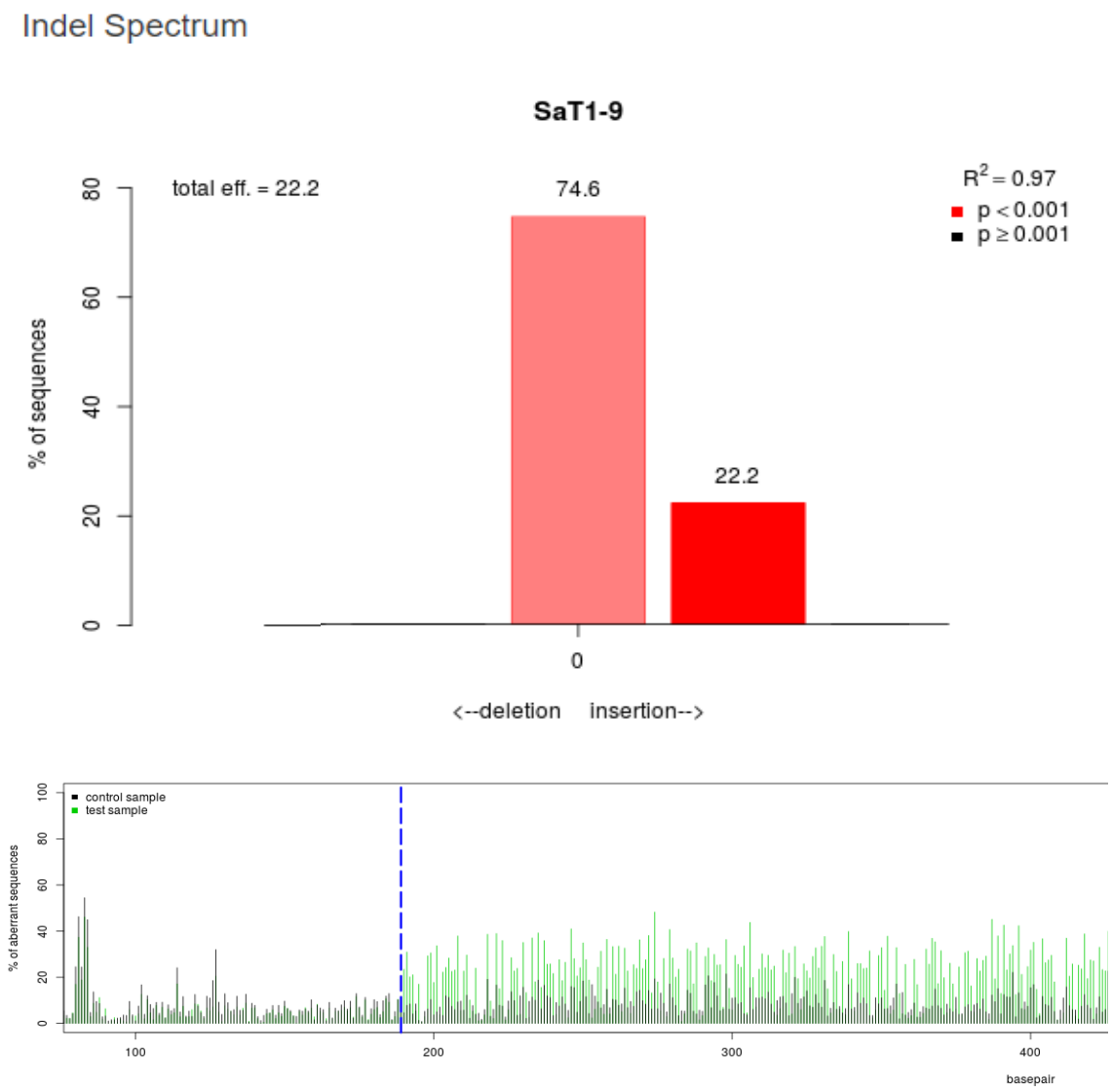

Figure 4.5: Verified editing with *S. aureus* demonstrated in soybean hairy root #9 as displayed with TIDE analysis. Bar graph displays the percentage of sequences that constitute each indel. Histogram compares sequence degeneracy between an unedited control and putative edit.

