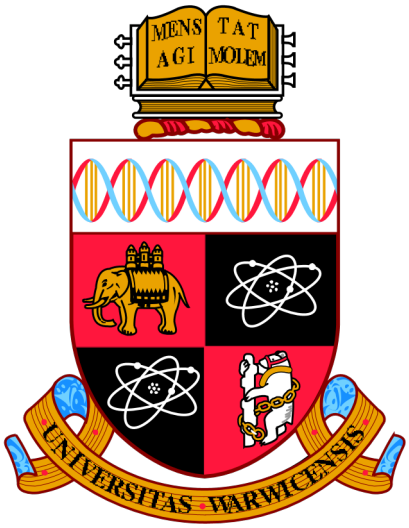


A non-integrative CRISPR/Cas9 genome-editing approach for use in vegetable crop breeding

by

Claudia Verónica Payacán Ortiz

A thesis submitted for the degree of

Doctor of Philosophy

University of Warwick

School of Life Sciences

July 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First, I would like to thank my supervisor, Dr. Stephen Jackson for his support during this PhD, his constant guidance and for giving me the opportunity to join his group. To former and current SJ group members: Thomas, Sally, Julien, Richa, Svenning, Josh, Huidong and Mohammad. All of you have contributed to my development as a scientist, helping me with your expertise and knowledge. To Dr. Yiguo Hong and my advisory panel Dr. Guy Barker, Dr. Brian Thomas and Dr. George Bassel for their expert advice and valuable feedback. To CONICYT for funding this PhD through the programme to pursue graduate studies abroad "Becas Chile" grant number 72180328.

To my beautiful C030 lab, especially Lorenzo, Javier, Yang-Seok, Ham, Natassa, Despoina, Alexia, Stefano, Ana, Daniela, and Bea. Thank you for sharing everyday with me with during a cup of tea, lunch breaks, talks, parties or simply enjoying the sunny days. All of you made my days happier and hopefully our paths will cross again in the future.

To my friends in Chile Sharim, Gabi, Leslie, Hector, Cris, Pancho, Javier, Carla, Yanara, Rocío and Pali for their remote support, for always being there, listening to me and believing in my capacities.

To my family and Mauricio, that were a massive support during this PhD. You were with me through thick and thin, always caring about my wellbeing and encouraging me to keep going, even if the road was rocky, because at the end, everything would be alright.

Finally, I would like to thank to my grandparents. Even though there are not physically with me anymore, they taught me to always follow my dreams, the importance of keep studying and being kind with others. No matter how difficult the path could seem, I know you are always by my side.

DECLARATIONS

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by myself.

ABSTRACT

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/CRISPR-associated nuclease 9 (Cas9) system has become the most widespread method to produce edited plants with improved traits. A DNA-free genome editing approach was developed using the plant virus vectors Potato virus X (PVX) and Tobacco rattle virus (TRV) to deliver and express Cas9 and sgRNAs. Splitting two Cas9 orthologs was explored to avoid viral instability due to the delivery of large transgenes. Manipulating flowering time in crops is a long-term goal in plant biotechnology. To achieve this, targeted genome edits were introduced in the SD flowering *Nicotiana tabacum* var. Maryland Mammoth *FLOWERING LOCUS T 4* (*NtFT4*) gene.

The expression of the split Cas9 was corroborated in *N. benthamiana* leaves and *in vivo* activity was validated in *N. tabacum* protoplasts by edition of the *PDS* gene. A single transcription unit was constructed encoding the *Cas9* and the sgRNA flanked by ribozymes. Mature sgRNAs were detected after self-cleavage of the ribozymes, enabling them to guide the Cas9 to the *NtFT4* target sites. Genome editing was found in *N. tabacum* leaves inoculated with TRV full-length or split Cas9, with frequencies between 3.03% to 0.29%, depending on the enzyme and target site. A delay in flowering time was observed in some plants, however other factors, such as strong viral symptoms, cannot be discarded. Cas9 protein was detected only in inoculated leaves with TRV-full-length *Sp*Cas9, and TRV or PVX vectors expressing the *Sa*Cas9 C-terminal end, while *Cas9* mRNA was found in all inoculated leaves. Even though systemic symptoms of TRV viral infection were seen, full-length or split *SpCas9* mRNA was not found, while weak expression of full-length and N-terminal *SaCas9* mRNA was observed. In contrast, *SaCas9* C-terminal mRNA was detected strongly in TRV systemic leaves, whilst weak expression was found in one PVX systemic leaf sample. Thus, a cargo below 1.3 kb is suggested for viral stability. To obtain fully edited plants, shoots from inoculated tobacco leaves with the different viral constructs were regenerated, but only non-edited plants were found.

In summary, this work demonstrates that viral-mediated genome editing is feasible. Improvements to this system, such as the use of newly developed Cas enzymes, are also discussed and may prove useful in generating crops with new desirable agronomical traits.

ABBREVIATIONS

(-) ssRNA	negative single-stranded RNA
(+) ssRNA	positive single-stranded RNA
µg	micrograms
µL	microliters
µM	micromolar
°C	degree Celsius
ABE	Adenine base editors
Asn	Asparagine
ATP	Adenosine triphosphate
*At*U6-26	*Arabidopsis thaliana* U6-26 promoter
BAP	6-benzylaminopurine
BNYVV	*Beet necrotic yellow vein virus*
bp	base pair
BSA	Bovine serum albumin
BSMV	*Barley stripe mosaic virus*
BYSMV	*Barley yellow striate mosaic virus*
$CaCl_2$	Calcium chloride
CaMV 35S	Cauliflower mosaic virus 35S promoter
CaMV	*Cauliflower mosaic virus*
CAPS	Cleaved Amplified Polymorphic Sequences assay
CBE	Cytosine base editors
cDNA	Complementary DNA
cm	Centimetre
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
cRT-PCR	Circular reverse transcription PCR
CTAB	Cetyltrimethylammonium bromide
CTP	Cytidine triphosphate
Cys	Cysteine
dATP	Deoxyadenosine triphosphate
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
dpi	Days post-infection
DSB	Double strand break
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FOMV	*Foxtail mosaic virus*
FT	*Flowering locus T* gene
g	grams
g	Relative centrifugal force

gDNA	Genomic DNA
GFP	Green fluorescent protein
Glu	Glutamic acid
GMO	Genetically modified organism
gRNA	guide RNA
GTP	Guanosine triphosphate
GUS	*β-glucuronidase* gene
H	hour
HDR	Homology directed repair
IAA	Isoamyl alcohol or 3-Methylbutan-1-ol
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KCl	Potassium chloride
kDa	kiloDalton
LB	media/LB agar Luria-Bertani media or agar
LD	Long day
LiCl	Lithium chloride
M	molar
M	molar
MES	2-(N-morpholino)ethanesulfonic acid
mg	milligrams
$MgCl_2$	Magnesium chloride
Min	minute
mM	millimolar
NAA	1-naphthaleneacetic acid
NaCl	Sodium chloride
NaOCl	Sodium hypochlorite
ng	nanograms
NGS	Next Generation Sequencing
NHEJ	Nonhomologous end joining
nm	nanometre
nM	nanomolar
NosT	*Nopaline synthase* gene terminator
Nt	*Nicotiana tabacum*
OD	Optical density
ORF	Open Reading Frame
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
PDS	*Phytoene desaturase* gene
PEBP	Phosphatidyl ethanolamine-binding protein
PEBV	*Pea early browning virus*
PEG	Polyethylene glycol
pegRNA	Prime editing guide RNA
PMSF	Phenylmethylsulfonyl fluoride
PTGS	Post-transcriptional gene silencing

PVX	*Potato virus X*
RGR	ribozyme-sgRNA-ribozyme unit
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT-PCR	Reverse transcription PCR
RVD	Repeat Variable Diresidue
s	seconds
SD	short day
SDS	Sodium dodecyl sulfate
sgRNA	single guide RNA
SDN	Site-directed Nucleases
ssDNA	single-stranded DNA
STU	Single transcript unit
SYNV	*Sonchus yellow net rhabdovirus*
T7EI	T7 endonuclease I assay
Ta	Annealing temperature
TAE	Tris acetate EDTA buffer
TALENS	Transcription Activator-Like Effector Nucleases
TALEs	Transcription activator-like effectors
TBS-T	Tris-buffered saline tween buffer
TMV	*Tobacco mosaic virus*
TOM1	*TOBAMOVIRUS MULTIPLICATION 1* gene
ToMV	*Tomato mosaic virus*
tracrRNA	trans-activating crRNA
TRBO	TMV RNA-based overexpression
Tris	Tris(hydroxymethyl) aminomethane
Tris-HCl	Tris(hydroxymethyl) aminomethane hydrochloride
TRV	*Tobacco rattle virus*
TuMV	*Turnip mosaic virus*
USDA	United States Department of Agriculture
UTP	Uridine triphosphate
V	Volts
WT	Wild type
X-gal	5-bromo-4 chloro-3-indolyl-β-Dgalactopyranoside
ZFN	Zinc Fingers Nucleases

CHAPTER 1: INTRODUCTION

One of plant biotechnology's key challenges is the development of sustainable agriculture in order to feed the rising global population. The genetic improvement of plants enables important traits, such as flowering in crops, to be manipulated in order to create improved varieties. However, classical methods need extensive breeding programs, or the stable integration of foreign DNA into the plant's genome. As an alternative, genome editing technologies offer the possibility of introducing simple and rapid mutations into target genes of a desired plant species.

1.1 GENOME EDITING TOOLS

1.1.1 Double strand break repair mechanisms

Genome editing is defined by the World Health Organization as a "method for making specific changes to the DNA of a cell or organism. It can be used to add, remove, or alter DNA in the genome" (1). These changes in the genome can be achieved by site-directed nucleases (SDN), engineered enzymes that introduce a double-strand break (DSB) into the DNA, which is repaired by the own cell machinery. DSB repair mechanisms are divided in non-homologous end joining (NHEJ) and homology directed repair (HDR) pathways (Figure 1).

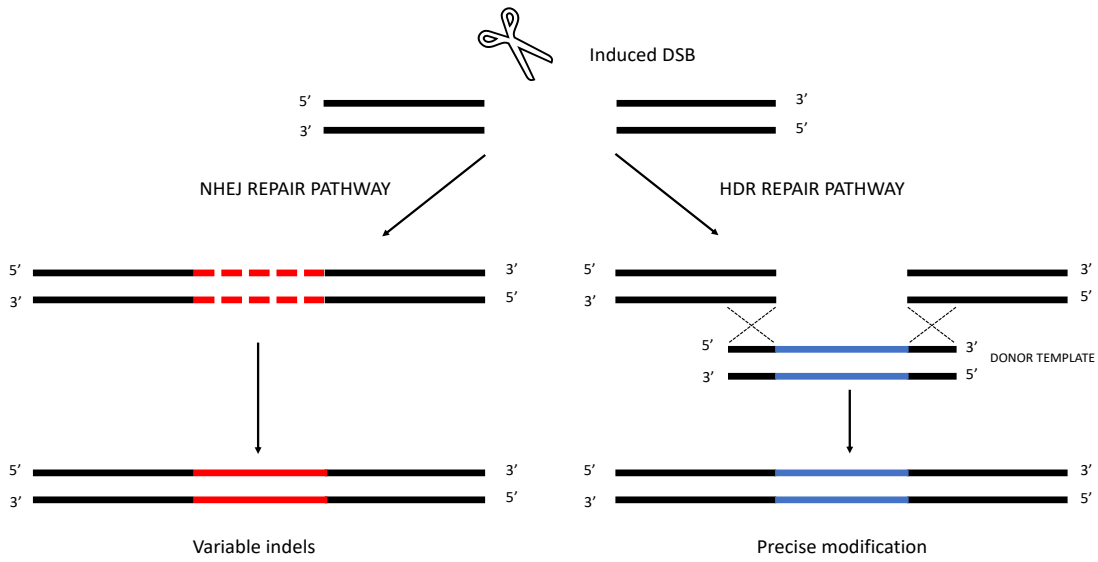


Figure 1. Double-strand break repair mechanisms. Nonhomologous end joining (NHEJ) pathway repairs the DSB introducing variable indels, while Homology directed repair (HDR) uses a template, creating a precise modification.

The NHEJ repair pathway generates deletions or insertions randomly. If this happens in a certain gene sequence, it could cause a frameshift of the reading frame. On the other hand, in the HDR pathway the DNA cleavage is repaired using a template, which could be an exogenous sequence (to induce from single nucleotide changes to bigger insertions/replacements) or using the sister chromatid. It has been reported that NHEJ is more frequent than HDR, due to the first one taking place throughout the cell cycle, meanwhile HDR happens only during S and G2 phases (2).

Therefore, site-directed nucleases can produce DNA alterations including insertions (both of cis and foreign DNA), deletions, a combination of those (indels) and edits of a few or single base pairs. Depending on the outcome of the DNA repair, SDN are classified into three categories. SDN-1 are based on the NHEJ mechanism, leading to a site-specific random mutation of a few bp or short indels, while SDN-2 and SDN-3 are based on the HDR mechanism. In the case of SDN-2, defined point mutations or substitution of less than 20 bp are achieved using a homologous repair template, obtaining a precise mutation or edit without the insertion of a foreign gene. SDN-3 technologies allow the insertion of a cisgenic or transgenic sequence which can be up to several kilobases long (3, 4).

Currently, the SDN technologies comprise Meganucleases, Zinc-Finger Nucleases (ZFN), transcription activator-like effectors nucleases (TALENs) and the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/CRISPR-associated nuclease 9 (Cas9) system.

1.1.2 Early genome editing tools

The first approaches for genome engineering were developed in the 1980s, wherein by using the cell's own machinery for homologous recombination, a defective gene in the chromosome of a mammalian cell was replaced with a donor DNA flanked by two homologous arms (5).

The earliest exploit of DSBs repair mechanism for genome editing was achieved using Meganucleases, also known as rare cutting endonucleases or homing endonucleases, enzymes of ~40 kDa (< 300 amino acids) which have a large recognition site between 12 to 40 bp (Figure 2) (6-8). These enzymes promote the lateral transfer of their own gene by first inducing site-specific DSBs and then,

repairing such breaks through recombination, in other words, “copying” the gene encoding them, a process known as “homing” (8). The most widespread meganucleases used for genome engineering are I-*Sce*I (from *Saccharomyces cerevisiae*) and I-*Cre*I (from *Chlamydomonas reinhardtii*). I-*Sce*I and I-*Cre*I are part of the LAGLIDADG family of homing endonucleases, called after the conserved domain found in all the proteins of this family (8).

However, there is a low probability of finding the restriction site for the homing nucleases in the genome (6), usually it was necessary to introduce the desired sites which had a low efficiency (9). To avoid this problem, hybrid endonucleases were developed. The first system was the Zinc Finger Nucleases by Kim *et al.* in 1996, where they reported the linkage between zinc fingers proteins and the cleavage domain of *Fok*I endonuclease (10). The binding to a specific DNA sequence is determined by the Cys_2His_2 zinc fingers, where each finger of ~30 amino acids and a zinc atom interacts with a specific nucleotide triplet. These can be engineered in arrays between 3 to 6 individual fingers, binding to 9 to 18 bp in total (9). On the other hand, *Fok*I is a type II restriction endonuclease, wherein each terminal of the protein has independent activity, its N-terminal is a DNA-binding domain while the C-terminal has nonspecific DNA-cleavage activity.

To cut the double strand DNA (dsDNA), *Fok*I must dimerize. To achieve this activity, two monomeric sets of ZFNs bind to opposite strands of DNA at a specific distance of 5 or 7 bp to allow the dimerization of the cleavage units (Figure 2) (9, 11). This last requirement is one of the advantages of the ZFNs, as monomeric units do not exhibit activity, and as each ZFNs comprises several zinc fingers that bind to consecutive nucleotides, two units would give to the system a high specificity (11).

In 2009 it was described how the transcription activator-like effectors (TALEs) from the plant pathogenic bacteria of the genus *Xanthomonas* can bind specifically to DNA. The effectors are comprised of a N-terminal translocation signal, a nuclear localization signal (NLS), an acidic transcriptional activation domain at the C-terminal and a central domain of tandem repeats, typically of 33 to 35 amino acids, where positions 12^{th} and 13^{th} are highly variable (known as Repeat Variable Diresidue or RVD). Each repeat binds to one nucleotide in the 15-20 bp target site, wherein the specificity of each repeat is determined by the RVD (Figure 2) (12, 13).

The correspondence between the RVDs and the nucleotides in the target DNA sequence was described by Boch *et al.* (2009) (12) and Moscou & Bogdanove (2009) (13) independently, developing a code to create custom made TALEs.

In 2010 Christian *et al*. (2010) fused two well-known TALEs to the catalytic domain of *Fok*I endonuclease to induce DNA DSBs in a target specific approach (Figure 2), creating the TALENs system. Moreover, using the code deciphered previously they created and tried custom TALEs repetitions to recognize new target sequences (14). Independently, Li *et al.* (2011) also fused two different TALEs from *Xanthomonas oryzae* pv. *oryzae* (Xoo) rice pathogen to the *Fok*I endonuclease (15). Following these two pieces of work, the use of the TALENs technology sky-rocketed, being quickly described in tobacco (16), mammalian cells (17), human cells and *A. thaliana* protoplasts (18).

1.1.3 CRISPR/Cas9 system and its use as a genome editing tool

An alternative method was proposed based on the CRISPR/Cas system. CRISPR was first identified in 1987, during the study of the gene related with the isozyme conversion of alkaline phosphatase in *E. coli*, where the authors noticed the same palindromic sequence in different clones (19). In 1993 Mojica *et al.* (1993) described similar repeats in the archaea *Haloferax mediterranei* during their studies of its adaptation to different salinities and in 2000 they proposed that these repeats found in archaea and bacteria would have a common function in prokaryotes (20, 21) . Then, in 2002 Jansen *et al*. (2002) proposed to name these repetitions according to the acronym CRISPR. Furthermore, in this study by comparing the genetic environment of the CRISPR loci they identified for the first time four CRISPR-associated genes (*cas1 – cas4*), which encode the proteins Cas1 to Cas4, possibly related to the function of the CRISPR loci (22). Successive papers would try to elucidate the role of CRISPR, proposing its origin as extrachromosomal or from bacteriophages (23-26). Finally, in 2007 it was experimentally demonstrated that CRISPR associated with the *cas* genes acts as an adaptive bacterial immune system, where small sequences from exogenous DNA are integrated into the bacterial genome, which when transcribed act as guides for the Cas proteins to cut the foreign DNA (27).