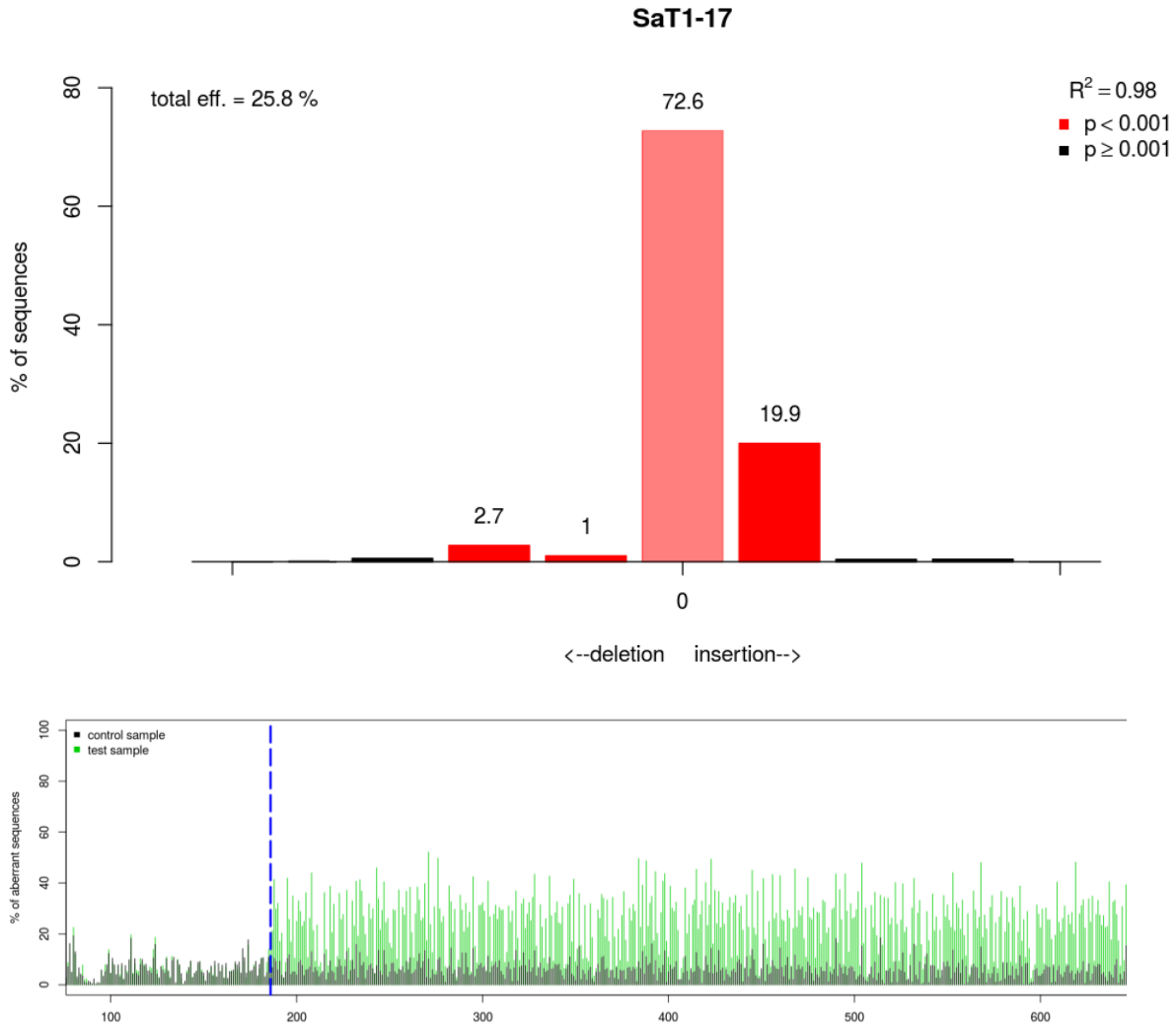

Figure 4.6: Verified editing with *S. aureus* demonstrated in soybean hairy root #17 as displayed with TIDE analysis. Bar graph displays the percentage of sequences that constitute each indel. Histogram compares sequence degeneracy between an unedited control and putative edit.

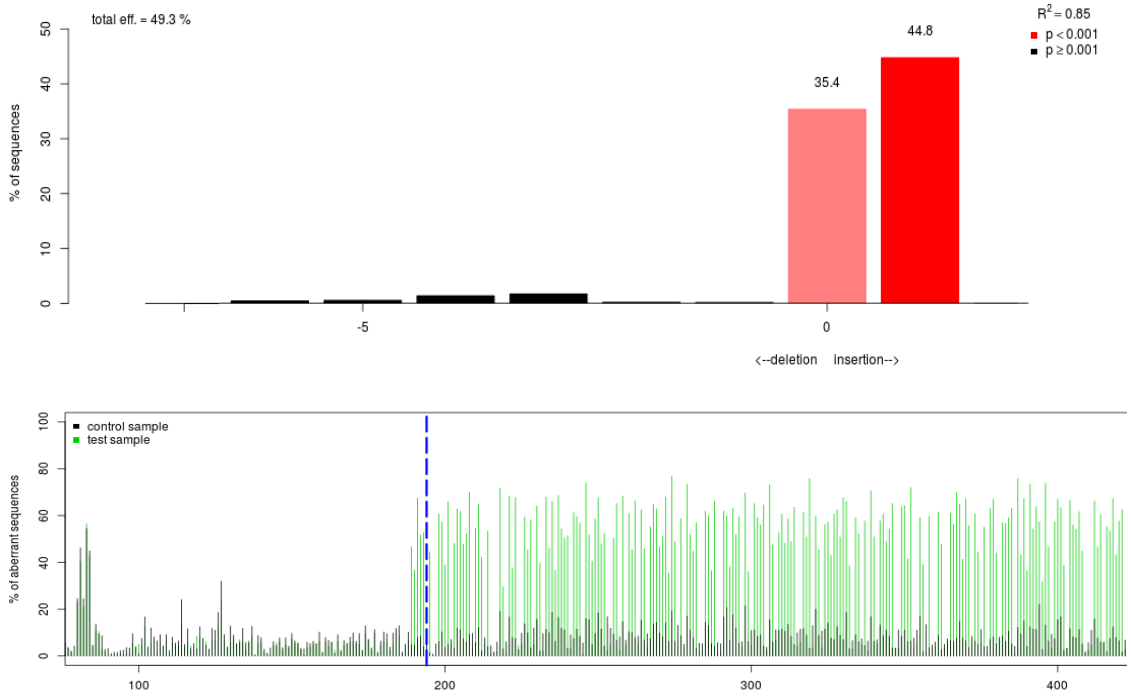

Figure 4.7: Verified editing with *S. thermophilus* demonstrated in soybean hairy root #16 as displayed with TIDE analysis. Bar graph displays the percentage of sequences that constitute each indel. Histogram compares sequence degeneracy between an unedited control and putative edit.