In August 2012, Jinek *et al.* (2012) published the use of the type II Cas9 protein as a programable endonuclease. The authors first demonstrated the cleavage activity of the protein purified from the organism *Streptococcus pyogenes* (*Sp*Cas9), which is directed by the crRNA (CRISPR RNA) complementary to the sequence to be cut and additionally by the trans-activating crRNA (tracrRNA). Furthermore, a protospacer adjacent motif (PAM) sequence of NGG adjacent immediately downstream to the crRNA binding region was required for target DNA binding. Moreover, the authors stated that Cas9 protein has homologous domains to HNH and RuvC endonucleases and described that the enzyme cuts dsDNA, wherein the HNH domain cleaves the complementary strand to the crRNA-tracrRNA and the RuvC-like domain cleaves the noncomplementary DNA strand. This cleavage is blunt-ended and 3-4 bp upstream of the PAM site. Finally, they created a single guide RNA (sgRNA) using a linker loop to connect the tracrRNA and the crRNA (Figure 2). The highlight of the paper was the ability to programme the system, demonstrated by the design of five chimeric sgRNAs to target the *GFP* gene, which efficiently directed the cleavage of the *GFP* dsDNA by Cas9. In summary, this paper demonstrated the potential of the Cas9 protein as a tool for gene editing (28).

Almost in parallel, in September of 2012, Gasiunas *et al*. (2012) published their work on the Cas9-crRNA complex from *S. thermophilus*. The authors isolated and purified the Cas9-crRNA complex and demonstrated its cleavage activity on the double strand DNA, creating blunt ends 3 bp upstream of the terminal end next to the PAM sequence. Furthermore, they demonstrated that the 42 bp crRNA includes 20 nucleotides which are complementary to the sequence in the target DNA. In similar experiments to Jinek *et al*. (2012), the authors also custom designed guide RNAs to cut a target site of their choice and determined that the HNH and RuvC domains of the Cas9 protein bind to each strand of the DNA to pursue its endonuclease activity (29).

In the following months, several groups would report the successful use of the CRISPR/Cas9 system for genome editing in human (30-32) and mouse cells (31). Improvements to the system included the use of a human codon optimized *Sp*Cas9, the addition of Nuclear Localization Signals (NLS) for ensuring the efficient translocation of *Sp*Cas9 to the nucleus of mammalian cells or the use of the RNA polymerase III (Pol III) promoters to drive the expression of the guide RNA (31, 32)

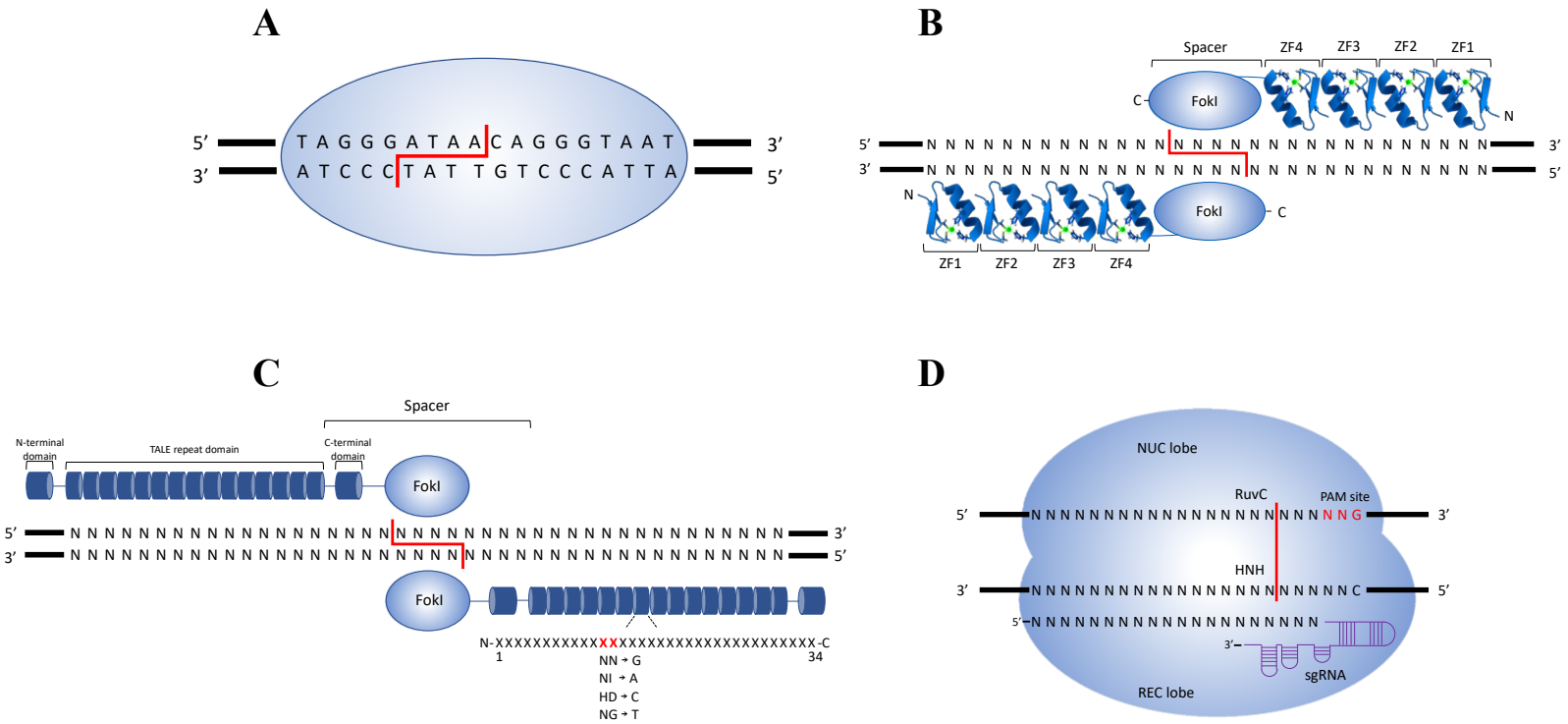


Figure 2. Schematic diagram of the genome editing technologies A. Meganucleases. The enzyme recognizes a sequence of 18 bp and cleaves the DNA creating sticky ends. **B. Zinc Finger Nucleases**, where each finger binds specifically to nucleotide triplets and are linked to the cleavage domain of *Fok*I endonuclease. **C. Transcription activator-like effector nuclease (TALEN)**. TALE repeats are comprised by a N-terminal, a C-terminal (both required for DNA binding) and a central repetition domain, which binds specifically to one nucleotide depending on two amino acids (in red) **D. CRISPR/Cas9**. Cas9 endonuclease consists of two lobes, where each of them binds to a DNA strand. The sgRNA guides the Cas9 to the complementary sequence to carry-out its cleavage activity. A PAM site is required as well for recognition site. Red lines indicate cut sites of each system

The use of a site directed mutated *Sp*Cas9 in the RuvC I domain (D10A) changes its activity from nuclease to nickase and promotes the HDR repairing system rather than the NHEJ mechanism, allowing a dsDNA donor construct to integrate into a desired site of the genome (31, 32). Finally, it has been possible to simultaneously express two guide RNAs, enabling efficient cleavage at different target sites (31).

In plants, three papers were published in August 2013 describing the use of CRISPR/Cas9 for genome editing. The authors tested a plant codon optimized *Sp*Cas9 with two NLS attached to each end and a sgRNA under the U3 or U6 RNA Pol III promoter to target the *phytoene desaturase (PDS)* gene in rice, *A. thaliana* and *N. benthamiana* protoplasts and plants obtaining highly efficient targeted mutagenesis rates (33-35). Moreover, the authors tried targeting different genes in protoplasts isolated from wheat, rice, *A. thaliana* and *N. benthamiana* finding different rates of efficiency (33, 34). Finally, as described in human cells, it was possible to integrate a donor DNA into the target site using the HDR pathway (33, 34).

Overall, the CRISPR/Cas9 system has proven to be the most convenient way to produce different genome editing and has proved to be an efficient, simple and low-cost method to modify gene function. For ZFNs and TALENs, the redesign of the target site domains requires the assemble of DNA regions between 500 to 1500 bp for each site, which is laborious and expensive. In contrast, the guide sequence of CRISPR/Cas9 is 20 bp long, so it is easier to redesign for different targets. This advantage raises the opportunity to use several guide RNAs in parallel to target multiple sites simultaneously, or to pursue precise deletions in a genomic region (36, 37). Another advantage of CRISPR/Cas9 over ZFNs and TALENs is the high amount of potential target sites. For *Sp*Cas9 the recognition site is 20 bp plus 3 bp for the PAM site, which appears on either strand of DNA on average once per 8 bp in the human genome (31). In contrast, for TALENs the target site frequency is estimated every ~35 bp, whilst for ZFNs it is every ~500 bp (18, 36).

A great disadvantage of the CRISPR/Cas9 system though, is the possibility of off-target cleavage. Even though this effect has also been reported for ZFNs (38, 39) and less prevalently for TALENs (40-42), due to its widespread use, several groups have studied the specificity of the CRISPR/Cas9 activity. This is of special concern

because of the potential use of the system for gene therapy. It has been described that the 5' distal region of the guide RNA is more tolerant for mismatches than the 3' proximal region in respect to the PAM. Nonetheless, this is target-site dependent since it has been reported that some guide RNAs tolerate single or double mismatches even in the 3' half while others are sensitive to changes in the 5' portion (43-45). Studies focused on the detection of editing of non-target sites found significant rates of off-target mutagenesis due to this mismatch tolerance (43, 44, 46, 47).

In order to minimize the off-target cleavage activity, it has been proposed to reduce the size of the guide RNA to increase its specificity (46, 48), decrease the amount of *Sp*Cas9 and guide RNA delivered (but this also leads to a decrease of the on-target editing) (44, 46), the use of the nickase Cas9 to induce the HDR repair pathway (49-51) or a catalytic inactive Cas9 fused to a *Fok*I endonuclease which, as mentioned before, must dimerize to exert its function (52, 53). In plants, off-target mutations can be removed or segregated away by backcrossing, however this is not possible for asexually propagated species (54, 55). Moreover, it has been suggested that Cas9 off-target activity is less frequent in plants, mostly due to low expression levels of Cas9 protein (56-58).

Another disadvantage of the CRISPR/Cas9 system is the size of the Cas9 enzyme which makes it more difficult to deliver using viral vectors due to their limited cargo capacity (eg. less than 5 kb for the adeno-associated virus - AAV) (37). *Sp*Cas9 gene is ~4.2 kb, while a TALEN monomer is ~3 kb and a ZFN monomer is ~1 kb, Alternative, smaller, Cas9 proteins have been isolated from different organisms in recent years. Two classes of CRISPR/Cas system have been described. Class 1 comprises type I, III and IV, characterized by the formation of a large and heteromeric Cas complex of 4 to 7 proteins. On the other hand, Class 2 includes type II, V and VI, distinguished for being a simple multidomain protein (59). Cas9 orthologs exhibit variable sizes and increased specificity, either due to more complex or larger PAM sites, thus increasing the versatility of the system.

Some of the orthologs described in the literature are listed in Table 1.

Table 1. Cas9 Orthologs

Organism	Acronym	Protein length	Protospacer length	PAM site	Reference
Streptococcus pyogenes	*Sp*Cas9	1368 aa	17 - 20 bp	5'-NGG-3'	(28)
Streptococcus thermophilus	*St*1Cas9	1121 aa	19 – 20 bp	5'-NNAGAAW-3'	(60-62)
	*St*3Cas9	1393 aa	19 bp	5'-NGGNG-3'	(29, 61)
Staphylococcus aureus	*Sa*Cas9	1053 aa	20-24 bp	5'-NGGRRT-3'	(60, 63)
Neisseria meningitidis	*Nm*Cas9	1082 aa	24 bp	5′-NNNNGATT-3′	(62, 64)
Campylobacter jejuni	*Cj*Cas9	984 aa	22 bp	5'-NNNNACAC-3' 5'-NNNNRYAC-3'	(65)
Streptococcus canis	*Sc*Cas9	1375 aa	20 bp	5'-NNG-3'	(66)
Geobacillus stearothermophilus	*Geo*Cas9	1087 aa	22 bp	5'- NNNNCRAA-3'	(67)
Francisella novicida	*Fn*Cas9	1629 aa	22 bp	5'-NGG-3'	(68)
Clostridium cellulolyticum	*Cc*Cas9	1027 aa	20 bp	5′-NNNNGNA-3′	(69)
Brevibacillus laterosporus	*Blat*Cas9	1092 aa	21 bp	5′-NNNNCNDD-3′	(70)
Francisella novicida	*Fn*Cas12a	1300 aa	23-25 bp	5′-TTN-3′ 5′-CTA-3′	(71)
Acidaminococcus sp BV3L6	*As*Cas12a	1307 aa	23 -25 bp	5′-TTTV-3′	(71)
Lachnospiraceae bacterium	*Lb*Cas12a	1228 aa	23 -25 bp	5′-TTN-3′	(71)
Bacillus hisashii	*Bh*Cas12b	1108 aa	20 bp	5'-TTN-3'	(72)
Alicyclobacillus acidiphilus	*Aa*Cas12b	1129 aa	20 bp	5'-TTN-3'	(73)
Biggiephage family	Cas12j	700-800 aa	14-20 bp	5′-TBN-3′	(74)
Deltaproteobacteria	*Dpb*CasX	986 aa	20 bp	5'- TTCN-3'	(75)
Planctomycetes	*Plm*CasX	~980 aa	20 bp	5'- TTCN-3'	(75)

aa: amino acid; bp: base pair; N: any nucleotide; **W**: A or T. **R**: A or G. **Y**: C or T.
D: A, G or T. **V**: A, C or G. **B**: G, T, or C

Cas12a (formerly known as Cpf1), Cas12b (previously known as C2c1) and CasX enzymes are class 2, type V endonucleases, which in contrast to Cas9, lack the HNH endonuclease protein domain. As shown in Table 1, their PAM sites are T rich domains, instead of G rich as Cas9. Moreover, these enzymes cut the dsDNA distant from the PAM site, leaving overhanging ends, in contrast to Cas9 which creates blunt ends (71-73, 75). From this group, Cas12a is the most studied as it does not need a tracrRNA and uses only a single 42-44 bp crRNA of which 23-25 bp correspond to the protospacer, in contrast to Cas9, Cas12b and CasX. Moreover, Cas12a enzymes also exhibit RNase activity, which allows to target multiples sites by delivering a long tandem of pre-crRNA, which will be processed to mature crRNA and used by the enzyme (71, 76).

Recently, a smaller Cas has been described, isolated from huge bacteriophages. CasΦ or Cas12j, is a compact enzyme of ~70 to 80 kDa, half of the size of Cas9 or Cas12a, making it ideal to be delivered by virus vectors. The DNA target recognition is by a crRNA spacer of 14 to 20 bp and cleavage products are staggered 5' overhangs of 8-12 bp. The authors showed successful genome editing events *in vitro*, HEK293 human cells and *A. thaliana* protoplast cells (74).

Finally, other Cas enzymes have been reported to have different activities. For example, class 2, type VI Cas13a (C2c2) has an RNA-guided RNA endonuclease activity towards single stranded RNA (77).

1.1.4 Protein structure of Cas9

The characterization of new Cas proteins was possible due to the publication of the crystal structure of *Sp*Cas9 in complex with the sgRNA and the DNA target site by Nishimasu *et al*. (2014) (78). The authors described a bilobed architecture of the protein, consisting of a target recognition (REC) lobe and a nuclease (NUC) lobe, which comprise the HNH and RuvC nuclease domains plus a carboxy-terminal PAM interacting (PI) domain. The connection of both lobes is through an arginine cluster known as the bridge helix and a linker. Moreover, they described that the negatively charged sgRNA:DNA complex is placed between both lobes in a positively charged groove.

In parallel, Jinek *et al.* (2014) published their work about the characterization of the crystal structures of *Sp*Cas9 and *Actinomyces naeslundii* Cas9 (*Ana*Cas9). The authors also proposed a conserved bilobed structure between both enzymes comprises by a nuclease domain lobe and a α-helical lobe (recognition lobe in Nishimasu's work), wherein this region is shorter in *Ana*Cas9. Moreover, electron microscopy (EM) studies revealed a conformational rearrangement of the Cas9 loaded with the sgRNA forming a central channel for the interaction with the target DNA (79).

Next, Nishimasu *et al.* (2015) published the crystal structure for *Staphylococcus aureus* Cas9 (*Sa*Cas9), which as described in Table 1, is smaller than *Sp*Cas9 and recognizes the target DNA by a different PAM site. The authors reported that *Sa*Cas9 is a bilobed protein, comprising a REC lobe and a NUC lobe, connected by an arginine-rich bridge helix and a linker loop. Between these two lobes is placed the heteroduplex of sgRNA and the target DNA. The NUC lobe is divided in the RuvC, HNH, WED and PI domains, which is also split into a Topoisomerase II (TOPO) homology region and a C-terminal domain (CTD). Moreover, it is described that *Sa*Cas9 also requires a conformational change in its structure to exert its function. Whilst the general structure is similar between *Sp*Cas9 and *Sa*Cas9, differences in the REC, WED and PI domains explain the size and PAM recognition differences. REC and WED regions interact with the tracrRNA, which is comprised of two stem loops instead of three as is the case for *Sp*Cas9. Furthermore, differences in the length of a region of the PI domain explains the longer PAM for *Sa*Cas9 (80).

Figure 3 shows a comparison between both structures, highlighting the key domains of both proteins.

1.1.5 Rational modification of the Cas9 protein

The structural studies of Cas9 enzymes contributed to the rational modification of the proteins. As mentioned previously, one of the main disadvantages of the method is the size of the *Sp*Cas9 enzyme and even though new proteins have been discovered and characterized, a second approach has been described. This consists of splitting the Cas9 protein into two lobes to diminish its size, which can be delivered by two separated vectors. Several papers have reported this approach, demonstrating its effectiveness.

Wright *et al*. (2015) reported the division of *Sp*Cas9 into the α-helical lobe and the nuclease lobe (The RuvCI and RuvCII domains were fused by a glycine-serine-serine linker), where both halves were stably expressed and purified. Moreover, the authors stated that each lobe does not exhibit cleavage activity on its own, but they do after dimerization by the formation of a ternary complex using the sgRNA as a scaffold. Interestingly, even though the cleavage activity of the re-constituted enzyme is initially ~10-fold slower than the WT, after about 5 minutes both reach the same end point (81).

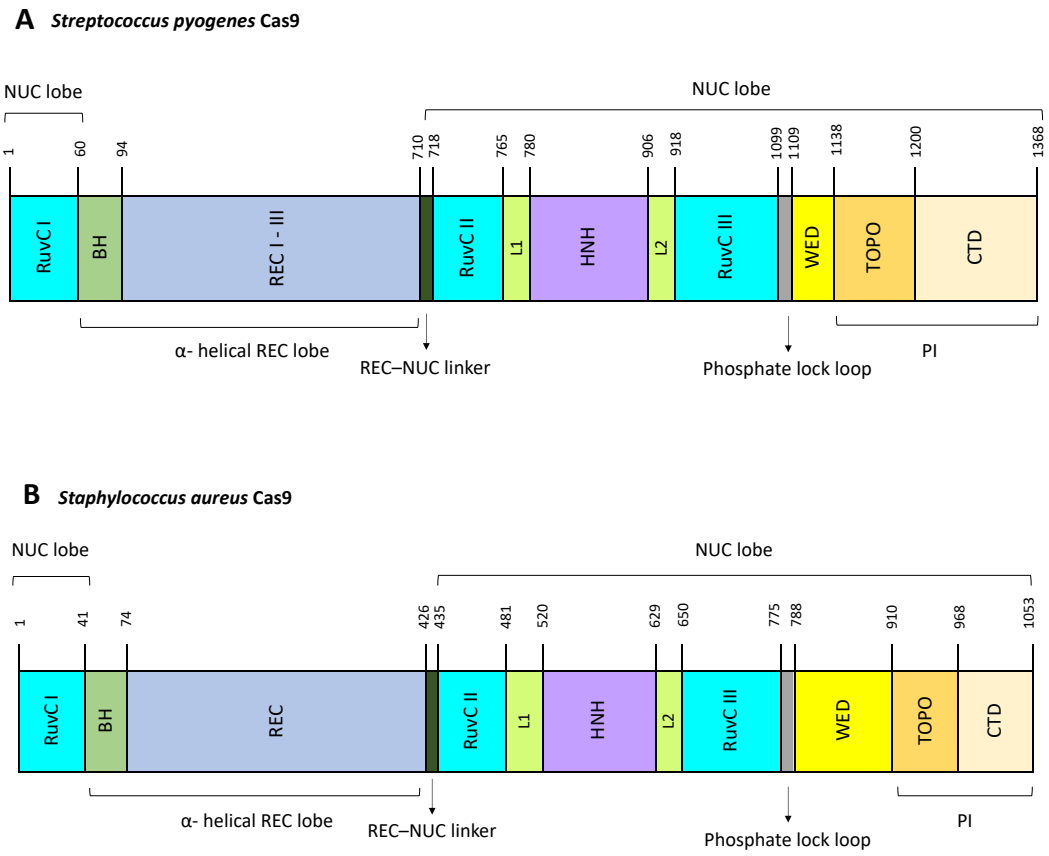


Figure 3. Protein domain organization of Cas9 endonucleases. A. *Sp*Cas9. B. *Sa*Cas9
Proteins are organized in a Nuclease (NUC) and target recognition α- helical (REC) lobes linked by an arginine rich bridge helix (BH) and a linker loop. The NUC lobe comprises the RuvC, HNH, WED and PI regions. L1 and L2 linkers connect the RuvC and HNH. The WEDGE (WED) domain is a highly divergent region among Cas9 proteins. The PAM Interacting (PI) domain is divided into a Topoisomerase II (TOPO) homology region and a C-terminal domain (CTD). Even though their domain organization is similar, differences in the length of the REC, WED and PI domains explain the size and PAM recognition differences. Numbers indicate amino acid position.

In order to increase the efficiency of the re-association of both monomers and, to have some control over the dimerization, Zetsche *et al.* (2015) identified eleven split sites in loops or unstructured regions and fused each N- and C- terminal portions to the FK506 binding protein 12 (FKBP) and FKBP rapamycin binding (FRB) domains, respectively. The re-assembly could then be induced by rapamycin, which stimulates the association of FKBP-FRB domains. The group detected targeted mutagenesis mediated by all split-*Sp*Cas9 sets in cells treated with rapamycin, but the efficiency varied depending on the split site used (82).

On the other hand, Nishimasu *et al.* (2015) split *Sa*Cas9 at two flexible regions (amino acids 430/431 and 739/740), where both showed robust cleavage activity. Then, the authors tested the induction of the assemble using the abscisic acid (ABA) sensing system and two versions of the FKBP/FRB system, where all of them exhibit cleavage activity (80).

Furthermore, several groups have been developing a strategy based on re-assembling the split-Cas9 using inteins (83-87). This system enables the delivery of the split-Cas9 in two separate vectors, with the split Cas9 protein domains reconstituting post-translationally. Inteins are known as protein introns due to their capacity to self-splice themselves out of a sequence after translation and form a covalent peptide bond between the remaining flanking regions (exteins) without leaving a scar. In order to increase the efficiency of the splicing, it is crucial that the split sites are surface exposed and that the first amino acid of the C-terminal portion of the exteins must be Cys, Ser or Thr to ensure an effective nucleophilic attack. The mechanism of the protein splicing is divided into the following steps: 1) association between the inteins, 2) N-S acyl rearrangement, where the amide peptide bond between the N-extein and the N-intein is converted to a (thio) ester bond and then the Cys, Ser or Thr residues of the C-extein nucleophilic attack this new bond, forming a linear (thio) ester intermediate, 3) transesterification step and formation of a branch (thio) ester intermediate, where the N-extein is released from the N-intein to join the C-extein by a (thio) ester bond 4) Resolution of the branch intermediate by the cyclization of a highly conserved asparagine (Asn) at the end of the C-intein, which leads to the separation of the inteins and exteins and 5) Formation of the peptide bond between both exteins through an N-S acyl shift from the free amino group of the C-extein to the (thio) ester bond formed in step 3 (88, 89). This reaction

mechanism is illustrated in Figure 4. Interestingly, several authors have demonstrated by *in vitro* assays at 37°C that splicing of the inteins is a fast process, being able to detect the re-assemble exteins in less than 60 seconds (90-92), whilst *in vivo* assays in *E.coli* shown re-associated units after 2 hours (93).

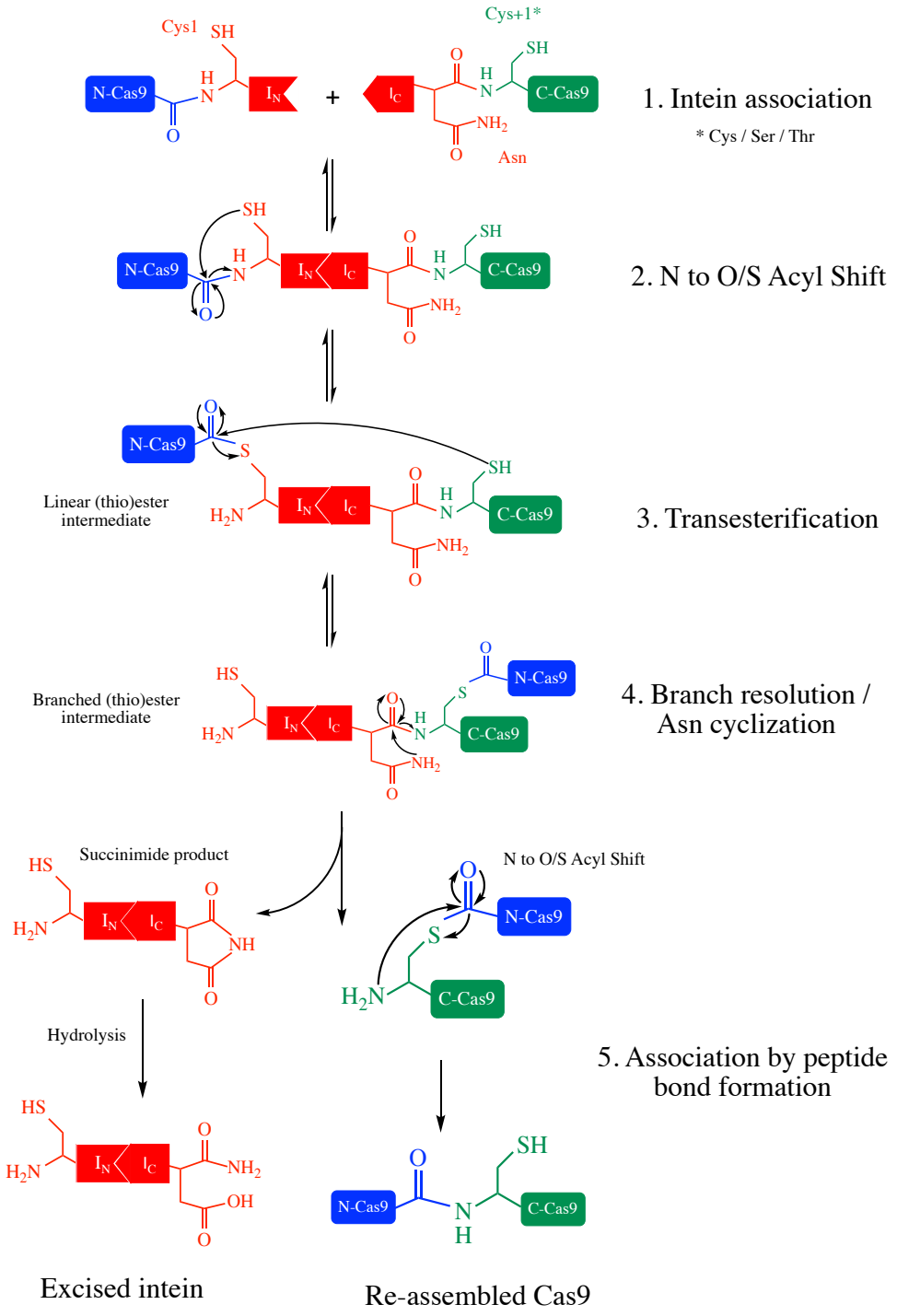


Figure 4. Re-association of a split *Sp*Cas9 protein by inteins. The mechanism is divided in five steps, leading to the re-assembly of the N-*Sp*Cas9 (blue) and C-*Sp*Cas9 (green) by a covalent peptide bond and the release of the re-associated intein (red). Cys and Asn amino acids are key to carry out different steps of the reaction.

1.1.6 Improvements to the sgRNA expression

The expression of multiple sgRNAs to edit different genomic sites simultaneously have also been developed. Traditionally, sgRNAs are produced *in vivo* using a vector containing a promoter for the RNA Pol III, which transcribes small RNA. It had been described that several sgRNAs can be expressed in tandem under this system. For example, Le Cong *et al*. (2013) expressed two sgRNAs to target two different human *loci* or a single human *locus* to create a larger deletion (31). In plants, Xing *et al*. (2014) and Ma *et al*. (2016) have developed several sgRNA module vectors where two or more modules can be assembled by cloning (94, 95).

Alternatively, sgRNAs can be expressed simultaneously separated by tRNAs. Xie *et al*. (2015) created a polycistronic unit of up to eight tRNA-gRNA, where endogenous cell RNases, such as RNase P and RNase Z in plants, cleave both ends of the tRNA precursor, releasing the individual sgRNAs. Moreover, the authors show that tRNA genes can enhance the activity of the RNA Pol III, since they contain internal promoter elements to recruit the polymerase (96). Moreover, in transgenic plants expressing Cas9, tRNAs fused to sgRNAs promote their cell-to-cell movement, improving genome editing efficiency, which is transmissible to the next generation (97).

However, the use of RNA Pol III promoters has some disadvantages. Usually, the U6 or U3 snRNA housekeeping gene promoters are used, which doesn't allow spatiotemporal, or specific cell/tissue mutations. As an alternative, the expression of the sgRNA using RNA Pol II promoters have been suggested using self-cleaving ribozymes to create an artificial gene (Ribozyme-sgRNA-Ribozyme) (98). It had been described that ribozymes have a rate of self-cleavage of 1.2 min^{-1} and since they self-excise from the construct, any exogenous machinery is needed (99). Considering the advantages of this approach, it has been described its use to deliver multiple sgRNAs in tandem (100), or optimizing the CRISPR/Cas9 system, creating a single transcription unit with the *Cas9* (101, 102) or *Cas12a* gene (103)

Until now, the use of a split Cas9 fused to the sgRNA flanked by ribozymes has not been tested in plants. Considering the encouraging results of this method in general, but particularly the intein studies, this thesis describes the adaptation of this system to edit a desired gene target in a crop model.

1.2 PLANT VIRUSES AS VECTORS FOR GENE EXPRESSION

Classically, the delivery of a desired gene of interest in plants has been done through techniques which lead to the stable integration of a transgene in the plant's genome. Usually, the delivery methods are: 1) physical, comprising biolistic and electroporation, which have a low integration efficiency of the delivered DNA into the genome and cause cell damage or toxicity, 2) chemical, such as polyethylene glycol (PEG) transformation of protoplasts, and 3) biological, via *Agrobacterium tumefaciens* transformation. The latter is a low cost and efficient method of delivering a gene of interest into plant cells for the expression of recombinant proteins.

The insertion of the gene of interest into the genome of the host plant by these methods leads to the creation of a stable, transgenic line. Even though this process allows the production of heterologous proteins over successive generations, creating these transgenic lines is time-consuming and generally requires tissue culture regeneration. To overcome this, the possibility of transiently expressing the gene of interest, avoiding its integration into the genome, has been described. In this case, the production of a recombinant protein can be detected even three hours after the delivery of the gene, and up to ten days post infection, with the peak of expression being between 18 to 48 hours, but just in the infiltrated leaves (104, 105).

The use of plant virus vectors has also been proposed for transient gene expression in plants. Plant virus vectors have been developed from double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), and positive single-stranded RNA (+ssRNA) viruses. The first generation of virus expression systems are based on the use of the full-length cDNA of the virus, wherein the gene of interest is expressed under the promoter of the viral coat protein (CP), or from a sub-genomic promoter (sgP), inserted into the viral genome (106). A great advantage of the use of viral vectors is the spread of the virus expressing the transgene into mature leaves within days, rather than weeks or months as for transgenic plants (107). In some cases, systemic infection into meristems (108) and roots (109) has even been reported. Moreover, RNA viruses do not integrate their genome into the plant genome, making them a good option for non-integrative gene editing approaches.

Nevertheless, a negative correlation between the cargo capacity of viral vectors and their stability has been reported. For example, the dsDNA *Cauliflower mosaic virus* (CaMV) has a restricted amount of space inside the capsid. On the other hand, (+) ssRNA viruses build their capsid around their genome independently of its size, so they do not have this limitation. But an instability of the insert has been reported in vectors containing duplicated sgP since the viral replicase changes templates at such repeated sequences leading to mutations or the loss of the transgene, particularly when it has a large size (110). To overcome this, different sgP from related viruses have been used in these vectors (109, 111). Another option is fusing the protein of interest to the CP using a 2A catalytic peptide, which promotes the cleavage of the fused protein during the translation (105, 111).

Second generation virus vectors, known as "deconstructed vectors", comprise the minimum genes for replication of the virus and the gene of interest. These vectors do not have size, host or tissue limitations, however due to the lack of genes essential for transport and assembly, the virus must be delivered, eg. via *Agrobacterium*. Some of these vectors have also been engineered for systemic infection by retaining some movement genes (112).

1.2.1 Potato virus X as a vector for gene expression in plants

Chapman *et al*. (1992) described the development of *Potato virus X* (PVX) as a vector for gene expression (107). PVX is a monopartite, single-stranded, positive-sense RNA virus from the genus *Potexvirus* in the family *Flexiviridae*. The PVX genome encodes five open reading frames (ORF), the 5' end has a methylguanosine cap and the 3' end has a polyadenylated tail. The first ORF is the 166 kDa viral RNA dependent RNA polymerase (RdRp) for viral RNA synthesis and replication. Next, three overlapping ORFs of 25 kDa, 12 kDa and 8 kDa known as the triple gene block encode the virus movement proteins. The final ORF corresponds to the coat protein, which is also needed for transport through the plant (105, 107, 111). *Potato virus X* infects more than 240 species in 16 families, mostly members of the *Solanaceae family*, such as potato, tomato *N. benthamiana*, and *N. tabacum* by mechanical transmission and plant-to-plant contact (113).

In Chapman's study, the authors duplicated the coat protein promoter to drive the expression of the *β-glucuronidase* (GUS) gene in a full-length cDNA clone of PVX,

reporting high levels of GUS in systemic and infected leaves of *N. clevelandii* and *N. tabacum* cv. Samsun plants (107).

Several PVX vectors have been developed with different characteristics. For example, PVX pP2C2S was created by adding a T7 RNA polymerase promoter before the RdRp ORF, a multiple cloning site to clone the desired transgene under the duplicated coat protein promoter and a *Spe*I restriction site at the end of the polyA tail, in order to linearize the vector and synthetize viral copies by *in-vitro* transcription, which enables infection of plants by mechanical rubbing (Figure 5) (114).

To avoid the transgene becoming unstable, heterologous sgP introduced upstream of the *cp* gene had been successfully tested (111). A binary T-DNA vector for *Agrobacterium* transformation has been used, wherein the PVX cDNA was positioned between a CaMV 35S promoter and a nopaline synthase terminator. Furthermore, as in pP2C2S, this vector contains a duplicated CP promoter followed by a polylinker, for expression of desired transgenes (115). Finally, this last PVX vector was adapted for easy cloning of transgenes using the Gateway™ cloning system, wherein the gateway cassette was placed downstream of the duplicated CP promoter (110, 116).

Some disadvantages in the use of PVX as a viral vector have been reported. For example, PVX is unable to enter the meristem, so the transgenes are not inherited (108, 110). Moreover, PVX is highly infectious due to its mechanical transmission, so special care must be taken to prevent the escape of the virus into the environment (110).