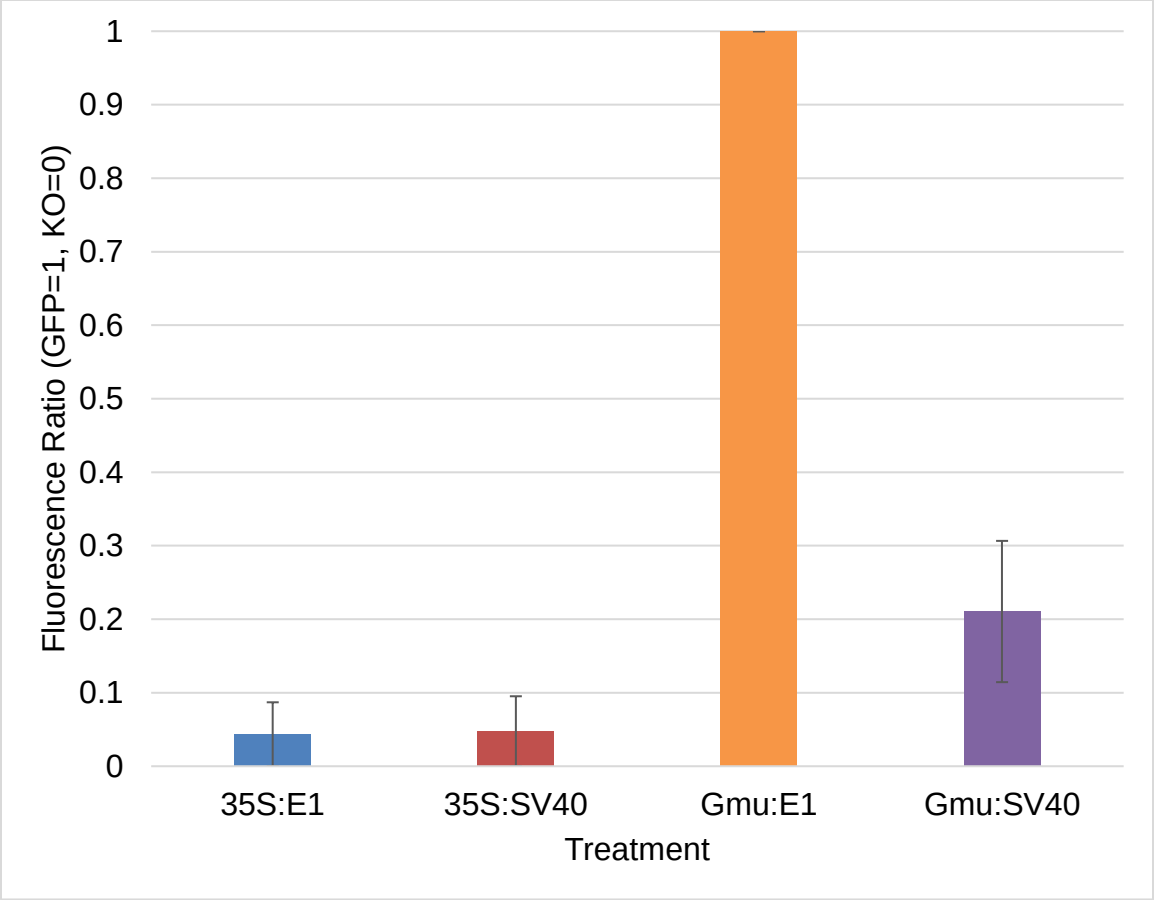

Figure 4.8: Interval Plot of GFP vs. Construct at 6 weeks. GFP score of 0 represents a loss-of-signal, and a score of 1 is producing a signal

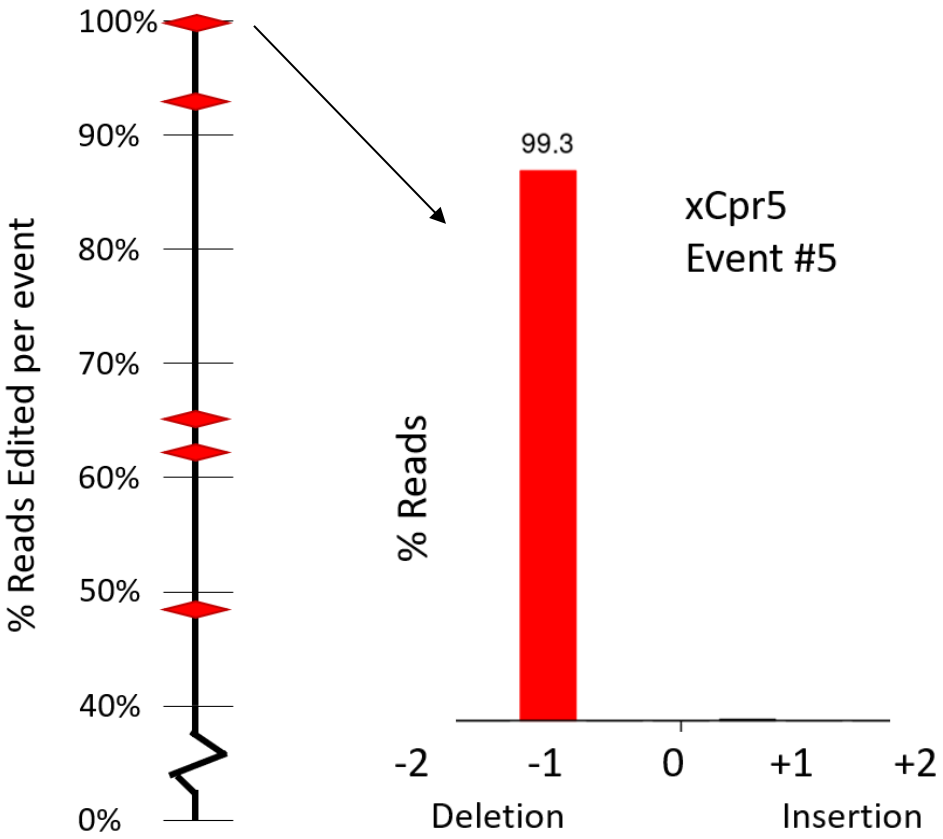

Figure 4.9: Total editing in 5 xCpr5 events at 10 weeks. Red diamonds represent individual transgenic events.

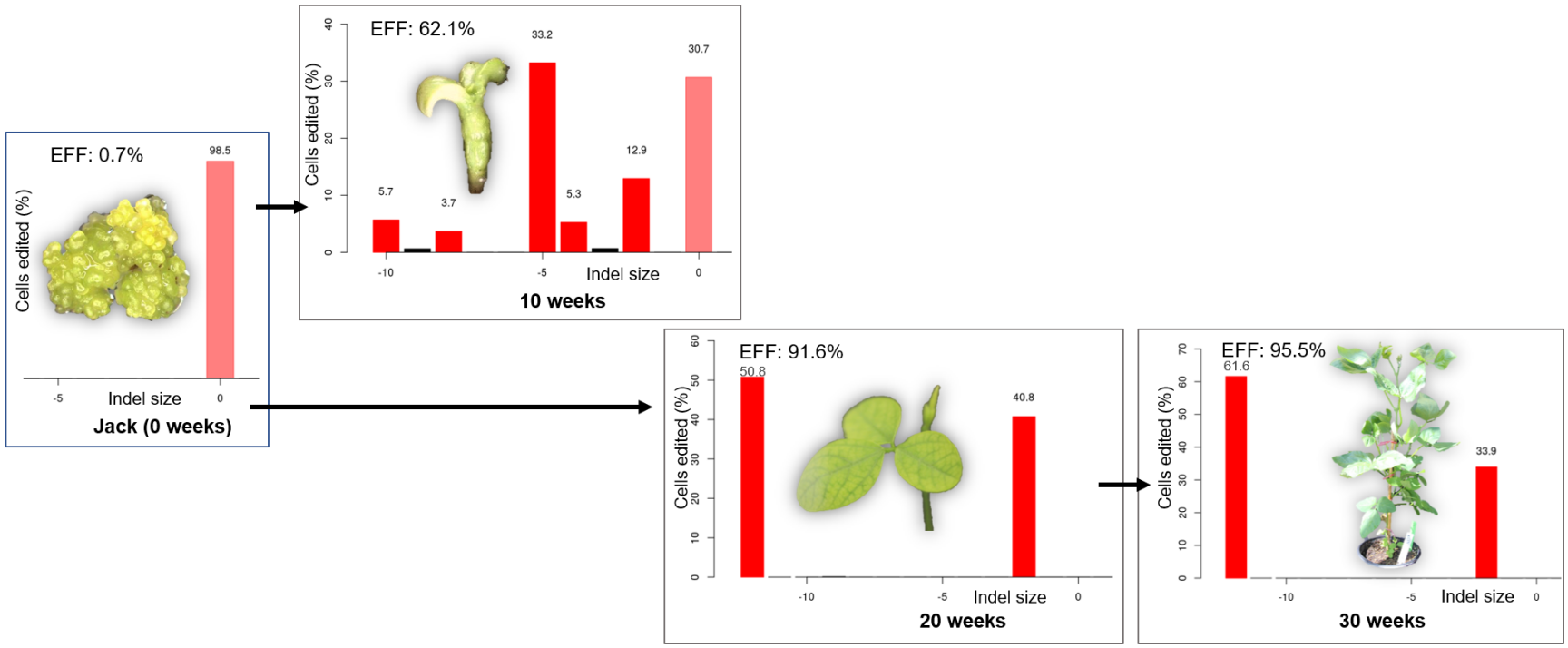

Figure 4.10: Editing over time in a developing plant xCpr5 #3-1. TIDE output shows the composition of indels detected in the amplicons produced at the target site in *Cpr5*. Photographs indicate developmental stage and phenotype of the event.

Figure 4.11: xCPR5 event #3 plant 1 photographed at 30 weeks.

Figure 4.12: Leaf and stem trichome detail on plants derived from xCpr5 event #3 and a control cv "Jack."

REFERENCES

340, C. 2004. Code of federal regulations.

Bernard, R. 1971. Two major genes for time of flowering and maturity in soybeans. Crop Science 11: 242-244.

Bolon, Y.T., W.J. Haun, W.W. Xu, D. Grant, M.G. Stacey, R.T. Nelson, et al. 2011. Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant Physiology 156: 240-253. doi:10.1104/pp.110.170811.

Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S.D. Ehrlich, S. Kulakauskas, et al. 2004. Complete sequence and comparative genome analysis of the dairy bacterium streptococcus thermophilus. Nat Biotechnol 22: 1554-1558. doi:10.1038/nbt1034.

Bowling, S.A., J.D. Clarke, Y.D. Liu, D.F. Klessig and X.N. Dong. 1997. The cpr5 mutant of arabidopsis expresses both npr1-dependent and npr1-independent resistance. Plant Cell 9: 1573-1584. doi:10.1105/tpc.9.9.1573.

Brinkman, E.K., T. Chen, M. Amendola and B. van Steensel. 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Research 42: e168. doi:10.1093/nar/gku936.

Burch-Smith, T.M., J.C. Anderson, G.B. Martin and S.P. Dinesh-Kumar. 2004. Applications and advantages of virus-induced gene silencing for gene function studies in plants. The Plant Journal 39: 734-746. doi:10.1111/j.1365-313X.2004.02158.x.

Chari, R., N.C. Yeo, A. Chavez and G.M. Church. 2017. Sgrna scorer 2.0: A species-independent model to predict crispr/cas9 activity. ACS Synth. Biol. 6: 902-904. doi:10.1021/acssynbio.6b00343.

Chatterjee, B. and C. Chakraborti. 1995. Non-sporing anaerobes in certain surgical group of patients. Journal of the Indian Medical Association 93: 333-335, 339.

Cho, H.J., S.K. Farrand, G.R. Noel and J.M. Widholm. 2000. High-efficiency induction of soybean hairy roots and propagation of the soybean cyst nematode. Planta 210: 195-204. doi:10.1007/pl00008126.

Cooper, J.L., B.J. Till, R.G. Laport, M.C. Darlow, J.M. Kleffner, A. Jamai, et al. 2008. Tilling to detect induced mutations in soybean. Bmc Plant Biology 8: 10. doi:10.1186/1471-2229-8-9.

Coruzzi, G., R. Broglie, C. Edwards and N.H. Chua. 1984. Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. The EMBO Journal 3: 1671-1679.

Covert, S.F., P. Kapoor, M.H. Lee, A. Briley and C.J. Nairn. 2001. Agrobacterium tumefaciens-mediated transformation of fusarium circinatum. Mycol Res 105. doi:10.1017/s0953756201003872.

Cradick, T.J., E.J. Fine, C.J. Antico and G. Bao. 2013. Crispr/cas9 systems targeting beta-globin and ccr5 genes have substantial off-target activity. Nucleic Acids Research 41: 9584-9592. doi:10.1093/nar/gkt714.

Cui, Y.Y., S. Barampuram, M.G. Stacey, C.N. Hancock, S. Findley, M. Mathieu, et al. 2013. Tnt1 retrotransposon mutagenesis: A tool for soybean functional genomics. Plant Physiology 161: 36-47. doi:10.1104/pp.112.205369.

Curtin, S.J., F. Zhang, J.D. Sander, W.J. Haun, C. Starker and N.J. Baltes. 2011. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiol 156. doi:10.1104/pp.111.172981.

Deltcheva, E., K. Chylinski, C.M. Sharma, K. Gonzales, Y.J. Chao, Z.A. Pirzada, et al. 2011. Crispr rna maturation by trans-encoded small rna and host factor rnase iii. Nature 471: 602-+. doi:10.1038/nature09886.

DiCarlo, J.E., J.E. Norville, P. Mali, X. Rios, J. Aach and G.M. Church. 2013. Genome engineering in saccharomyces cerevisiae using crispr-cas systems. Nucleic Acids Res 41. doi:10.1093/nar/gkt135.

Doench, J.G., N. Fusi, M. Sullender, M. Hegde, E.W. Vaimberg, K.F. Donovan, et al. 2016. Optimized sgrna design to maximize activity and minimize off-target effects of crispr-cas9. Nature Biotechnology 34: 184-+. doi:10.1038/nbt.3437.