1.2.2 Tobacco rattle virus as a vector for gene expression in plants

Another virus engineered as viral vector is *Tobacco rattle virus* (TRV). It has been reported that this virus can infect approximately 400 plant species from 50 families, a wider host range than PVX, although in some occasions the infection remains in the roots (113). TRV is a bipartite, positive single-stranded RNA virus from the genus *Tobravirus* in the family *Virgaviridae*. Of the two viral genomic RNAs, RNA1 (also known as TRV1) encodes a protein of 134 kDa with homology to a helicase, and immediately downstream in the same reading frame a 194 kDa RNA

dependent RNA polymerase (RdRp). Next is encoded a 29 kDa movement protein and a 16 kDa cysteine-rich protein. On the other hand, RNA2 (or TRV2) encodes the coat protein and two additional proteins (2b and 2c) downstream involved in the transmission of the virus by nematodes of the family *Trichodoridae* (117, 118). Early TRV virus expression vector cDNA clones were constructed by replacing the 2b and 2c genes of TRV2 with a gene of interest under the control of the *Pea early browning tobravirus* (PEBV) CP promoter, and *in vitro* transcribing the viral RNA using the T7 RNA Polymerase. In contrast, TRV1 transcripts were synthetized from full-length cDNA clones of isolated strains or purified from natural virus infections (109, 119). Next, binary T-DNA vectors were developed, where both TRV1 and TRV2 viral cDNAs were cloned between the CaMV 35S promoter and a nopaline synthase terminator. A self-cleaving ribozyme (Rz) was added to release exact 3' end of viral RNAs (Figure 5) (120, 121).

As in the early TRV2, the 2b and 2c were replaced with a multiple cloning site for insertion of transgenes, but, since this vector was developed for gene silencing, a duplicate sub genomic promoter was not included (120, 121). Furthermore, a TRV2 vector compatible with the Gateway™ cloning system was developed for gene silencing (122) and gene expression by inserting the PEBV CP subgenomic promoter (123). Also, a Golden Gate Assembly TRV2 vector is available for easy modular cloning of desired fragments for gene silencing (124).

It has been reported that the 16 kDa protein encoded in TRV1 acts as a suppressor of RNA silencing and allows the virus to infect meristems of *N. benthamiana* plants. Moreover, the authors demonstrate that this protein can act in *trans*, allowing a heterologous virus to invade the meristem as well (108).

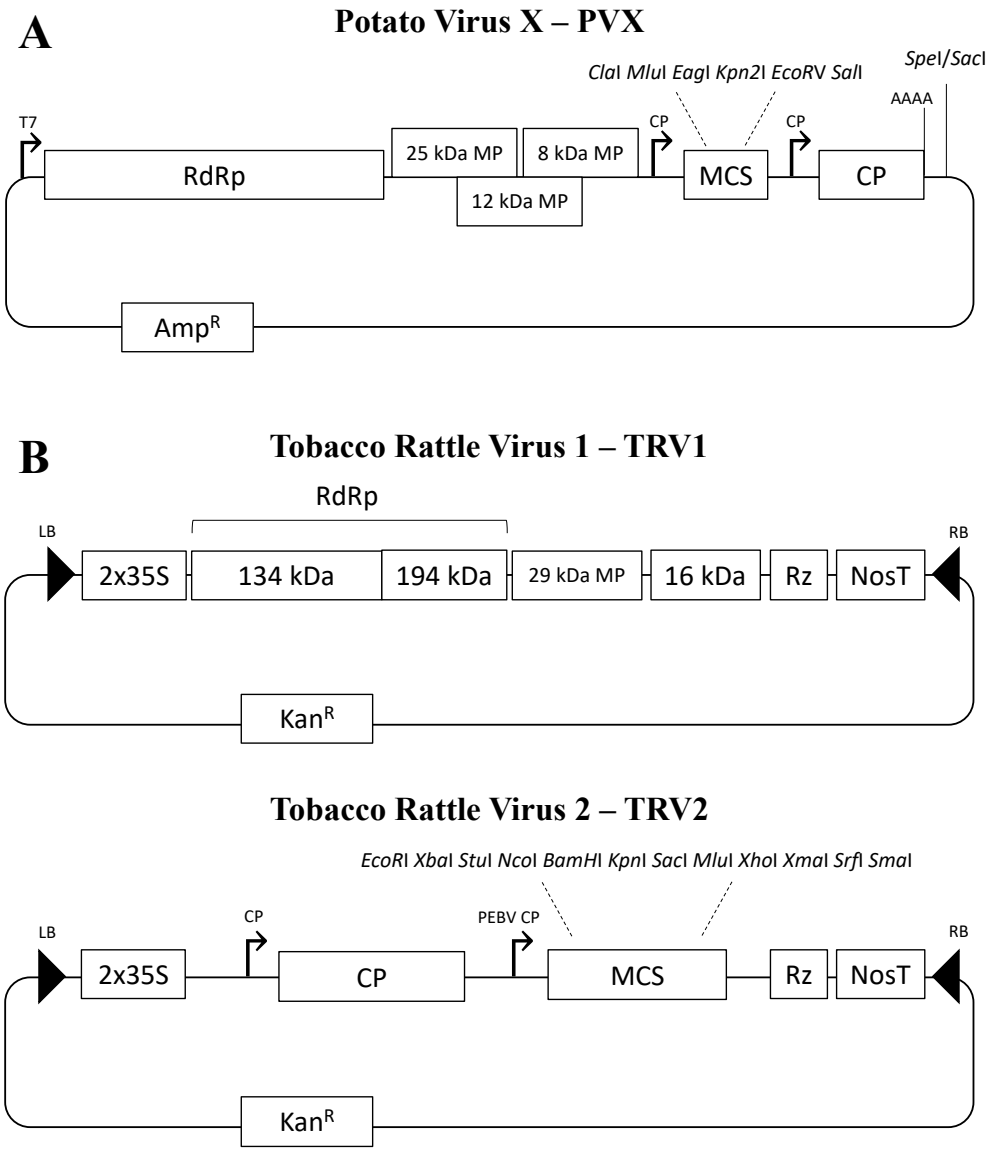


Figure 5. Genome organization of PVX and TRV viral vectors. A. PVX is comprised by the *RdRp* gene, three ORFs for movement proteins (MP) and the c*oat protein* (CP) gene. The expression of the transgene is under a duplicated coat protein promoter. For infection, the vector is linearized with *Spe*I and *in vitro* transcribed using the T7 RNA Polymerase promoter (T7). **B.** TRV is a bipartite vector, consisting in TRV1, encoding the *RdRp* gene, a movement protein gene and 16 kDa cysteine-rich protein gene. TRV2 encodes the *coat protein* gene and is used as vector. The transgene is expressed under a PEBV coat protein promoter. Both plasmids had been engineered for *Agro*infiltration, where TRV cDNA was cloned between the a duplicated 35S CaMV promoter (2x35S) and a Nopaline synthase terminator (NosT). A self-cleaving ribozyme (Rz) was included to release exact 3' end viral RNAs. Viral ORFs are illustrated as boxes. The restriction sites at the multiple cloning sites (MCS) are shown in italic. Promoters are depicted as arrows. LB: Left border. RB: Right border. AmpR. Ampicillin resistance gene. KanR. Kanamycin resistance gene.

1.2.3 Genome editing tools delivered via plant viral vectors

Lately, genome editing tools have been in the public eye due to the advantages of the system to improve desired traits in plants. As stated above, viral vectors are a good option to deliver the components of the different systems into plant cells, particularly if it is possible to cross into the meristem, producing heritable mutations. Moreover, it has been discussed that the delivery of the genome editing tools using RNA viral vectors will be categorized as non-transgenic, since the viral RNA genome does not integrate into the plant genome, a long-term goal in creating genome edited crops.

In 2010, Marton *et al.* (2010) reported the use of TRV to transiently deliver ZFNs. The authors described the systemic infection of TRV and the ability to detect targeted mutagenesis mediated by ZFN in a variety of tissues and cells, even in buds and newly developed infected tissues isolated from *N. tabacum* and *Petunia hybrida* plants. Sequence analysis shown the transmission of the mutations to the next generation plants (125). Furthermore, in 2015 Honing *et al*. (2015) demonstrated the feasibility of introducing mutations into the *DIHYDROFLAVONOL 4- REDUCTASE* gene using a meganuclease delivered by a TRV vector into the purple-flowered ornamental tobacco *Nicotiana alata*. The authors described a visible phenotype in the mutated plants, consisting in a reduction in the purple pigmentation of flower petals. Also, as stated by Marton *et al*. (2010), this study demonstrated inheritance of the mutations by at least two further generations without detecting TRV transcripts on its progeny (126).

In 2015, Ali *et al*. (2015) published three studies indicating the use of TRV as vector to deliver gRNA into *N. benthamiana* plants already stably transformed to overexpress Cas9. The authors reported the possibility of targeting the *PDS* and *PROLIFERATING CELL NUCLEAR ANTIGEN* genes individually, and by co-delivering both targeting gRNAs, finding a persistence of the viral activity for up to 30 days post infection (dpi). Also, as in previous studies, they detected the presence of genomic mutations in seed progeny. In addition, after analysing putative off-target sites, no mutations were detected, demonstrating the specificity of the system. Finally, they described the use of this system to confer resistance against the *Tomato Yellow Leaf Curl*, *Beet Curly Top* and *Merremia mosaic* DNA viruses (127-129).

In recent years, different authors have published their results delivering just the guide RNA using viral vectors into plants stably expressing Cas9: A second generation TMV to target multiples genes in *N. benthamiana* (130), PEBV to deliver one or multiple sgRNAs into *N. benthamiana* and *A. thaliana* plants (131), *barley stripe mosaic virus* (BSMV) to target genes of wheat and maize (132) and the *Beet necrotic yellow vein virus* to introduce mutations into the *PDS* gene of *N. benthamiana* plants (133). Recently the use of TRV has been proposed to deliver single and multiple sgRNAs fused to a *FT* movement sequence, or tRNAs, to promote cell-to-cell movement and access to the shoot apical meristem to generate heritable mutations in transgenic plants expressing Cas9 (97). Furthermore, the use of PVX was described to deliver single or multiple guides, finding heritable genome edits in regenerated shoots from infected leaves or in seeds when the *FT* movement sequence was fused to the sgRNAs (as described previously), but not by PVX itself, since this virus doesn't enter the meristem (134).

Nonetheless, a constant disadvantage of the virus delivery system is the cargo capacity of the vectors. It has been described that the cargo capacity of TRV is limited to 2-3 kb (129), whilst Baulcombe's group indicated that the longest gene successfully expressed by them in PVX vector is GUS (2 kb) (107, 135). However, some recent papers have indicated the successful expression of the full-length *SpCas9* gene (4.2 kb) into virus vectors. Zhang *et al.* (2019) engineered the *Foxtail mosaic virus* (FoMV), a monocot- and dicot-infecting *Potexvirus*, to co-deliver the *Sp*Cas9 and sgRNA separately into two FoMV vectors, reporting successful mutations in the target *PDS* gene of *N. benthamiana* plants (136). Ma *et al*. (2020) engineered the *Sonchus yellow net rhabdovirus* (SYNV), a (-) ssRNA virus, to deliver a cassette of *Sp*Cas9 with tRNA-sgRNA-tRNA, targeting multiple sites simultaneously (137). Gao *et al.* (2019) engineered as a vector a different rhabdovirus, the *Barley yellow striate mosaic virus* (BYSMV) to simultaneous deliver the *Sp*Cas9 and sgRNA, introducing genome edits into the *GFP* gene in *N. benthamiana* 16c constitutively expressing GFP (138). Ariga *et al.* (2020) used PVX to deliver the Cas9 gene and sgRNA, finding genome editing events in both target *TOBAMOVIRUS MULTIPLICATION 1* (*TOM1*) and *PDS* genes when *N. benthamiana* leaves were infiltrated or mechanically inoculated with PVX-Cas9-*TOM1* sgRNA virions from previously infiltrated leaves. Moreover, fully edited

plants were obtained by shoot regeneration from infiltrated leaves, where a higher genome editing frequency was detected compared with its parental tissue (139).

Considering the problems posed by the size limitation, delivering Cas9 as split domains using virus vectors is a possible approach to obtain efficient genome editing in plants in a non-integrative manner. However, it has been reported that in plants viral cross protection occurs to prevent infection with the same virus twice (140, 141). For example, Marton *et al*. (2010) reported that co-inoculation of two TRV2 viruses with different transgenes resulted in less efficient expression of both (125). To avoid this problem, each split portion can be delivered using different virus vectors, eg. PVX and TRV2, that have been shown to infect simultaneously and, due to the *trans* activity of the 16 kDa protein of TRV2, enables them both to infect meristematic cells (108).

This split Cas9 approach has been already tested in plants by Kaya's research group. In 2016 they successfully induced targeted mutagenesis in *N. tabacum* and rice using a codon optimized *Sa*Cas9 protein (142). Subsequently, in a second study they reported splitting of the *Sa*Cas9 and induced targeted mutagenesis in the *PDS* gene in *N. benthamiana*. The group evaluated two split sites, at 430N/431C and 739N/740C, described previously in human cells by Nishimasu *et al.* (2015) (80), and delivered one fragment of the split-*Sa*Cas9 using the *Tomato Mosaic Virus* (ToMV) and the other fragment by *Agrobacterium* infiltration of a plant expression plasmid. The group found that by using both split sites, it has been possible to detect targeted mutagenesis, but site 739N/740C exhibited a higher genome editing efficiency compared to site 430N/431C. Interestingly, when leaves were inoculated with ToMV expressing the full-length *Sa*Cas9, it was not possible to detect targeted mutagenesis, indicating that the full-length protein might be too large to be expressed by the virus (143).

1.3 FLOWERING TIME IN PLANTS

The global population is continually increasing, creating significant challenges in producing sufficient food and reducing food wastage. Traditionally, the use of agrochemicals and the development of high yield crop varieties by selective breeding has solved this problem. However, these methods are slow, time consuming and chemicals are now not well accepted by the consumer. Additionally, climate change adds a new challenge, since drought, extreme heat and pollution greatly affect crop yields. Thus, a major aim of plant biotechnology is to develop new methods to maximise the production of food in a sustainable manner (144).

Classically, breeding techniques involves the cross-pollination between genotypes with traits of interest. Due to the random recombination and undirected mutagenesis, the process requires several generations of backcrossing and selection before obtaining a new elite crop variety, which can take years (145). Crop varieties obtained by conventional breeding present a loss of fitness and genetic diversity, which in cultivated crops is already limited by the genetic bottlenecks due to its domestication (146, 147). In contrast, genetic engineering has been rated as the "fastest developing technology in agriculture" (146). A first approach relied on the transfer of a gene of interest into a crop, conferring it with beneficial agronomical traits. The organisms obtained by these technologies are called "genetically modified organisms" (GMO) and their commercialization is highly regulated to avoid potential harm to the consumer and the environment, ensuring their biosafety. Even though, a major setback is the public concern and acceptance of these new varieties (146, 148). A new alternative has arisen with the development of site-directed nucleases, which allow the precise edit of the crop genome, without introducing foreign DNA. Particularly, the CRISPR/Cas9 system has been widely used in several crops. This topic is further explored by Zhu *et al.* (2020) and Karkute *et al.* (2017), which reviews the current applications of CRISPR/Cas9 in agriculture to increase yield, improve quality, and confer disease and herbicide resistance by silencing undesired traits or introducing gain-of-function mutations (54, 148). Moreover, the CRISPR/Cas9 technology has been combined with conventional breeding methods by the precise target of reproduction-related genes for haploid induction or the generation of male sterile lines (54). Crops edited by SDN are monitored GMO crop legislations. In 2016, a CRISPR-edited mushroom was granted with a non-regulated

status, allowing its cultivation and commercialization in the US (149). However, in countries with a process-oriented regulation, their commercialization is not possible due to the genome manipulation to obtain a new trait (150).

Most food crops are flowering plants, and fruits and seeds constitute a major component of the human diet. For example, cereal crops such as bread wheat, barley, rice and maize are staple crops for human nutrition and the developmental switches from vegetative to reproductive and grain-filling phases is essential for their production (151). The manipulation of flowering has been used since ancient times to domesticate and spread many crop species into new climatic regions. Jung and Müller reviewed the implications of modifying flowering time for crop improvement (152). In cereals, early flowering is desirable in order to prolong the grain filling phase and avoid certain climate conditions that can affect production. On the other hand, in plants used for biofuel production, animal feeding (forage crops) or crops grown for their leaves, delayed flowering is preferable to achieve higher biomass yields. Also, a delayed flowering phenotype is wanted for certain types of trees that grow in cold environments which can damage their flowers and fruits, but in others where the vegetative phase lasts many years an acceleration of flowering would be advantageous. So, understanding the different pathways that regulate this process is crucial for crop improvement.

At the genetic level, flowering is a complex tightly regulated process. The environmental pathways comprise genes that respond to ambient temperature (such as hot weather), light quality, exposure to long periods of cold before flowering (vernalization), and daylength (photoperiod) signals (153). Short day (SD) plants (ie *N. tabacum*) grow mainly in tropical regions and flower when days become shorter during cold seasons to avoid summer hot temperatures. In contrast, long day (LD) plants (ie *A. thaliana*) grow in temperate climates and flowering is regulated by the exposure to cold temperatures and photoperiod, where floral induction is enhanced during long day conditions after winter has passed (151).

Endogenous factors are also involved in the control of flowering, these consist of the autonomous, circadian clock, plant age, gibberellin, and sugar pathways (153). For example, the flowering of day neutral plants (such as maize) is mainly controlled by the autonomous pathway genes (151).