Doench, J.G., E. Hartenian, D.B. Graham, Z. Tothova, M. Hegde, I. Smith, et al. 2014. Rational design of highly active sgrnas for crispr-cas9-mediated gene inactivation. Nature Biotechnology 32: 1262-U1130. doi:10.1038/nbt.3026.

Esvelt, K.M., P. Mali, J.L. Braff, M. Moosburner, S.J. Yaung and G.M. Church. 2013. Orthogonal cas9 proteins for rna-guided gene regulation and editing. Nature methods 10: 10.1038/nmeth.2681. doi:10.1038/nmeth.2681.

Feng, Z.Y., B.T. Zhang, W.N. Ding, X.D. Liu, D.L. Yang, P.L. Wei, et al. 2013. Efficient genome editing in plants using a crispr/cas system. Cell Res. 23: 1229-1232. doi:10.1038/cr.2013.114.

Finnegan, J. and D. Sherratt. 1982. Plasmid cole1 conjugal mobility - the nature of bom, a region required in cis for transfer. Mol. Gen. Genet. 185: 344-351. doi:10.1007/bf00330810.

Friedland, A.E., R. Baral, P. Singhal, K. Loveluck, S. Shen, M. Sanchez, et al. 2015. Characterization of staphylococcus aureus cas9: A smaller cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. Genome Biology 16: 10. doi:10.1186/s13059-015-0817-8.

Garbarino, J.E. and W.R. Belknap. 1994. Isolation of a ubiquitin-ribosomal protein gene (ubi3) from potato and expression of its promoter in transgenic plants. Plant Mol Biol 24. doi:10.1007/bf00040579.

Garneau, J.E., M.E. Dupuis, M. Villion, D.A. Romero, R. Barrangou, P. Boyaval, et al. 2010. The crispr/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468: 67-+. doi:10.1038/nature09523.

Gilbertson, R.L., M. Sudarshana, H. Jiang, M.R. Rojas and W.J. Lucas. 2003. Limitations on geminivirus genome size imposed by plasmodesmata and virus-encoded movement protein: Insights into DNA trafficking. The Plant Cell 15: 2578-2591. doi:10.1105/tpc.015057.

Girard, P.M., B. Kysela, C.J. Harer, A.J. Doherty and P.A. Jeggo. 2004. Analysis of DNA ligase iv mutations found in lig4 syndrome patients: The impact of two linked polymorphisms. Hum. Mol. Genet. 13: 2369-2376. doi:10.1093/hmg/ddh274.

Hancock, C.N., F. Zhang, K. Floyd, A.O. Richardson, P. LaFayette, D. Tucker, et al. 2011. The rice miniature inverted repeat transposable element mping is an

effective insertional mutagen in soybean. Plant Physiology 157: 552-562. doi:10.1104/pp.111.181206.

Haun, W., A. Coffman, B.M. Clasen, Z.L. Demorest, A. Lowy and E. Ray. 2014. Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnol J 12. doi:10.1111/pbi.12201.

Hernandez-Garcia, C.M., R.A. Bouchard, P.J. Rushton, M.L. Jones, X. Chen, M.P. Timko, et al. 2010. High level transgenic expression of soybean (glycine max) gmerf and gmubi gene promoters isolated by a novel promoter analysis pipeline. BMC Plant Biology 10: 1-16. doi:10.1186/1471-2229-10-237.

Hernandez-Garcia, C.M., A.P. Martinelli, R.A. Bouchard and J.J. Finer. 2009. A soybean (glycine max) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. Plant Cell Reports 28: 837-849. doi:10.1007/s00299-009-0681-7.

Horvath, P. and R. Barrangou. 2010. Crispr/cas, the immune system of bacteria and archaea. Science 327: 167-170. doi:10.1126/science.1179555.

Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, et al. 2013. DNA targeting specificity of rna-guided cas9 nucleases. Nature Biotechnology 31: 827-+. doi:10.1038/nbt.2647.

Hu, X., C. Wang, Q. Liu, Y. Fu and K. Wang. 2016. Targeted mutagenesis in rice using crispr-cpf1 system. Journal of Genetics and Genomics. doi:10.1016/j.jgg.2016.12.001.

Jacobs, T.B., P.R. LaFayette, R.J. Schmitz and W.A. Parrott. 2015. Targeted genome modifications in soybean with crispr/cas9. BMC Biotechnology 15: 1-10. doi:10.1186/s12896-015-0131-2.

Jacobs, T.B., N.J. Lawler, P.R. LaFayette, L.O. Vodkin and W.A. Parrott. 2016. Simple gene silencing using the trans-acting sirna pathway. Plant Biotechnol. J. 14: 117-127. doi:10.1111/pbi.12362.

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

Kalderon, D., B.L. Roberts, W.D. Richardson and A.E. Smith. 1984. A short amino-acid sequence able to specify nuclear location. Cell 39: 499-509. doi:10.1016/0092-8674(84)90457-4.

Kaya, H., M. Mikami, A. Endo, M. Endo and S. Toki. 2016. Highly specific targeted mutagenesis in plants using staphylococcus aureus cas9. Scientific Reports 6: 9. doi:10.1038/srep26871.

Kereszt, A., D.X. Li, A. Indrasumunar, C.D.T. Nguyen, S. Nontachaiyapoom, M. Kinkema, et al. 2007. Agrobacterium rhizogenes - mediated transformation of soybean to study root biology. Nat. Protoc. 2: 948-952. doi:10.1038/nprot.2007.141.

Kim, G.B. and Y.W. Nam. 2013. Isolation and characterization of medicago truncatula u6 promoters for the construction of small hairpin rna-mediated gene silencing vectors. Plant Mol. Biol. Rep. 31: 581-593. doi:10.1007/s11105-012-0528-1.

Kosugi, S., M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita, et al. 2009. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. Journal of Biological Chemistry 284: 478-485. doi:10.1074/jbc.M807017200.

Leenay, Ryan T., Kenneth R. Maksimchuk, Rebecca A. Slotkowski, Roma N. Agrawal, Ahmed A. Gomaa, Alexandra E. Briner, et al. 2016. Identifying and visualizing functional pam diversity across crispr-cas systems. Molecular Cell. doi:10.1016/j.molcel.2016.02.031.

Li, Z., Z.-B. Liu, A. Xing, B.P. Moon, J.P. Koellhoffer, L. Huang, et al. 2015. Cas9-guide rna directed genome editing in soybean. Plant Physiology. doi:10.1104/pp.15.00783.

Li, Z.J., S. Jayasankar and D.J. Gray. 2001. Expression of a bifunctional green fluorescent protein (gfp) fusion marker under the control of three constitutive promoters and enhanced derivatives in transgenic grape (vitis vinifera). Plant Science 160: 877-887. doi:10.1016/s0168-9452(01)00336-3.

Liu, Y.L., M. Schiff and S.P. Dinesh-Kumar. 2002. Virus-induced gene silencing in tomato. Plant J. 31: 777-786. doi:10.1046/j.1365-313X.2002.01394.x.

Makarova, K.S., Y.I. Wolf, O.S. Alkhnbashi, F. Costa, S.A. Shah, S.J. Saunders, et al. 2015. An updated evolutionary classification of crispr-cas systems. Nat Rev Microbiol 13: 722-736. doi:10.1038/nrmicro3569.

Mali, P., L.H. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, et al. 2013. Rna-guided human genome engineering via cas9. Science 339: 823-826. doi:10.1126/science.1232033.

Mankin, S., D. Hill, O. Paula, T. Effie, W. Allan, N. Lawrence, et al. 2007. Disarming and sequencing of agrobacterium rhizogenes strain k599 (ncppb2659) plasmid pri2659. In Vitro Cellular & Developmental Biology Plant 43: 521-535.

Meng, Z., C. Ruberti, Z.Z. Gong and F. Brandizzi. 2017. Cpr5 modulates salicylic acid and the unfolded protein response to manage tradeoffs between plant growth and stress responses. Plant J. 89: 486-501. doi:10.1111/tpj.13397.

Mitsuhara, I., M. Ugaki, H. Hirochika, M. Ohshima, T. Murakami, Y. Gotoh, et al. 1996. Efficient promoter cassettes for enhanced expression of foreign genes in dicotyledonous and monocotyledonous plants. Plant Cell Physiol. 37: 49-59.

Moreno-Mateos, M.A., J.P. Fernandez, R. Rouet, M.A. Lane, C.E. Vejnar, E. Mis, et al. 2017. Crispr-cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. bioRxiv.

Murray, M.G. and W.F. Thompson. 1980. Rapid isolation of high molecular-weight plant DNA. Nucleic Acids Res 8. doi:10.1093/nar/8.19.4321.

Nishimasu, H., F.A. Ran, P.D. Hsu, S. Konermann, S.I. Shehata, N. Dohmae, et al. 2014. Crystal structure of cas9 in complex with guide rna and target DNA. Cell 156: 935-949. doi:10.1016/j.cell.2014.02.001.

Olhoft, P.M., L.E. Flagel, C.M. Donovan and D.A. Somers. 2003. Efficient soybean transformation using hygromycin b selection in the cotyledonary-node method. Planta 216.

Orf, J.H. and R.L. Denny. 2004. Registration of 'mn1302' soybean. Crop Science 44: 693-693.

Parrott, W.A., G. Dryden, S. Vogt, D.F. Hildebrand, G.B. Collins and E.G. Williams. 1988. Optimization of somatic embryogenesis and embryo germination in soybean. In Vitro Cellular & Developmental Biology 24: 817-820.

Pedelacq, J.D., S. Cabantous, T. Tran, T.C. Terwilliger and G.S. Waldo. 2006. Engineering and characterization of a superfolder green fluorescent protein. Nature Biotechnology 24: 79-88. doi:10.1038/nbt1172.

Pflieger, S., S. Blanchet, C. Meziadi, M.M.S. Richard, V. Thareau, F. Mary, et al. 2014. The "one-step" bean pod mottle virus (bpmv)derived vector is a functional genomics tool for efficient overexpression of heterologous protein, virus-induced gene silencing and genetic mapping of bpmv r-gene in common bean (phaseolus vulgaris l.). Bmc Plant Biology 14: 16. doi:10.1186/s12870-014-0232-4.

Puchta, H. 2005. The repair of double-strand breaks in plants: Mechanisms and consequences for genome evolution. J. Exp. Bot. 56: 1-14. doi:10.1093/jxb/eri025.

Ran, F.A., L. Cong, W.X. Yan, D.A. Scott, J.S. Gootenberg, A.J. Kriz, et al. 2015. In vivo genome editing using staphylococcus aureus cas9. Nature 520: 186-191. doi:10.1038/nature14299.

Ran, F.A., P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott and F. Zhang. 2013. Genome engineering using the crispr-cas9 system. Nat. Protocols 8: 2281-2308. doi:10.1038/nprot.2013.143

Ratcliff, F., A.M. Martin-Hernandez and D.C. Baulcombe. 2001. Tobacco rattle virus as a vector for analysis of gene function by silencing. Plant J. 25: 237-245. doi:10.1046/j.0960-7412.2000.00942.x.

Ruiz, M.T., O. Voinnet and D.C. Baulcombe. 1998. Initiation and maintenance of virus-induced gene silencing. Plant Cell 10: 937-946.

Samoylov, V.M., D.M. Tucker and W.A. Parrott. 1998. Soybean glycine max (l). Merrill embryogenic cultures: The role of sucrose and total nitrogen content on proliferation. In Vitro Cell. Dev. Biol.-Plant 34: 8-13.

Schardl, C.L., A.D. Byrd, G. Benzion, M.A. Altschuler, D.F. Hildebrand and A.G. Hunt. 1987. Design and construction of a versatile system for the expression of foreign genes in plants. Gene 61: 1-11. doi:10.1016/0378-1119(87)90359-3.

Schmidt, M.A., D.M. Tucker, E.B. Cahoon and W.A. Parrott. 2005. Towards normalization of soybean somatic embryo maturation. Plant Cell Rep 24: 383-391. doi:10.1007/s00299-005-0950-z.

Schwarzacher, T., M.L. Wang, A.R. Leitch, N. Miller, G. Moore and J.S. HeslopHarrison. 1997. Flow cytometric analysis of the chromosomes and stability of a wheat cell-culture line. Theor. Appl. Genet. 94: 91-97.

Semenova, E., M.M. Jore, K.A. Datsenko, A. Semenova, E.R. Westra, B. Wanner, et al. 2011. Interference by clustered regularly interspaced short palindromic repeat (crispr) rna is governed by a seed sequence. Proceedings of the National Academy of Sciences of the United States of America 108: 10098-10103. doi:10.1073/pnas.1104144108.