1.3.1 *FLOWERING LOCUS T* as a gene integrator of signalling pathways for flowering

The pathways controlling flowering all converge on just a few genes, called floral integrators, that regulate the floral transition from the vegetative to a reproductive stage. Of these genes, the transcriptional regulator *FLOWERING LOCUS T* (*FT*) plays a central role in the flowering process (153, 154). *FT* mRNA is synthesized and translated in the phloem companion cells of leaf veins which is where perception of photoperiod takes place (155). From there, the FT protein is transported through the phloem sieve elements to the shoot apical meristem (SAM) where floral transition occurs (156, 157). In the cells of the apical meristem, FT binds to the bZIP transcription factor FLOWERING LOCUS D (FD), activating the expression of several genes involved in flowering. The FT-FD interaction activates the expression of a MADS-box transcription factor encoding gene *APETALA1* (*AP1*), a floral meristem identity gene. Moreover, the FT-FD complex also activates the expression of the MADS box transcription factor *SUPPRESSOR OF OVEREXPRESSION OF CONSTANTS 1* (*SOC1*) gene, which interacts with another MADS box transcription factor AGL24 (AGAMOUS-LIKE 24) to promote activation of the transcription of *LEAFY* (*LFY*), another meristem identity gene involved in the initiation of flower development. Furthermore, it has been shown that the interaction between FT and FD also activates the expression of the *SQUAMOSA BINDING PROTEIN LIKE* (*SPL*) transcription factors, which regulate the expression of *SOC1*, *FUL* (*FRUITFULL* or *AGL8*, another MADS box gene), *LFY* and *AP1*. Furthermore, LFY and AP1 also repress negative regulators of *FT*, such as TERMINAL FLOWER1 (TFL1) (154). Finally, the meristem identity genes *AP1* and *LFY* activate the downstream floral meristem identity genes. Figure 6 summarizes the pathways regulating the transition from a vegetative to a floral meristem.

Since FT plays such a fundamental role in the flowering process, modifying its function by alterations to its gene sequence using gene editing approaches would enable us to manipulate the flowering time of plants.

The plant model *N. tabacum*, a member of the *Solanaceae* family, is one of the most extensively cultivated non-food crops. Whilst native to tropical regions, it is now cultivated commercially worldwide in ~120 countries to produce tobacco. The size

of its genome is 4.5 GB and it has evolved from the hybridization of the ancestors *Nicotiana sylvestris* (2n=24, maternal donor, LD plant) and *Nicotiana tomentosiformis* (2n=24, paternal donor, SD plant) (158). Most *N. tabacum* varieties are day-neutral, although the variety Maryland Mammoth flowers only under SD conditions (159).

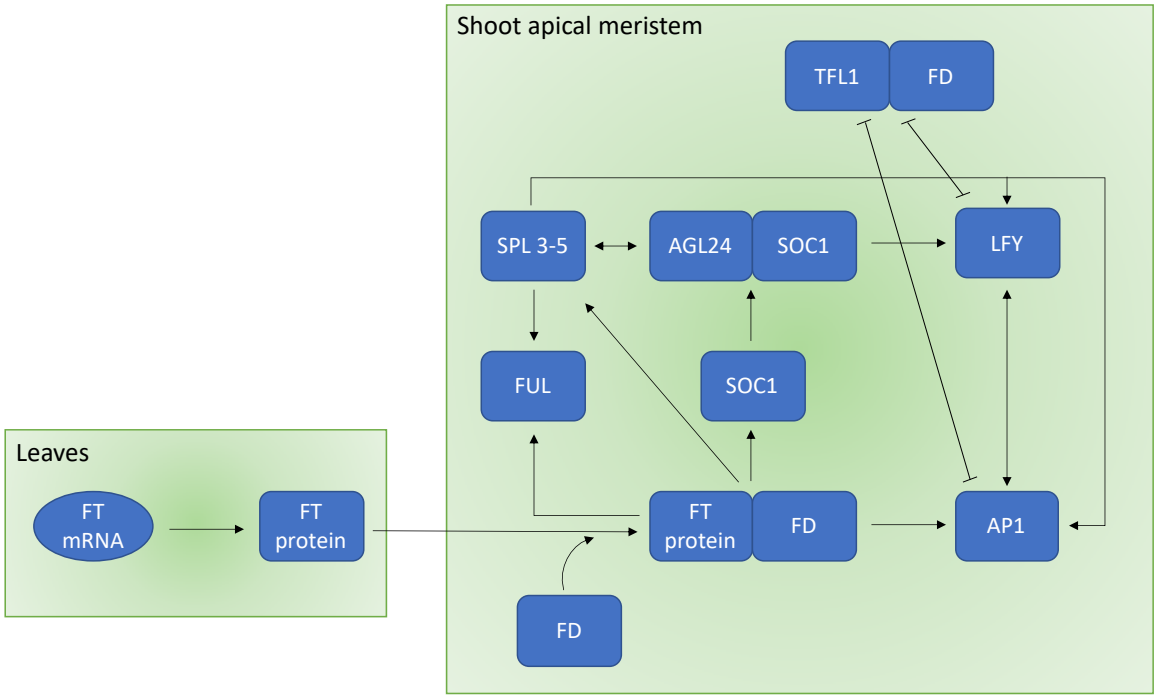


Figure 6. Signalling pathways which induces flowering in plants. FT is synthesized in leaves and moves to the shoot apical meristem, where forms a complex with FD, activating the expression of several genes, such as *SOC1*, *FUL*, *SPL3-5* and *AP1.* These leads to flowering by the floral meristem identity genes *LFY* and *AP1*, which also repress *TFL1*, a negative regulator of *FT.* Arrows indicates interaction, while blunt ends illustrate repression

1.3.2 Flowering regulation in the model plant *Nicotiana tabacum*

In angiosperms, the phosphatidyl ethanolamine-binding protein (PEBP) family comprises several proteins involved in the control of flowering. In the plant model *A. thaliana*, six PEBP family proteins have been described, which are grouped in three main clades: FT-like clade, comprising the FLOWERING LOCUS T (FT) and TWIN SISTER OF FT (TSF) proteins; TFL1-like clade containing the TERMINAL FLOWER 1 (TFL1), BROTHER OF FT AND TFL1 (BFT) and ARABIDOPSIS THALIANA CENTRORADIALIS (ATC) proteins and the MFT-like clade which includes the MOTHER OF FT AND TFL1 (MFT) protein (159, 160). The FT-like and MFT-like proteins act mostly as floral promoters, while TFL-1 like proteins act mostly as floral repressors (159). The origin of these three subfamilies has been proposed to be due to the duplication and divergence of ancestral PEBP genes in plants and coincides with the evolution of the plant kingdom. Additionally, PEBP paralogs have been described among the subfamilies, with different spatiotemporal expression patterns as well as different functions (161).

In *N. tabacum*, Harig *et al.* (2012) have described four *FT*-like genes, each with a different function: *NtFT1*, *NtFT2* and *NtFT3* act as floral inhibitors (atypical function of *FT*) and *NtFT4* is a floral inducer (159). In addition, Beinecke *et al.* (2018) identified more *FT*-like genes, *NtFT5* (a floral inducer, also characterized by Wang *et al.* (2018) (162)) and eight more, named as *NtFT6* to *NtFT13*. For *NtFT8*, *NtFT9*, *NtFT10* and *NtFT13* their function could not be predicted, whereas *NtFT6*, *NtFT7*, *NtFT11* and *NtFT12* were excluded from their analysis (163). Phylogenetic analysis of PEPB family members have shown that *NtFT1-5* genes cluster together in the *FT* clade, separated from the *TFL1* and *MFT* clades (159, 162). A phylogenetic comparison between *NtFT1-12* and *FT* members from different species showed their similarity, however no *TFL1* and *MFT* outgroups were included in this analysis (163).

Conserved regions characterize the members of the PEBP family, such as the DPDxP and GxHR motifs, necessary for anion-binding activity (164). Moreover, specific amino acids distinguish FT protein activity from their counterparts, the floral repressors TFL proteins. Amino acids H88 and D144 form a hydrogen bond required for TFL1 activity, while those positions in FT proteins have a Y85 and a

Q140 (or E in NtFT2 and NtFT3), which don't interact. All NtFT proteins contains these residues. In contrast, the segment B region (LGRxTVYAPGWRQN) is shared between FT floral inducers, such as AtFT, NtFT4 and NtFT5, whilst this sequence is different in the floral repressors NtFT1, NtFT2 and NtFT3. Protein alignments of the segment B of NtFT8, NtFT9, NtFT10 and NtFT13 could not predict their function as floral inducer or repressors (163). Specifically, substitutions of three amino acids on segment B have shown to be able to convert an activator (Y134, G137 and W138) into a repressor (N134, Q137 and Q138), and vice versa. A triad of XYN in the segment C is conserved among FT floral inducers, however its presence on the sugar beet protein BvFT1, a floral repressor, suggest that it isn't required for floral inducing activity (159, 162, 163).

Beinecke *et al.* (2018) reported that FT genes are expressed differently under SD or LD conditions. The authors determined that *NtFT1* and *NtFT4* are predominantly expressed under SD conditions, but a very weak expression of the genes can also be detected in LD conditions. Moreover, *NtFT5* can be detected in both conditions, *NtFT2* is expressed only under SD conditions, whilst *NtFT3* was almost undetectable. As mentioned, the genome of *N. tabacum* is a hybridization of the ancestors *N. sylvestris* and *N. tomentosiformis*. The authors demonstrated that the origin of *NtFT1*, *NtFT4* and *NtFT5* is *N. tomentosiformis, while N. sylvestris is the* origin of *NtFT2 and NtFT3*. Finally, the authors stated that in tobacco, floral-activating and repressor *FT* genes are expressed simultaneously, but in other crops this expression is mutually exclusive. They have described three functional FD proteins, NtFD1, NtFD3 and NtFD4, which interact with the floral repressor *Nt*FT2 and the floral inducers NtFT4 and NtFT5 in a non-selective way. Nonetheless, a bias for the formation of the floral activator complex was observed, which might explain how NtFT4/NtFT5 can surpass the inhibitory effect of the other NtFTs (163).

Recently, Schmidt *et al.* (2019) reported the use of the CRISPR/Cas9 system to generate a *NtFT5* knock-out gene, indicating a loss of the ability of flowering during LD conditions. In contrast, heterozygous *Ntft5*$^{-}$ */NtFT5*$^{+}$ plants show a delay of flowering of around 2 days, which was enough to confer beneficial agronomic traits such as an increase in the vegetative leaf biomass, the production of more seeds and better performance under abiotic stress, such as salinity, drought and heat stress.

However, in this work, knock-out plants were produced through the stable integration of the CRISPR/Cas9 components in transformed tobacco (165).

The encouraging results of targeting the *NtFT5* gene arises the question whether targeting the floral inducer *NtFT4* gene, expressed mostly under SD, would also confer similar beneficial impacts in the plant model *N. tabacum* var Maryland Mammoth. To answer this, the use of an RNA viral non-integrative to deliver the CRISPR/Cas9 system can be tested, including the split Cas9 approach to address the cargo capacity limitation.

1.4 HYPOTHESIS

The use of RNA viral vectors to express Cas9 and sgRNA transcripts in plants will enable non-integrative genome editing to occur in virus infected tissues throughout the plant.

This hypothesis will be tested by constructing virus CRISPR/Cas9 vectors designed to genome edit the floral inducer *NtFT4* gene in the SD flowering plant *Nicotiana tabacum* var Maryland Mammoth to alter flowering time in these plants.

1.5 OBJECTIVES

- Assessment of *in vivo* activity of re-assembled split *Sp*Cas9 fused with inteins and split *Sa*Cas9 domains by genome editing of the *PDS* target gene in *N. tabacum* var. Maryland Mammoth protoplasts.

- Establishment of *Potato virus X* and *Tobacco rattle virus* expression vectors for the delivery of a single transcription unit encoding a full-length or split Cas9 orthologs and ribozyme-sgRNA-ribozyme into *N. tabacum* var. Maryland Mammoth plants.

- Screen for targeted mutations in *NtFT4* gene in leaves of plants inoculated with virus vectors expressing both full-length and re-assembled Cas9 enzymes.

- Shoot regeneration of virus-free genome edited plants from leaves infected with TRV and PVX virus vectors expressing full-length or split Cas9-*NtFT4* sgRNAs cassettes.

CHAPTER 2: MATERIALS AND METHODS

2.1 PLANT MATERIAL

N. tabacum var. Maryland Mammoth and *N. benthamiana* tobacco seeds were grown in F2 + S soil (Levington Advance) and incubated in a plant growth cabinet (Panasonic Biomedical, catalogue #MLR-352H-PE) at 22°C under long day (16 h light and 8 h dark), or short day (8 h light and 16 h dark) conditions. Fluorescent bulbs provide light at an average of 123.93 µmol/m2/s. Plants that were 2-3 weeks old were used for transformation.

2.2 PLASMIDS AND VIRUS VECTORS

A map of all the vectors constructed in this thesis and heterologous sequences are shown in Appendix A1

2.2.1 Expression vectors

To express *Sp*Cas9 full-length, N-*Sp*Cas9 N-intein and C-intein C-*Sp*Cas9 the vectors pBlueScript II SK(+) and pCAMBIA1300 were used, where each construct was cloned between the CaMV 35S promoter and the NosTer.

The vectors to express *Sa*Cas9 full-length, *Sa*Cas9 739N and *Sa*Cas9 740C were kindly gifted by Seiichi Toki (Ryukoku University and NARO, Japan). The Cas9 genes were cloned between the CaMV 35S promoter and the terminator region from the *A. thaliana HEAT SHOCK PROTEIN 18.2* gene (HSP terminator). To enhance the transcription of the genes, the untranslated region of the *A. thaliana ALCOHOL DEHYDROGENASE* gene (*AtADH* 5'-UTR) was positioned upstream of the transgene (142, 143).

To express the sgRNAs, the vectors pBlueScript II SK(+) and a sgRNA expression vector (backbone pUC19, kindly provided by Seiichi Toki) were used. Both vectors contain an RNA Pol III promoter from *A. thaliana U6 SMALL NUCLEOLAR RNA26* (*AtU6-26*) to drive the expression of small RNAs. On each vector, a *Bbs*I restriction site is positioned between the promoter and the Cas9 scaffold RNA to clone the corresponding guide RNA.

2.2.2 Inteins plasmids

The plasmids expressing both inteins were a gift from Hideo Iwai (University of Helsinki, Finland). The vector pHYRSF1 contains the segment for the DnaE N-intein (Addgene plasmid #34549; http://n2t.net/addgene:34549; RRID: Addgene_34549) (166), while the vector pSKBAD2 bears the segment for the DnaE C-intein (Addgene plasmid #15335; http://n2t.net/addgene:15335; RRID: Addgene_15335) (93).

2.2.3 Virus vectors

The virus vectors Potato Virus X (PVX) and Tobacco Rattle Virus (TRV) were used for this thesis work and were provided by Dr. Stephen Jackson. The expression of heterologous genes in PVX is driven by a duplicated viral coat protein sgP upstream of the multiple cloning site. PVX comprises a T7 RNA Polymerase promoter upstream of the *RNA dependent RNA polymerase* gene in order to be able to synthetize viral RNA copies from the cDNA clone by *in-vitro* transcription. This transcribed RNA can be used to infect plants by mechanical rubbing. On the other hand, TRV is a bipartite vector engineered as a binary T-DNA vector for *Agro*infiltration, where both TRV1 and TRV2 viral cDNAs are cloned between the CaMV 35S promoter and a Nos terminator.

A sgP from the PEBV coat protein promoter was cloned downstream of the TRV2 *coat protein* gene to express heterologous genes (109). The promoter was amplified from the P1406 pTRV-2b-GW+ vector, which was a gift from Stuart MacFarlane (The James Hutton Institute, Scotland UK) using the TRV2 promoter F and TRV2 promoter R primers (sequences described in Appendix B1) and cloned between the *EcoR*I (GAATTC) and *Xba*I (TCTAGA) restriction sites of the TRV2 polylinker.

A schematic representation of both vectors is shown in Figure 5

2.3 MOLECULAR BIOLOGY AND CLONING METHODS

2.3.1 Polymerase chain reaction

2.3.1.1 PCR from plasmids for cloning

Every insert for cloning was obtained by PCR, performed using 10 ng of DNA plasmid, 1X of Phusion® HF 5X Buffer, 0.2 mM of each dNTP, 0.5 μM of each primer and 0.02 U/μL of Phusion® High-Fidelity DNA Polymerase (New England Biolabs #M0530). The amplification program was an initial denaturation of 98°C for 30 s, followed by 35 cycles with a denaturation step at 98°C for 10 s, an annealing step at a primer-specific temperature listed on Appendix B for 30 s, an extension at 72°C for 30 s/kb and a final extension at 72°C for 5 min.

2.3.1.2 Colony PCR

To screen for positive colonies, colony PCR was performed using the primers listed in Appendix B1. The PCR mixture consisted of 1X of Green GoTaq® Reaction 5X Buffer, 2.5 mM of $MgCl_2$, 0.25 mM of each dNTP, 0.125 μM of each primer and 0.025 U/μL of GoTaq® G2 Flexi DNA Polymerase in a final volume of 10 μL. The thermocycling program was an initial denaturation at 94°C for 5 min, followed by 35 cycles with a denaturation step at 94°C for 30 s, an annealing step at the temperature listed on Appendix B1 for 30 s, an extension at 72°C for 1 min and a final extension at 72°C for 7 min. Positive clones were sent for sequencing using the appropriate primers.

2.3.2 Visualisation and Purification of PCR products and digested vectors

PCR and other DNA samples were run in a 1% w/v agarose gel prepared with 1x Tris-acetate EDTA (TAE) buffer (40 mM Tris, 20 mM acetic acid, and 1 mM of Ethylenediaminetetraacetic acid (EDTA) and dyed with 1X GelRed® Nucleic Acid Gel Stain (Biotium, catalogue #41003). The DNA was visualized under UV light and pictures were taken using a G:BOX GelDoc (Syngene). The 1 kb plus DNA ladder was used as size comparison (Invitrogen #10787018).

If multiple bands were obtained, the desired band was isolated by excising it from the gel on a UV light box and cleaned up using QIAquick® Gel Extraction Kit (Qiagen, catalogue #28704) following the manufacturer instructions.

If a single band was obtained after the PCR, or after a digestion using restriction enzymes, the DNA was directly purified using the QIAquick® PCR Purification Kit (Qiagen, catalogue #28104) as the manufacturer recommended.

In both cases, the purified DNA was eluted in 30 µL of autoclaved Milli-Q water and the concentration and quality of the obtained DNA was measured using a NanoDrop™ ND-1000 Spectrometer (Thermo Scientific).