Shah, S.A., S. Erdmann, F.J.M. Mojica and R.A. Garrett. 2013. Protospacer recognition motifs: Mixed identities and functional diversity. RNA Biology 10: 891-899. doi:10.4161/rna.23764.

Shan, Q.W., Y.P. Wang, J. Li, Y. Zhang, K.L. Chen, Z. Liang, et al. 2013. Targeted genome modification of crop plants using a crispr-cas system. Nature Biotechnology 31: 686-688. doi:10.1038/nbt.2650.

Soltis, D.E., V.A. Albert, J. Leebens-Mack, C.D. Bell, A.H. Paterson, C.F. Zheng, et al. 2009. Polyploidy and angiosperm diversification. Am. J. Bot. 96: 336-348. doi:10.3732/ajb.0800079.

Soriano, F., C.E. Shelburne and M. Gokcen. 1974. Simian virus 40 in a human cancer. Nature 249: 421-424. doi:10.1038/249421a0.

Steinert, J., S. Schiml, F. Fauser and H. Puchta. 2015. Highly efficient heritable plant genome engineering using cas9 orthologues from streptococcus thermophilus and staphylococcus aureus. Plant J. 84: 1295-1305. doi:10.1111/tpj.13078.

Sternberg, S.H., S. Redding, M. Jinek, E.C. Greene and J.A. Doudna. 2014. DNA interrogation by the crispr rna-guided endonuclease cas9. Nature 507: 62-+. doi:10.1038/nature13011.

Tang, X., L.G. Lowder, T. Zhang, A.A. Malzahn, X.L. Zheng, D.F. Voytas, et al. 2017. A crispr-cpf1 system for efficient genome editing and transcriptional repression in plants (vol 3, 17018, 2017). Nat. Plants 3: 1. doi:10.1038/nplants.2017.103.

Thompson, J. and W. Parrott. 1998. Pmeca: A size-based, blue/white selection multiple common and rare-cutter general cloning and transcription vector. BioTechniques 24: 922-927.

Thomson, J.M., P.R. Lafayette, M.A. Schmidt and W.A. Parrott. 2002. Artificial gene-clusters engineered into plants using a vector system based on intron- and intein-

encoded endonucleases. In Vitro Cell. Dev. Biol.-Plant 38: 537-542. doi:10.1079/ivp2002329.

Trick, H.N., R.D. Dinkins, E.R. Santarém, R. Samoyolov, C. Meurer, D. Walker, et al. 1997. Recent advances in soybean transformation. Plant Tissue Culture and Biotechnology 3: 9-26.

Tricoli, D.M., K.J. Carney, P.F. Russell, J.R. McMaster, D.W. Groff, K.C. Hadden, et al. 1995. Field-evaluation of transgenic squash containing single or multiple virus coat protein gene constructs for resistance to cucumber mosaic-virus. Bio-Technology 13: 1458-1465. doi:10.1038/nbt1295-1458.

vanderBij, A.J., L.A. deWeger, W.T. Tucker and B.J.J. Lugtenberg. 1996. Plasmid stability in pseudomonas fluorescens in the rhizosphere. Appl. Environ. Microbiol. 62: 1076-1080.

Wang, F., L. Wang, L. Qiao, J. Chen, M.B. Pappa, H. Pei, et al. 2017. Arabidopsis cpr5 regulates ethylene signaling via molecular association with the etr1 receptor. J Integr Plant Biol. doi:10.1111/jipb.12570.

Wang, S., Y.N. Gu, S.G. Zebell, L.K. Anderson, W. Wang, R. Mohan, et al. 2014. A noncanonical role for the cki-rb-e2f cell-cycle signaling pathway in plant effector-triggered immunity. Cell Host Microbe 16: 787-794. doi:10.1016/j.chom.2014.10.005.

Wang, Y.P., X. Cheng, Q.W. Shan, Y. Zhang, J.X. Liu, C.X. Gao, et al. 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nature Biotechnology 32: 947-951. doi:10.1038/nbt.2969.

Watson, N. 1988. A new revision of the sequence of plasmid-pbr322. Gene 70: 399-403. doi:10.1016/0378-1119(88)90212-0.

Xia, Z.J., S. Watanabe, T. Yamada, Y. Tsubokura, H. Nakashima, H. Zhai, et al. 2012. Positional cloning and characterization reveal the molecular basis for soybean maturity locus e1 that regulates photoperiodic flowering. Proceedings of the National Academy of Sciences of the United States of America 109: E2155-E2164. doi:10.1073/pnas.1117982109.

Xie, K. and Y. Yang. 2013. Rna-guided genome editing in plants using a crispr-cas system. Mol Plant 6: 1975-1983. doi:10.1093/mp/sst119.

Zetsche, B., J.S. Gootenberg, O.O. Abudayyeh, I.M. Slaymaker, K.S. Makarova, P. Essletzbichler, et al. 2015. Cpf1 is a single rna-guided endonuclease of a class 2 crispr-cas system. Cell 163: 759-771. doi:10.1016/j.cell.2015.09.038.

Zetsche, B., M. Heidenreich, P. Mohanraju, I. Fedorova, J. Kneppers, E.M. DeGennaro, et al. 2017. Multiplex gene editing by crispr-cpf1 using a single crrna array. Nat Biotech 35: 31-34. doi:10.1038/nbt.3737.

Zhang, H., J. Zhang, P. Wei, B. Zhang, F. Gou and Z. Feng. 2014. The crispr/cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol J 12. doi:10.1111/pbi.12200.

Zhou, X., T.B. Jacobs, L.J. Xue, S.A. Harding and C.J. Tsai. 2015. Exploiting snps for biallelic crispr mutations in the outcrossing woody perennial populus reveals 4-coumarate: Coa ligase specificity and redundancy. New Phytologist 208: 298-301.

APPENDICES

Table 5.1: Primers

sGFP amplified adding E1NLS	
sGFP- F	ATTACCATGGTGAGCAAGGGCGA
sGFP- e1L-R	ATTCTGCAGTCAGAATCTTCTCCTTGATGTCCTAAAGTTAGAGGCTTCGCATATGGTGGATTTCCTCTTCTTCTTGTACAGCTCGTCCATGCCG
Gmubi amplified from pGmubi	
GmUbl3P-F +SpeI	TAACTAGTGCATGCGGGCCCAATATAA
GmUbl3P-R +NheI, AscI	ATTAGGCGCGCCATGCTAGCTGGATCCTGTCGAGTCAACAAT
RbcsT amplified from pGmute	
infRbcsT- R	CTGTTATCCCTAGATTAATTAAGGATTGATGCATGTTGTCA
infAscRbcsT- F	TAGCATGGCGCGCCTACCTAGGTTCGAGTATTATGGC
RGENs amplified for insertion into plant-derived cassettes, adding E1NLS	
pStuGmu SaurCas9 Fwd	CTCGACAGGATCCAGCTAGCATGAAGCGGAACTACATCCTG
pStuGmu SaurCas9 Rev	TGCCATAATACTCGAACCTAGGTCATCTTCTCCTTGATGTCCTAAAGTTAGAGGCTTCGCATATGGTGGATTTCCTCTTCTTGCCCTTTTTGATGATCTG
PStuGmu SthermCas9 Fwd	CTCGACAGGATCCAGCTAGCATGAGCGACCTGGTGCTG
pStuGmu SthermCas9 Rev	TGCCATAATACTCGAACCTAGGTCATCTTCTCCTTGATGTCCTAAAGTTAGAGGCTTCGCATATGGTGGATTTCCTCTTCTTGAAGTCCAACTTGGGCTTG
gibson AsCfp1 Fwrd	TGACTCGACAGGATCCAGCTAGCATGACACAGTTCGAGGGC
gibson AsCfp1 Rev	CAATGCCATAATACTCGAACCTAGGTCATCTTCTCCTTGATGTCCTAAAGTTAGAGGCTTCGCATATGGTGGATTTCCTCTTCTTGTTGCGCAGCTCCTGGATGT
GmuCas9NheI	CTCGACAGGATCCAGCTAGCATGGACAAGAAGTA
cas9frag2:e1InLs-R avrII	ATCCTAGGTCAGAATCTTCTCTTGATGTCCTAAAGTTAGAGGCTTCGCATATGGTGGATTTCCTCTTCTTGTCTCCACCGAGCTGAGAGAGG

Clean cassette with RGENs amplified for insertion into p201Npt	
p201Spe-Gmubi 091216	CTATGTGCTTTGGATCGATCTGCCCACTGTGGGCCCAATATAACAACGACGTCGT
p201I-SceI-RbcsT 091216	TAAAACGACGGCCATGCCAAGCTTAATTACCCTGTTATCCCTAATTGATGCATGTTGTCAATCAATTG
MtU6 amplification for insertion at SpeI	
Spe2689-MtU6F	TAACTATGTGCTTTGGATCTGCCCAATGCCTATCTTATATGATCAATGAGG
MtU6R	AAGCCTACTGGTTCGCTTGAAG
SpyScaffold amplification for insertion at SpeI	
SpyScaffoldF	GTTTTAGAGCTAGAAATAGCA
GASpeIScaffoldR	GTTGTTATATTGGGCCCGACTAGAAAAAAAGCACCGACTCGGTG
ssDNA oligos for gRNA scaffolds with overhang for Gmubi	
AsCpf1 Target-Scaffold Oligo	CTTCAAGCGAACCAGTAGGCTTGAATTTCTACTGTTGTAGATNNNNNNNNNNNNNNNNNNNNTTTTTTCTAGTCGGGCCCAATATAACAAC
SaScaffold	TTTCCAGAGTACTAAAACGGTGAGCAAGGGCGAGGAGCAAGCCTACTGGTTCGCTT
St1Scaffold pt.1	AATGACAGGGTGTTGATTTCGGCATGAAGCCTTATCTTTGTAGCTTCTGCATTTTTGAGAGTACAAAAAC
GASt1Scaffold pt. 2	GAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTCGTTATTTAACTAGTCGGGCCCAATATAACAAC
ssDNA oligos for Targets	
SpyT1	AAGCGAACCAGTAGGCTTGCCGTGAGTGATCCCGGCGGGTTTTAGAGCTAGAAATA
SpyT2	AAGCGAACCAGTAGGCTTGCTGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATA
SpyT3	AAGCGAACCAGTAGGCTTGAAGGGCATCGACTTCAAGGGTTTTAGAGCTAGAAATA
SaT1	TTTCCAGAGTACTAAAACGGTGAGCAAGGGCGAGGAGCAAGCCTACTGGTTCGCTT
SaT2	TTTCCAGAGTACTAAAACCACAACATCGAGGACGGCAGCAAGCCTACTGGTTCGCTT
SaT3	TTTCCAGAGTACTAAAACTCCGCCCTGAGCAAAGACCCCAAGCCTACTGGTTCGCTT
St1T1	AAGCGAACCAGTAGGCTTGGATCCATGGTGAGCAAGGGGTTTTTGTACTCTCAAAA
St1T2	AAGCGAACCAGTAGGCTTGCCCTGAGCAAAGACCCCAAGTTTTTGTACTCTCAAAA
St1T3	AAGCGAACCAGTAGGCTTGTCTATATCATGGCCGACAAGTTTTTGTACTCTCAAAA
AsCpf1 Target-Scaffold 1	CTTCAAGCGAACCAGTAGGCTTGAATTTCTACTGTTGTAGATACCGGGGTGGTGCCCATCCTTTTTTTCTAGTCGGGCCCAATATAACAAC

AsCpf1 Target-Scaffold 2	CTTCAAGCGAACCAGTAGGCTTGAATTTCTACTGTTGTAGATCGTCGCCGTCCAGCTCGACCTTTTTTCTAGTCGGGCCCAATATAACAAC
AsCpf1 Target-Scaffold 3	CTTCAAGCGAACCAGTAGGCTTGAATTTCTACTGTTGTAGATCTCAGGGCGGACTGGGTGCTTTTTTTCTAGTCGGGCCCAATATAACAAC
2015T2Target	AAGCGAACCAGTAGGCTTGCTGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATA
NLS Swap primers	
Cas9NLSSwitchForward	GTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTT
SV40Reverse-AvrIIOverhang	CCAATGCCATAATACTCGAACCTAGGTCACACCTTCCTCTTCTT
E1NLSReverse-SacIOverhang	CGCGTAAGCTTCTGCAGATATCCGCGGTCATCTTCTCCTTGATGTCCT