2.3.3 Adenylation of PCR products

To clone PCR blunt end products into vectors with a single T 3´-overhang at the insertion site, an adenylation step was performed. A reaction mixture of 20 µL was prepared using 13.3 µL of purified PCR product, 1X of Colorless GoTaq® Reaction 5X Buffer, 2.5 mM of $MgCl_2$, 0.25 mM dATP and 0.05 U/µL of GoTaq® G2 Flexi DNA Polymerase (Promega, catalogue #M7801). Samples were incubated at 72°C for 30 min to allow the addition of an A to the 5' ends of the PCR products and then cloned it into a suitable vector.

2.3.4 Dephosphorylation of vectors

When destination vectors were digested with a single restriction enzyme, a dephosphorylation step was carried out to prevent its recircularization. The reaction was pursued using a Calf Intestinal Alkaline Phosphatase (CIAP, Promega #M182A). The enzyme was diluted to 0.01 U/µL using 1X CIAP 10X Reaction Buffer. Then, 20 µL of purified vector were mixed with 2.5 µL of CIAP 10X Buffer and 2.5 µL of diluted CIAP enzyme. The mixture was incubated at 37°C for 30 min and then another 2.5 µL of diluted CIAP enzyme were added, to continue the incubation for 30 min.

2.3.5 Ligation

The ligation reaction was performed using a ratio of insert to vector of 3:1 calculated using the formula

$$ng\ insert = \frac{kb\ insert}{kb\ vector}\ x\ ng\ of\ vector\ x\ R$$

where "ng" represents the nanograms of insert or vector, "kb" is the size of insert or vector and "R" is the desired ratio.

The mixture was ligated using 1x of Ligase 10X Buffer and 1 μL of T4 DNA Ligase (Promega #M1801) and incubated overnight at room temperature.

2.3.6 Transformation of *E. coli* competent cells

The ligated products were transformed into *E. coli* TOP10 chemo-competent cells. Up to 10 μL of the ligation was mixed with 50 μL of competent cells and placed on ice for 20 min. Then, the cells were transformed by heat shock at 42°C for 45 s and placed immediately after on ice for 2 min. Next, 300 μL of LB media (Tryptone 10 g/L, NaCl 10 g/L, Yeast extract 5 g/L) was added to the cells and they were incubated at 37°C for 1 h with continuous shaking at 200 rpm. Finally, the cells were plated on LB agar plates with the appropriate antibiotic and placed overnight at 37°C. In the case of pGEM®-T Easy vector, 0.1 mM IPTG and 60 mg/μL X-gal was also added for blue/white selection.

2.3.7 Transformation of competent *Agrobacterium tumefaciens* cells

Plasmid expression vectors pCAMBIA1300 and pRI plus the viral vectors TRV1 and TRV2 were transformed into *Agrobacterium tumefaciens* strain GV3101 by electroporation. 50 μL of the competent cells were thawed and mixed with 1 μL of plasmid and placed on ice. The mixture was placed on a 0.2 cm electroporation cuvette and electroporated using the "Agr" program in a Micro Pulser Electroporator (Bio-rad, catalogue #1652100) set at 2.2 kV. Immediately 300 μL of LB media was added to the cuvette and transferred to a 2 mL tube. Cells were recovered at 28°C during 3 h with continuous shaking at 200 rpm. Finally, the mixture was plated on LB agar plates with kanamycin and incubated at 28°C for 2 days.

2.3.8 Miniprep, sequencing and glycerol stocks

Colonies were picked and inoculated into 5 mL of LB media with the appropriate antibiotic and grown overnight at 37°C with continuous shaking at 200 rpm. Next day, the cells were pelleted at 2850xg for 10 min, the supernatant was discarded, and the plasmid was extracted using the QIAprep® Spin Miniprep Kit following the manufacturer instructions.

Plasmids were sequenced at GENEWIZ (South Plainfield, NJ). 10 µL of reaction was set up mixing 2.5 µL of the appropriate primer at a stock concentration of 10 µM, 1 µL of plasmid (30-100 ng/µL) and 6.5 µL of sterile water. The sequences files were analysed using SnapGene Viewer v6.0 software (from Insightful Science; available at snapgene.com).

Cells from positive colonies, and *Agro*-transformed cells, were stored in 15% v/v glycerol, mixing 375 µL of 80% v/v sterile glycerol and 1625 µL of liquid cell culture. Samples were snap-frozen in liquid nitrogen and stored at -80°C.

2.4 DESIGN OF GUIDE RNAs AND CONSTRUCTION OF THE sgRNA

For *Sp*Cas9 two guide sequences were designed to target the *N. tabacum FT4* gene in exon 2 (TobFT1) and exon 4 (TobFT4) and had been cloned into pGEM®-T Easy vector (Promega, catalogue #A1360) prior to this thesis work.

For *Sa*Cas9, three guide sequences were designed to target the *N. tabacum FT4* gene in the same positions previously described for *Sp*Cas9 using the online tool Cas-Designer from the webpage CRISPR RGEN Tools (http://www.rgenome.net/cas-designer/) selecting 5'-NNGRRT-'3 (R=A or G) as a PAM site (167, 168). The target sites of each guide RNA for *Sp*Cas9 and *Sa*Cas9 are illustrated in Figure 7.

Off-target activity of TobFT1 and TobFT4 gRNAs was assessed by sequence alignment using the Clustal Omega algorithm (169). *NtFT*1-13 gene sequences were download from the GenBank database using the published accession numbers (159, 163).

Guides RNA were synthesized as forward and reverse oligonucleotides by Integrated DNA Technologies IDT (Coralville, IA), adding to each end the restriction site overhangs for cloning into the destination vector, and an additional G at the 5'-end necessary for higher efficiency of the *At*U6-26 promoter. Each primer is listed in Appendix B2. Forward and reverse primers were annealed in a 10 µL reaction mixing 10 µM of forward oligo, 10 µM of reverse oligo, 1X of T4 ligation 10X buffer (final concentration of 30 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM of Dithiothreitol (DTT) and 1 mM ATP) (Promega, catalogue #C126) and 1 µL of T4 polynucleotide kinase (Promega, catalogue #M4101). Samples were incubated at

37°C for 30 min then at 95°C for 5 min and then allowed to cool down on the bench until reach room temperature for correct annealing.

To create the single guide RNA (sgRNA), the annealed guides were fused to their respective scaffold by cloning into the *Bbs*I site (New England Biolabs, catalogue #R0539) of their respective sgRNA expression vector described previously in section 2.2.1.

As positive control of Cas9 gene editing activity, a guide RNA to target exon 5 of the *N. tabacum PDS* gene with *Sp*Cas9 and *Sa*Cas9 were synthesized. These sequences were described by Kaya *et al.* (2016) (142) and cloned as described above to create the sgRNA.

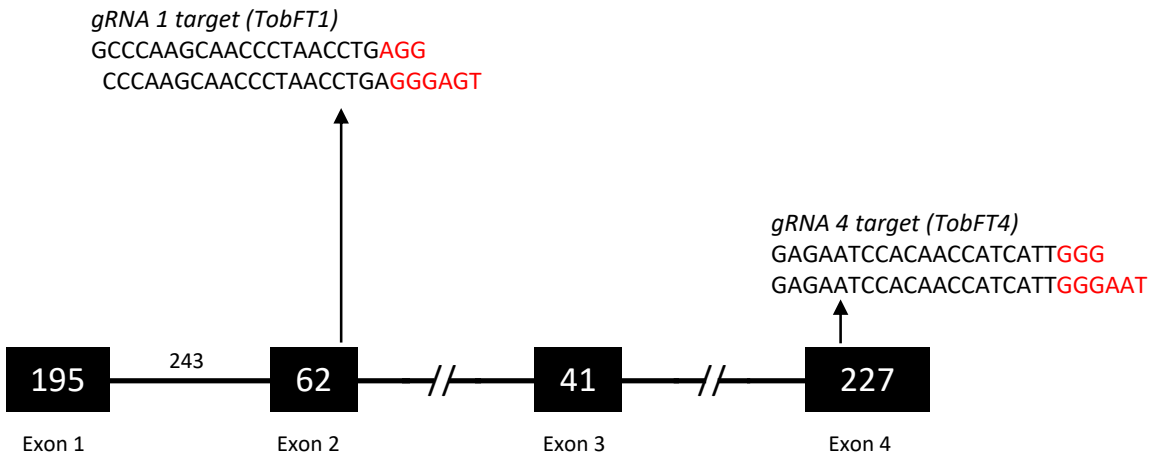


Figure 7. Design of two sgRNA targeting *N. tabacum FT4* gene. Black letters indicated the guide sequences for *Sp*Cas9 (20 bp) and *Sa*Cas9 (21 bp). The PAM sites for *Sp*Cas9 (NGG) and *Sa*Cas9 (NNGRRT) are shown in red.

2.5 CONSTRUCTION OF THE SINGLE TRANSCRIPT UNIT (STU)

2.5.1 Fusion of split *Sp*Cas9 ends to inteins by overlapping extension PCR

An overlapping extension PCR was performed to fuse each *Sp*Cas9 terminal to the respective intein (170). Basically, this protocol consisted of a first-round amplification of each sequence using primers with the overlapping segment, followed with a second-round amplification to fuse the fragments (Figure 8).

The split N-terminal and C-terminal *Sp*Cas9 domains were amplified from the full-length gene of 4272 bp, which comprised at the 5' end a 3xFLAG tag and a SV40 NLS and at the 3' end a nucleoplasmin NLS. The full-length protein was split at the position (Glu573)/(Cys574), as described elsewhere (82, 83). The DnaE N-intein was amplified from the vector pHYRSF1, while DnaE C-intein was amplified from the vector pSKBAD2, described in section 2.2.2. Each PCR was performed following the conditions described in section 2.3.1.1 and primers listed in Appendix B3.

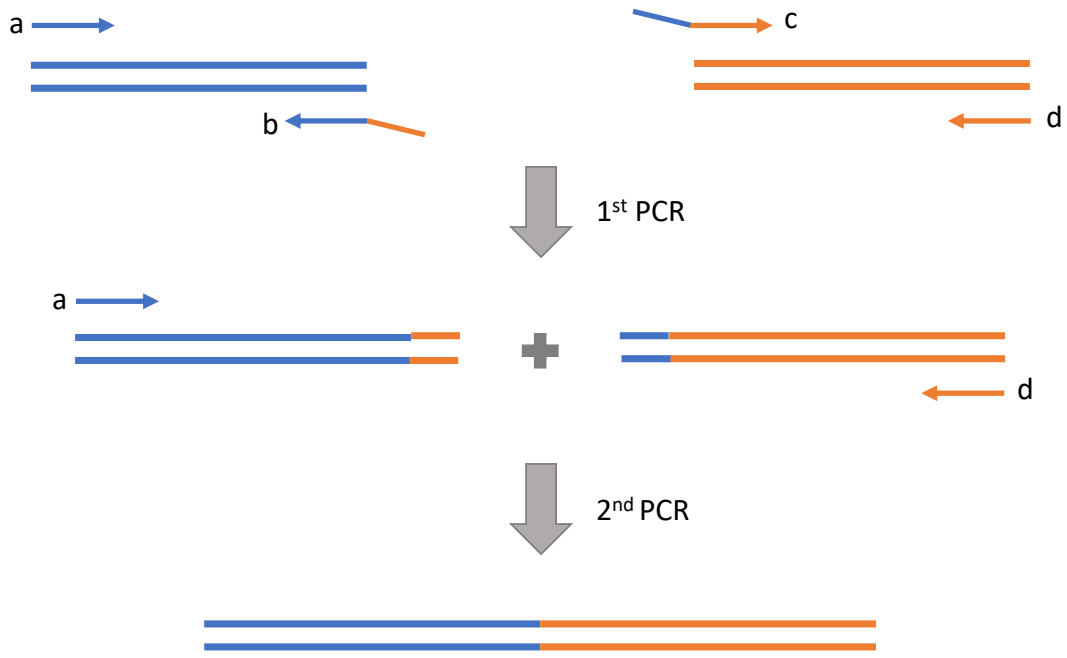


Figure 8. Overlapping extension PCR (OE-PCR) protocol. Each sequence to fuse is amplified independently with primers with an overlap segment (b and c in the figure). In a second PCR, both PCR fragments are fused using the external primers "a" and "d".

A second round OE-PCR (see Figure 8) was carried out where the DNA templates to fuse were mixed in a 1:1 ratio. For the fusion of N-*Sp*Cas9 with N-intein the forward primer was N-intein Cas9 F1 and the reverse primer was N-intein Cas9 R. For the synthesis of C-intein C-*Sp*Cas9 the forward primer was C-intein Cas9 F1 and reverse primer was C-intein Cas9 R. The sequence of these primers is listed in Appendix B3. The PCR mixture and amplification program were the same as mentioned in section 2.3.1.1 using 20 ng of each template. Afterwards, the PCR products were purified and cloned into pGEM®-T Easy vector.

2.5.2 Construction of the expression vectors pCAMBIA1300 *Sp*Cas9 full-length, N-*Sp*Cas9 N-intein and C-intein C-*Sp*Cas9

Full-length *Sp*Cas9 was subcloned from the pBluescript II SK(+) 2x CaMV 35S - *Sp*Cas9 -NosT plasmid into pCAMBIA1300 plant expression vector using the *Hind*III – *EcoR*I sites (Promega #6041 and #6011, respectively).

The N-*Sp*Cas9 N- intein and C-intein C-*Sp*Cas9 sequences were amplified from their respective pGEM®-T Easy vectors using the PCR primers listed in Appendix B3, where the *Xho*I (CTCGAG) restriction sites added are shown in bold. The PCR conditions used are described in section 2.3.1.1. Afterwards, these PCR products were digested using *Xho*I (Promega #R6161) and cloned between the CaMV 35S promoter and the CaMV poly(A) signal of the pCAMBIA1300 vector, replacing the *HYGROMYCIN B* resistance gene. Positive clones were screened using internal primers for each end of the gene (Appendix B1) as described in section 2.3.1.2 and sent for sequencing.

2.5.3 Construction of the Ribozyme-sgRNA-Ribozyme (RGR) unit

For each sgRNA designed to target the *FT4* gene, hammerhead ribozymes (HHRz) were attached at each end to generate the ribozyme-sgRNA-ribozyme (RGR) unit (Figure 9).

The first six nucleotides of the 5'- hammerhead ribozyme (shown in green in Figure 9) are reverse complementary to the first six nucleotides of each sgRNA (orange) to generate the hammerhead structure (98). At the 3'-end of the guide RNA scaffold was added a hammerhead ribozyme (green) from the tobacco ringspot virus satellite RNA (99).

To construct the RGR units, two consecutives PCR were set-up. A first PCR was performed to add the first half of the sequence of each hammerhead, following the PCR conditions described in section 2.3.1.1. Subsequently, a second round PCR was carried out to add the rest of the hammerhead sequences using 2 µL of the previous PCR mixture and the same PCR conditions.

The primers used to create the RGR unit are listed in Appendix B4 and B5. The six complementary nucleotides necessary for the formation of the hammerhead structure are shown in italics. The restriction sites for *Sal*I (GTCGAC), *Mlu*I (ACGCGT), *Xma*I (CCCGGG) and *Xho*I (GTCGAG) added for downstream cloning are in bold. The additional nucleotides added to create the hammerhead sequences in the RGR unit are underlined.

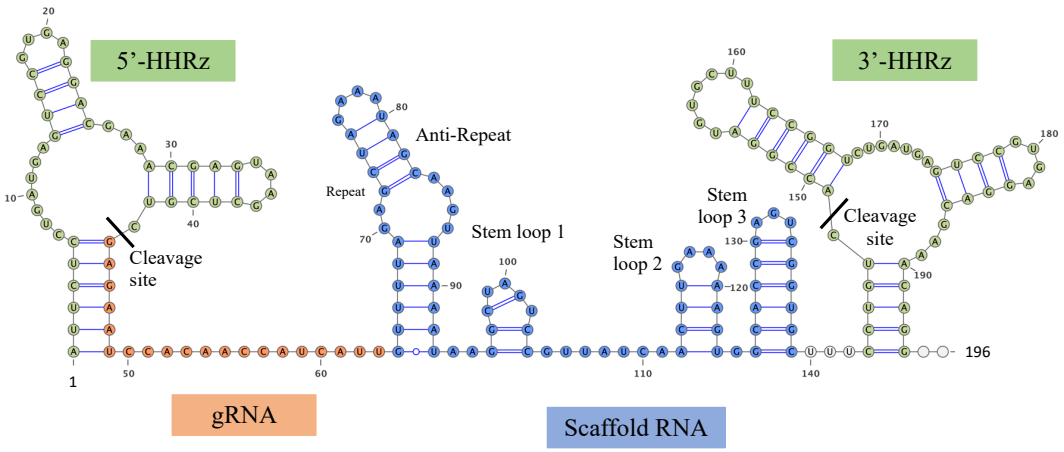


Figure 9. Secondary structure of the ribozyme-sgRNA-ribozyme (RGR) unit. Green nucleotides denote each hammerhead ribozyme with their cleavage site indicated. Orange nucleotides denote the guide RNA, while the blue nucleotides are the guide RNA scaffold or tracrRNA. *Sa*Cas9 RGRs follow the same structure, except that the guide RNA is 21 bp and the scaffold exhibits two stem loops. The figure was created using the VARNA v3.9 software (171). HHRz: Hammerhead Ribozymes.

2.5.4 Cloning of the Cas9-RGR single transcript unit into virus vectors

The Single Transcript Unit (STU) was created by the fusion of the *Cas9* gene with the Ribozyme-sgRNA-Ribozyme in the virus vectors PVX and TRV2

2.5.4.1 Cloning of *Sp*Cas9 constructs fused to sgRNAs flanked by hammerhead ribozymes

*Sp*Cas9 full-length was cloned into TRV2 at the *Mlu*I site of the vector before this thesis work. sgRNAs flanked by HHRz were amplified from pGEM®-T Easy vector using the primers listed on Appendix B4, which have at each end an *Xma*I (CCCGGG) restriction site for cloning downstream of *Sp*Cas9.

To clone into PVX, the single transcript unit of *Sp*Cas9 full-length and sgRNAs flanked by HHRz were amplified from TRV2 using the primers listed on Appendix B6 and cloned between the *Cla*I (ATCGAT) and *Sal*I (GTCGAC) restriction sites of the vector.

N-*Sp*Cas9 N-intein and C-intein C-*Sp*Cas9 were cloned into the polylinker of pGEM®-T Easy vector as described in section 2.5.1. On the other hand, the RGR constructs were amplified from the sgRNA expression vector (backbone pBlueScript II SK(+)) using two rounds of consecutive PCRs, as described in section 2.5.3 with the primers listed in Appendix B4 and cloned downstream of each fused intein-*Sp*Cas9 terminal at the *Sal*I (GTCGAC) and *Mlu*I (ACGCGT) restriction sites of the pGEM®-T Easy vector.

For cloning into PVX vector, each single transcript unit was digested from the pGEM®-T Easy vector using the restriction enzymes *Cla*I and *Mlu*I, while for cloning into TRV2 vector the enzymes used were *Kpn*I and *Mlu*I. Positive colonies were screened as described in section 2.3.1.2 using as a forward an internal primer pairing in each *Sp*Cas9 construct and as a reverse a PP82 R or TRV2 R seq (Appendix B1), accordingly.

2.5.4.2 Cloning of *Sa*Cas9 fused to sgRNAs flanked by hammerhead ribozymes

To clone into TRV2, *Sa*Cas9 full-length, *Sa*Cas9 739N and *Sa*Cas9 740C constructs were amplified from the pRI expression vectors mentioned in section 2.2.1 using the amplification conditions described in section 2.3.1.1. The primers for this reaction are listed in Appendix B7, where the restriction sites *Kpn*I (GGTACC) and *Sac*I (GAGCTC) were added for downstream cloning.

The TRV2 vector and PCR products of each *Sa*Cas9 construct were digested using *Kpn*I and *Sac*I restriction enzymes and ligated as described in section 2.3.5. Screening of positive colonies was carried out as described in section 2.3.1.2, using a forward internal primer for each *Sa*Cas9 domain and the TRV2 R seq reverse primer (Appendix B1).

RGR constructs were amplified from the sgRNA expression vector depicted in section 2.2.1 using two rounds of consecutive PCRs, as described in section 2.5.3 with the primers listed in Appendix B5. For cloning into TRV2 vector, the PCR products were digested using *Mlu*I and *Xho*I. Positive colonies were screened using as a forward an internal primer pairing in each *Sa*Cas9 construct and as a reverse TRV2 R seq (Appendix B1).

To clone the *Sa*Cas9-RGR STU into PVX viral vector, two approaches were followed. First, TRV2 *Sa*Cas9 740C with each RGR unit were used as template using the PCR conditions detailed in section 2.3.1.1 and the primers listed on Appendix B8. Then, the obtained PCR product was cloned between *Cla*I and *Sal*I restriction sites of PVX. Secondly, *Sa*Cas9 full-length with each RGR and *Sa*Cas9 739N with each RGR were cloned into PVX using the NEBuilder® HiFi DNA Assembly kit (New Englands Biolads, catalogue # E5520S). Each insert was amplified using the primers listed on Appendix B8, using the TRV2 constructs as templates and the PCR conditions described in section 2.3.1.1, while PVX was digested using the enzymes *Eag*I and *Sal*I. The insert and vector were mixed in 10 µL following a molar ratio of 2:1 pmol and then 10 µL of NEBuilder® HiFi DNA Assembly Master Mix was added. The reaction was incubated at 50°C for 15 min and then 2 µL of the reaction was transformed into TOP10 chemo-competent cells.

In both cases, positive colonies were screened using the par of primers PP82 that bind outside the multiple cloning site of PVX to amplify the full insert (Appendix B1).

The maps of all the constructed vectors are shown in Appendix A1

2.6 *N. tabacum* PROTOPLASTS ISOLATION AND TRANSFORMATION

Tobacco protoplasts were isolated from young leaves of 3-week-old plants. The leaves were cut in strips diagonally to the main vein and digested in 20 mL of enzyme solution (10 mM of 2-(N-morpholino) ethanesulfonic acid (MES) pH 5.7, 10 mM $CaCl_2$, 20 mg/mL Cellulase R10, 10 mg/mL Macerozyme, 2 mg/mL of Bovine serum albumin (BSA), 0.6 M mannitol) at room temperature for 5 h or overnight with continuous shaking. Next day, the enzyme solution was passed through a 100 µM cell strainer sitting in a 50 mL falcon tube and protoplasts were pelleted at 100xg for 5 min at room temperature. Protoplasts were washed using 25 mL of chilled W5 solution (2 mM MES pH 5.7, 154 mM NaCl, 125 mM $CaCl_2$ and 5 mM KCl) followed by a centrifugation at 100xg for 5 min to collect them. This step was repeated twice and afterwards the protoplasts were resuspended in 5 mL of chilled W5 solution and incubated on ice for 30 min. Next, the protoplast solution was centrifugated at 100xg for 5 min, the pellet was resuspended in 1300 µL of MMG solution (4 mM MES pH 5.7, 0.4 M mannitol and 15 mM $MgCl_2$) and the quantity of protoplasts was calculated using a haemocytometer.

For transformation, 30 µg of total DNA was mixed with 300 µL of MMG protoplast suspension and 330 µL of PEG4000 solution (40% w/v PEG4000, 0.2 M mannitol, 100 mM $CaCl_2$) and incubated for 15 min at room temperature. As control of transformation, protoplasts were transformed with pGY-1 mCherry vector (172). Subsequently, 5 mL of W5 solution was slowly added and protoplasts were collected at 100xg for 3 min. This step was repeated twice. Finally, protoplasts were resuspended in 3.5 mL of W5 solution and cultured in darkness in 12- well plates pre-treated with 1% BSA solution. mCherry fluorescence was visualized using an inverted fluorescence microscope (Model IX70, Olympus USA) after 24 and 48 h. Then the protoplasts were pelleted at 17000xg for 1 min, supernatant was discarded, and they were frozen for further experiments.

2.7 PLANT INOCULATIONS

Three independent plants (biological replicates) were infected per virus construct. Three independent plants were inoculated with water only, corresponding to mock controls (WT).

2.7.1 Inoculation by *Agro*infiltration of plant expression vectors and TRV viral vectors

Plasmid expression vectors pCAMBIA1300 and pRI, plus the viral vectors TRV1 and TRV2, were grown from *Agrobacterium tumefaciens* glycerol stocks in LB broth with 50 µg/mL of kanamycin, 50 µg/mL of rifampicin and 25 µg/mL of gentamicin for 48 h. Cells were pelleted at 2850xg for 10 min and the supernatant was poured off. Cells were resuspended in 1 mL of Induction buffer (10 nM $MgCl_2$, 10 nM MES and 150 µM acetosyringone), OD_{600} was adjusted to 1.0 using the same buffer and samples were kept in darkness for at least 3 h. Afterwards, using a needleless 1 mL syringe *Agrobacterium* was infiltrated into the underside part of the leaves. For further analysis using the plasmid expression vectors, samples from the infiltrated leaves were collected after 3 days. On the other hand, when the TRV viral vector was used, samples from infiltrated leaves were collected after 7 days if *in vitro* regeneration of plants was followed or after 14 days for systemic infection analysis.

2.7.2 Viral infection using PVX vector

PVX vectors were linearized using *Spe*I or *Sac*I restriction enzymes (New England Biolabs, catalogue #R0133S and #R3156S respectively) in a reaction mixture consisting of 1 µg of plasmid DNA, 10 U (1 µL) of *Spe*I or *Sac*I enzyme, 1X of CutSmart® 10X Buffer and 0.1 mg/mL of acetylated BSA in 20 µL reaction and incubated at 37°C for 1 h. Next, the linearized plasmid was cleaned up using the QIAquick® PCR Purification Kit as described in section 2.3.2. *In vitro* transcription reactions were set up using the HiScribe™ T7 High Yield RNA Synthesis Kit (New England Biolabs, catalogue #E2040S), adding 1X of Reaction 10X Buffer, 2 mM of ATP, UTP and CTP, 0.2 mM of GTP, 0.5 mM of m7G(5')ppp(5')G RNA Cap Structure Analog (New England Biolabs, catalogue #S1404S) and 30 µL of linearized PVX plasmid. The mixture was incubated at 37°C for 5 min and then 4 µL

of T7 RNA Polymerase Mix was added, followed by incubation at 37°C for 25 min. Finally, 1 µL of GTP 100mM was added and a final incubation at 37°C for 35 min was performed.

In vitro transcribed mRNA was cleaned up adding RNAse free sterile water up to 100 µL followed by 100 µL of phenol:chloroform:isoamyl alcohol (IAA) (25:24:1), mixture was vortexed for 30 s and centrifugated at 17000xg for 3 min. The upper phase was transferred to a new tube, 100 µL of chloroform was added, vortex for 30 s and centrifuged at 17000xg for 3 min. Next, 100 µL of the upper phase were taken out and 0.1 of the volume (10 µL) of sodium acetate 3M pH 5.2 and 2.5 of the volume (250 µL) of absolute ethanol was added and incubated at -80°C for at least one hour. Then, samples were centrifuged at 17000xg for 18 min, the supernatant was poured off and the pellet was washed with 100 µL of Ethanol 70%, followed by a centrifugation at 17000xg for 5 min. Afterwards, the supernatant was discarded, the pellet was air dried and finally dissolved in 120 µL of RNAse free sterile water.

As an alternative, the RNA was cleaned-up by precipitation with LiCl. RNA in a total volume of 90 µL was mixed with 30 µL of 8 M LiCl for molecular biology to a final concentration of 2 M (Sigma, catalogue #L7026). The mixture was incubated at -20°C for at least 30 min and then centrifugated at 16000xg for 15 min at 4°C. Afterwards, the supernatant was discarded, and the pellet was washed with 500 µL of Ethanol 70% v/v followed by a centrifugation step at 16000xg for 10 min at 4°C. Then, the supernatant was discarded, the pellet was air dried and finally resuspended in 120 µL of RNAse free sterile water.

For transformation of *N. benthamiana* and *N. tabacum*, 20 µL of purified *in vitro* transcribed PVX vector was rubbed on top of the leaf dusted with carborundum powder as abrasive. After 7 to 10 days, symptoms of infection started to appear, such as chlorosis, mottles and rugose mosaic (113). Samples for further experiments were taken 14 dpi from both infected and systemic leaves.

2.8 SHOOT REGENERATION FROM VIRUS INFECTED LEAVES

Infected leaves with PVX or TRV vectors were collected 7 to 10 dpi. Leaves were cut into 1 cm squares, washed with 30 mL of 1% w/v Sodium hypochlorite (NaOCl) for 30 min and then washed three times with sterilized water. Afterwards, each

square was placed on non-selective shoot induction media, consisting of 4.4 g/L Murashige-Skoog basal salts, 20 g/L sucrose, 0.2 mg/L of 1-Naphthaleneacetic acid (NAA), 1 mg/L of 6-Benzylaminopurine (BAP) or kinetin, 250 mg/L of carbenicillin and 4 g/L of Gellan Gum (Alfa Aesar, catalogue #J63423), adjusting the pH to 5.7-5.8. Plates were kept in a plant growth cabinet (Panasonic Biomedical, catalogue #MLR-352H-PE) at 22°C under long day conditions (16 h light and 8 h dark). Fluorescent bulbs provide light at an average of 123.93 μmol/m2/s. After approximately a month, regenerated shoots were moved to non-selective regeneration media for root development (2.2 g/L Murashige-Skoog basal salts, 15 g/L sucrose, vitamin mix of 1 mg/L thiamine, 0.5 mg/L nicotinic acid and 0.5 mg/L pyridoxine and 4.4 g/L of Gellan Gum, adjusting the pH to 5.7-5.8), replacing the media every week until further analyses were pursued (139).

2.9 RNA ANALYSIS

2.9.1 RNA extraction

2.9.1.1 RNA extraction using TRIzol™ reagent

Frozen tissue samples (100 mg) were ground up using a DREMEL® workstation drill in liquid nitrogen until a fine powder was obtained and then 1 mL of TRIzol™ (Fisher Scientific, catalogue #15596026) was added to extract total RNA. The samples were incubated at room temperature for 5 min and then 200 μL of chloroform was added. Samples were incubated at room temperature for 3 min, then centrifuged at 16000xg for 15 min at 4°C. Afterwards, the aqueous phase was transferred to a new tube, 500 μL of isopropanol was added and the mixture was incubated at room temperature for 10 min. Nucleic acids were pelleted by centrifugation at 16000xg for 10 min at 4°C. The supernatant was removed, and the pellet was washed with 1 mL of Ethanol 75% v/v, centrifuged at 7500xg for 5 min at 4°C and air dried. Finally, the pellet was resuspended in 40 μL of RNAse free water and treated with Amplification Grade DNase I (Sigma, catalogue number AMPD1), mixing 5 μL of 10X Reaction Buffer and 5 μL of DNase I 1 unit/ μL, Amplification Grade. Samples were incubated at room temperature for 15 min, then 5 μL of Stop Solution (provided with the DNaseI) was added and then samples were incubated at 70°C for 10 min to stop the reaction.

2.9.1.2 RNA extraction using Monarch® Total RNA Miniprep Kit

Frozen tissue samples (100 mg) were ground using a DREMEL® workstation drill in liquid nitrogen until a fine powder was obtained and then 800 µL of 1X DNA/RNA Protection Reagent was added to the sample. Then, samples were centrifuged at 16000xg for 2 min, the supernatant was transferred to a new 2 mL tube, an equal volume of RNA Lysis Buffer was added, and samples were briefly vortexed. Next, 800 µL of the mixture was loaded into the gDNA removal column and samples were centrifugated at 16000xg for 30 s. The flow-through was saved, as it contains the RNA partitions. The step was repeated to collect the complete volume and then equal volume of absolute ethanol (1600 µL) was added. Afterwards, the mixture was transferred to the RNA purification column and centrifugated each time at 16000xg for 30 s until all the volume was passed through. Next, samples were treated with DNase I on the column, adding 500 µL of RNA Wash Buffer, centrifuged at 16000xg for 30 s and then a mixture of 5 µL DNase I with 75 µL DNase I Reaction Buffer. The reaction was incubated at room temperature for 15 min and then washed with 500 µL RNA Priming Buffer. Samples were centrifuged at 16000xg for 30 s, washed twice with 500 µL RNA Wash Buffer and then the RNA was eluted using 50 µL of RNAse free water.

2.9.2 Assessment of the quality of the RNA and storage

The RNA concentration was measured using a NanoDrop™ ND-1000 Spectrometer. The quality of the RNA was evaluated by the absorbance ratio 260 nm and 280 nm ratio, considering a ratio of 2.0 as "pure". Finally, samples were stored at -20°C for short time storage or -80°C for long time storage.

2.9.3 cDNA synthesis and Reverse Transcription PCR (RT-PCR)

The cDNA was synthesized using 4 µL of extracted RNA and 1 µL of random primers (Promega, catalogue #C1181), incubated at 70°C for 5 min and immediately placed on ice for at least 5 min. Next, a final volume of 20 µL was completed adding 1X GoScript™ 5X Reaction Buffer, 2.5 mM of MgCl2, 0.5 mM of each dNTP and 1 µL of GoScript™ Reverse Transcriptase (Promega, catalogue #A5003). Samples were annealed at 25°C for 5 min, followed by an extension step at 42°C for an hour and a final inactivation of the enzyme at 70°C for 15 min.

The RT-PCR was carried out in a final volume of 20 µL using 2 µL fresh synthesized cDNA, 1X of Green GoTaq® Reaction 5X Buffer, 2.5 mM of $MgCl_2$, 0.25 mM of each dNTP, 0.125 µM of each primer and 0.025 U/µL of GoTaq® G2 Flexi DNA Polymerase. The PCR protocol consisted in an initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, a specific annealing temperature (listed in Appendix B9) for 30 s, 72°C for 1 min and a final extension of 72°C for 7 min. As control of RNA quality the constitutively expressed housekeeping gene *ELONGATION FACTOR 1-ALPHA* (*EF1α*) was used. Finally, the PCR products were visualized in an 1% w/v agarose gel as described in section 2.3.2.

2.9.4 Assessment of the self-cleavage activity of the RGR unit by circular RT-PCR (cRT-PCR)

To check the self-cleavage activity of the ribozymes from the sgRNA, a circular RT-PCR was performed to check the flanking regions of the guide. A schematic representation of this procedure is shown in Figure 10.

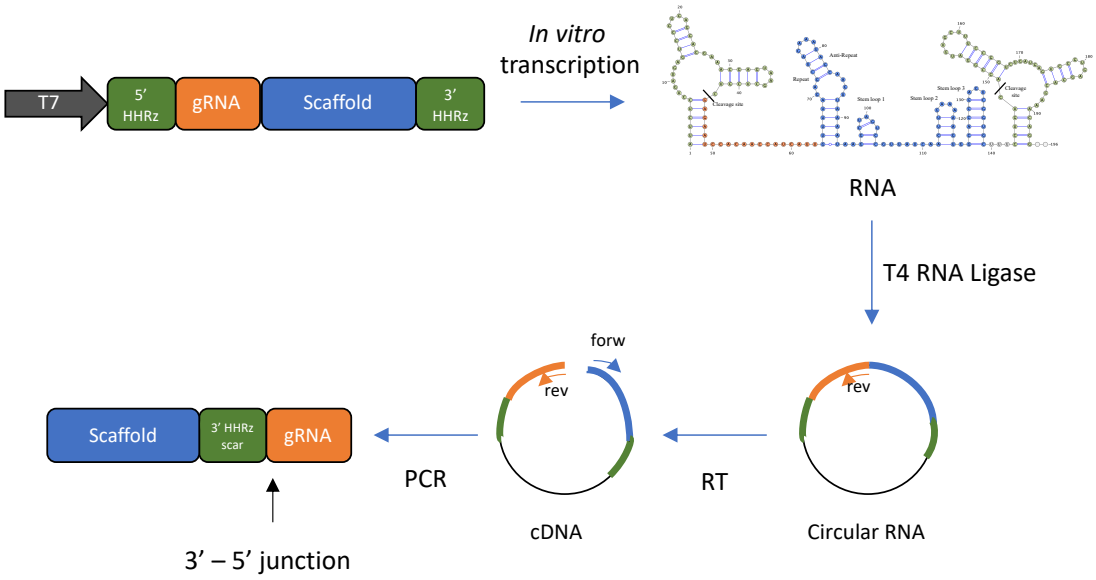


Figure 10. Schematic representation of the cRT-PCR analysis. RNA is synthesized by *in vitro* transcription using the T7 RNA Pol promoter. The obtained RNA is circularized by a T4 RNA ligase and used as template for cDNA synthesis primed by the gRNA reverse primer. RT-PCR reactions are carried out using as forward a primer binding to the RNA scaffold sequence and as reverse the same primer used in the previous step. Next the PCR product is cloned, and individual colonies are screened to assess the HHRz self-cleavage activity. HHRz: Hammerhead ribozymes.

As a test, PVX C-intein C-*Sp*Cas9 TobFT1 HH and PVX C-intein C-*Sp*Cas9 TobFT4 HH were linearized with *Spe*I and RNA was synthetized by *in vitro* transcription as described in section 2.7.2.

Next, the RNA was treated with DNase I and cleaned-up with TRIzol™ reagent, as described in section 2.9.1.1. For samples from leaf material, RNA was extracted using TRIzol™ reagent, as described in section 2.9.1.1.

The circularization of the RNA was carried out in a 20 µL reaction comprising 10 µL of RNA (1 µg), 1X of T4 RNA ligase 10X buffer, 50 µM of ATP, 10% w/v PEG8000, 20 U of RNasin (Promega, catalogue #N2111) and 10 U of T4 RNA ligase (New England Biolabs, catalogue M0204). The reaction was incubated at 25°C for 4 h and then inactivated at 65°C for 15 min. Afterwards, the obtained circular RNA was clean-up by completing the volume to 100 µL with sterile water and then precipitated with 0.2 volumes of sodium acetate 3M pH 5.2 and 2.5 volumes of absolute ethanol. Samples were incubated at -20°C overnight, followed by centrifugation at 17000xg for 15 min at 4°C to pellet the RNA. The supernatant was discarded, the pellet was washed with 500 µL Ethanol 70% v/v, followed by a second centrifugation at 17000xg for 15 min at 4°C and then air dried. Finally, the pellet was resuspended in 30 µL of RNAse free water.

Subsequently, the circular RNA was used as a template for RT-PCR. For the cDNA synthesis, the reaction mixture comprises 4 µL of cRNA and 0.5 µM of the gRNA spacer reverse primer and the protocol was as described in section 2.9.3. The RT-PCR was carried out in a final volume of 20 µL using 2 µL fresh synthesized cDNA, 1X of Q5® Reaction 5X Buffer, 0.2 mM of each dNTP, 0.5 µM of each abutting primers binding to the gRNA and RNA scaffold and 0.02 U/µL of Q5® High-Fidelity DNA Polymerase. The PCR protocol consisted of an initial denaturation at 98°C for 30 s, followed by 35 cycles of 98°C for 10 s, a specific annealing temperature for 30 s, 72°C for 30s and a final extension of 72°C for 2 min The sequence of the primers and the annealing temperatures is listed in Appendix B10. Finally, the cRT-PCR products were analysed in a 2% w/v agarose gel, cloned into pGEM®-T Easy vector and Sanger sequenced individual colonies.

2.10 PROTEIN ANALYSIS

2.10.1 Protein extraction

Leaves disks of 1.5 cm were frozen in liquid nitrogen and homogenised in 100 µL of total protein extraction buffer (0.1% v/v of NonidetP-40 (NP-40), 150 mM Tris-HCl pH 7.5, 5 mM EDTA pH 8, 150 mM NaCl, 10% v/v glycerol, 1 mM of Phenylmethylsulfonyl fluoride (PMSF), 1x Protease cocktail inhibitors and 5 mM DTT) using a DREMEL® workstation drill. Next, samples were centrifuged at 17000xg for 30 min at 4°C and supernatant was collected and stored at -20°C for short term storage, or -80°C for long term storage. If required, a Bradford Assay (Sigma, catalogue #B6916) was performed to quantify the concentration of the total protein extract using BSA 100 µg/mL as standard. For this assay, 0.5 µL of protein crude extract was mixed with 19.5 µL of water in triplicate and 200 µL of Bradford dye was added. Then, the absorbance was measured in a 96-well microplate reader (TECAN) at 595 nm using Magellan™ software v3.12.

2.10.2 Western blot analysis

Total protein extracts were separated by SDS-PAGE using 7.5% w/v acrylamide/ bis-acrylamide gels with 1% w/v of sodium dodecyl sulfate (SDS). Samples were incubated in 1x Laemmli 6x buffer (173) and 4.2 mM DTT at 95°C for 5 min and loaded on the gel, which was run in 1x running buffer (3 g/L Tris-HCl, 14.4 g/L Glycine, 1 g/L SDS) at 150V at 4°C. The Color Prestained Protein Ladder Broad Range (11–245 kDa) (New England Biolabs, catalogue #P7712S) or Color Prestained Protein Ladder Broad Range (10–250 kDa) (New England Biolabs, catalogue #P7719S) were used as a size standard. Proteins were transferred from the gel onto a 0.45 µm nitrocellulose membrane (Amersham™ Hybond™ ECL™, GE Healthcare catalogue #RPN303D) in 1x transfer buffer (3 g/L Tris-HCl, 14.4 g/L Glycine, 1 g/L SDS and 20% v/v methanol) for 1.5 h at 100V at 4°C. The membrane was then incubated with blocking solution (1x of Tris-buffered saline tween buffer (TBS-T) and 5% w/v skimmed milk powder) for 1 hour at room temperature on an orbital shaker. For immunodetection of both Cas9 proteins, the membrane was incubated with the primary antibody anti-FLAG® M2 1:1000 (Sigma-Aldrich, catalogue #F1804). Alternatively, for *Sp*Cas9 the Anti-CRISPR/Cas9 monoclonal antibody (Sigma, catalogue #SAB4200701) was used at a dilution of 1:1000 or

1:500. To detect the *Sa*Cas9 protein, the monoclonal antibodies for the C-terminal (Sigma, catalogue #MAC141) and for the N-terminal (Sigma, catalogue #MAC142) were used at a concentration of 1:1000. The incubation of all primary antibodies was carried out overnight at 4°C on an orbital shaker. Next day, the membrane was washed 3 times for 5 min each with 1x TBS-T and then incubated for one hour at room temperature on an orbital shaker with the secondary antibody Anti-Mouse IgG (whole molecule)–Peroxidase antibody produced in goat (Sigma, catalogue #A4416) diluted to 1:4000. Finally, the membrane was washed 3 times for 5 min each with 1x TBS-T and developed using the Amersham™ ECL™ Prime Western Blotting Detection Reagent (Cytiva, catalogue #RPN2232) for chemiluminescence detection using CL-XPosure™ Film (Thermo Scientific, catalogue #34089) and Xray film processor machine (Konica Minolta, catalogue SRX-101A).

2.11 CRISPR/Cas9-MEDIATED MUTATION ANALYSIS

2.11.1 DNA extraction from transformed protoplasts

DNA was extracted from protoplast samples using the Monarch® Genomic DNA Purification Kit (New England Biolabs, catalogue #T3010). Briefly, transformed protoplasts were resuspended in 30 µL of EB Buffer with 1 µL of Proteinase K and 3 µL of RNase A and mixed properly by vortexing. Then, 100 µL of Cell Lysis Buffer was added and the solution was incubated for 5 min at 56°C. Next, 400 µL of gDNA Binding Buffer was added, the mixture was transferred to a gDNA Purification Column and centrifuged first at 1000xg for 3 min to bind the gDNA, and then at 17000xg for 1 min. Afterwards, the column was washed twice with 500 µL of gDNA Wash Buffer, centrifuged at 17000xg for 1 min and finally the gDNA was eluted with 35 µL of sterile Milli-Q water.

2.11.2 DNA extraction from infected plants and regenerated shoots with virus vectors

When leaf material was analysed, DNA was extracted using an extraction method with cetyltrimethylammonium bromide (CTAB) based on Lodhi *et al.* (1994) (174). Samples (100 mg) were frozen in liquid nitrogen and homogenised in 800 µL of DNA extraction buffer (2% w/v CTAB, Tris-HCl 100 mM pH 8, NaCl 1.4 M, EDTA 20 mM pH 8 and 1% v/v of β-mercaptoethanol) and incubated at 60°C for 25 min.

Then, 600 µL of 24:1 chloroform: IAA was added, and samples were centrifugated at 3500xg for 15 min. 450 µL were recovered from the upper (aqueous) phase and half the volume (225 µL) of NaCl 5M plus twice of the volume (900 µL) of absolute ethanol was added to precipitate the DNA. Samples were incubated overnight at 4°C. Next day, two consecutive centrifugations were carried out at 870xg and 3500xg for 3 min each, the supernatant was discarded, and the pellet was washed with 500 µL of Ethanol 76% v/v. Samples were centrifugated at 3500xg for 5 min, supernatant was poured off and the pellet was air dry. Finally, DNA pellet was resuspended in 100 µL of sterile Milli-Q. water with 1 µL of RNase (final concentration of 0.1 mg/mL) at 37°C.

2.11.3 Assessment of genome editing by PCR analysis

DNA edited regions were amplified using the primers listed in Appendix B11. PCR was performed using 5 µL of protoplast DNA or 2.5 µL of leaf gDNA, 1X of Q5® Reaction 5X Buffer, 0.2 mM of each dNTP, 0.5 µM of each primer and 0.02 U/µL of Q5® High-Fidelity DNA Polymerase. The thermocycling conditions were an initial denaturation at 98°C for 30 s, 35 cycles of 98°C for 10 s, primer-specific annealing temperature as listed in Appendix B11 for 30 s and 72°C for 30 s and a final extension of 72°C for 2 min. The PCR products were run on a 1% agarose gel with 1X GelRed® Nucleic Acid Gel Stain and purified using QIAquick PCR Purification Kit, as described in section 2.3.2. Afterwards, the genome editing events were assessed by colony PCR and Sanger sequencing or Amplicon-EZ analysis.

2.11.3.1 Evaluation of genome editing by colony PCR

The obtained PCR products were adenylated, cloned into pGEM®-T Easy vector, and transformed into TOP10 chemo-competent cells, as described previously in section 2.3.6. Next, colony PCR was carried out using the M13 primers as described in section 2.3.1.2. As a primary screen, positive colony PCR products were cleaned up using the QIAquick® PCR Purification Kit and sent for sequencing using the M13 forward primer. The plasmids for clones showing mutations were purified using the QIAprep® Spin Miniprep Kit and checked by sequencing using the M13 forward primer.

2.11.3.2 Evaluation of genome editing by Amplicon-EZ NGS

Amplicon-EZ analysis enables a mixture of PCR products to be screened for CRISPR/Cas9 genome edits and it was performed by GENEWIZ (South Plainfield, NJ). For this analysis, PCR products must be between 150 – 500 bp, with a concentration of 20 ng/µL, a purity by ratio of 260nm/280nm of 1.8-2.0 and a total amount of 500 ng. The purity of the obtained PCR products was assessed using a NanoDrop™ ND-1000 Spectrometer, while the concentration of the samples was measured using Qubit™ dsDNA HS Assay Kit (Thermo Scientific, catalogue #Q32851), following the manufacturer's instructions. When the required amount of DNA was obtained, samples were submitted for sequencing. After submitting the samples, the company prepares the library, performs the sequencing, and sends a report with the results.

2.11.3.3 Assessment of genome editing by Cleaved Amplified Polymorphic Sequences (CAPS) assay

Genome editing was evaluated in regenerated shoots by CAPS assay. The *NtFT4* exon 2 target region was amplified using the primers listed in Appendix B11. The PCR was performed using 1.5 µL of gDNA extracted from regenerated shoots, 1X of Green GoTaq® Reaction 5X Buffer, 2.5 mM of $MgCl_2$, 0.25 mM of each dNTP, 0.125 µM of each primer and 0.025 U/µL of GoTaq® G2 Flexi DNA Polymerase in a final volume of 20 µL. For difficult templates, 1 mg/mL of BSA was added to the reaction mix to enhance PCR amplification. The thermocycling program was an initial denaturation at 94°C for 5 min, followed by 35 cycles with a denaturation step at 94°C for 30 s, an annealing step at a temperature according to primers listed on Appendix B11 for 30 s, an extension at 72°C for 1 min and a final extension at 72°C for 7 min. Once the reaction finished, 5 U (0.5 µL) of *Dde*I enzyme (Promega, catalogue #R6291) was directly added to the mixture and samples were incubated at 37°C for 1 h. Finally digested PCR products were run on a 1.5% w/v agarose gel with 1X GelRed® Nucleic Acid Gel Stain.

CHAPTER 3: ACTIVITY ASSESSMENT OF FULL-LENGTH AND RE-ASSEMBLED SPLIT Cas9 DOMAINS AND RGR SELF-CLEAVAGE

3.1 INTRODUCTION

3.1.1 Split Cas9 approach

The split Cas9 approach reduces the size of the protein enabling it to be delivered by the viral vector system. Wright *et al*. (2015) and Nishimasu *et al.* (2015) split the *Sp*Cas9 and *Sa*Cas9 proteins respectively at flexible linkers of the proteins and the dimerization of the halves was through the sgRNA (80, 81).

To improve the specificity of re-assembly of the protein domains, the use of protein introns, also known as inteins, has been reported (83-87). Inteins self-excise from a longer protein and enable re-association of the flanking regions (exteins) with a covalent peptide bond formed between them in the process.

Different studies applied this approach to re-associate both domains of split Cas9 proteins using inteins from different sources. Truong *et al*. (2015) described the use of the Npu DnaE split intein integrated in the catalytic subunit of DNA Polymerase III DnaE from *Nostoc punctiforme*. This study reported that by splitting the *Sp*Cas9 at sites Glu573/Cys574 and Lys637/Thr638 and fusing it with the corresponding intein, the enzyme is rapidly reconstituted with an activity comparable to the full-length *Sp*Cas9 protein. Moreover, the intein-splitCas9 was efficiently packaged and delivered by an rAAV virus vector (83). Fine *et al*. (2015) used the Mxe GyrA intein from *Mycobacterium xenopi*, but their results showed a lower efficiency compared with Truong's study (84). A third report, from Chew *et al*. (2016) described the fusion of each split *Sp*Cas9 domain to inteins from *Rhodothermus marinus* and their viral delivery by rAAV in postnatal mice (85). Ma *et al.* (2016) divided the Cas9 protein into three portions and reconstituted them in a circuit like-manner (86). Davis *et al.* (2015) used a ligand binding intein to modulate the activation of full-length Cas9 (87).

Therefore, the use of a split Cas9 approach fused with inteins will reduce the size of the endonuclease to be delivered by virus vectors, and specific re-assembly in the plants cells to exert its editing activity.

3.1.2 Assessment of *in vivo* activity of the CRISPR/Cas9 system by protoplasts transformation

Protoplasts are widely use for the assessment of genome editing activity in plants. Carlson (1973) described some advantages of their use, as for example the lack of cell wall allows the transformation of exogenous components into the cell, the possibility to isolate large amounts of them, and their ability to regenerate into fully grown plants (175). Specifically for genome editing, the use of protoplasts allows a rapid way of evaluating mutations introduced by the CRISPR/Cas9 system, since isolation and transformation of protoplasts can be performed in less than a week. Additionally, a direct delivery of the genome editing reagents is possible since no biological vectors, such as *Agrobacterium*, are necessary. Further, protoplasts can be exploited to produce several independent events, generating non-GMO plants by transient expression of the CRISPR/Cas9 components or by the transformation of protoplasts with ribonucleoprotein complexes containing the Cas9 protein and sgRNA, which don't integrate into the host genome. Lastly, transformed protoplasts can be detected by microscopy using a reporter gene and are manageable for cell sorting (176-178).

Numerous studies have been published using protoplasts to evaluate genome editing activity. Shan *et al.* (2013) and Li *et al.* (2013) were the firsts groups to test the targeted editing of plant genes using PEG-transformed protoplasts (33, 34). Shan *et. al.* (2013) targeted two sites of the rice *PDS* gene using a plant codon optimized *Sp*Cas9. Efficient targeted mutagenesis was detected starting at 18 h of cultivation and similar or higher frequencies of genome edits were found 24 h and 72 h after. Furthermore, the authors also tested the targeting of three more rice genes plus one wheat gene in protoplasts, reporting genome editing frequencies between 26.5% - 38% calculated by band intensity in PCR/RE assays (33). Li *et al.* (2013) also used a plant codon optimized *Sp*Cas9 to target the *A. thaliana PDS3* and *FLS2* (*FLAGELLIN SENSITIVE 2*) genes with mutagenesis frequencies of 5.6% and 1.1%, respectively, based on the number of mutated sequences among randomly selected amplicons. Then, the authors tested the targeted genome editing of the *N. benthamiana PDS* gene at two different sites, obtaining higher frequencies of mutation than in *A. thaliana* (37.7% and 38.5%).

The possibility of multiplex genome editing in *A. thaliana* protoplasts was tested. Successful mutations in two members of the *RECEPTOR FOR ACTIVATED C KINASE 1* (*RACK1*) family were found with similar frequencies (2.5 – 2.7%). When tandem gRNAs were delivered aiming for two juxtaposed targets in *AtPDS3* gene, a deletion of up to 48 bp with a frequency of 7.7% was observed (34). Shan *et al.* (2013) and Li *et al.* (2013) reported the targeted introduction of an intended mutation by gene replacement using the HDR pathway. In both cases, a restriction site was introduced, Shan *et al.* (2013) reported that two out of 29 colonies had the expected insertion into the rice *PDS* gene (33), while Li *et al.* (2013) reported a frequency of insertion of 10.7% into the *N. benthamiana PDS* gene (34). Lin *et al.* (2018) reported the improvement of protoplast isolation and transformation for several plant species. For example, longitudinal cutting was used to isolate protoplasts from rice, wheat, maize, millet and bamboo, while a "tape sandwich" was used for *A. thaliana* and many *Brassicaceae* species, such as *B. oleracea*, *B. napus*, *Cleome spinosa*, *C. monophilla*, and *C. gynadra*. For tomato the authors isolated protoplasts from a suspension cell line from callus derived from hypocotyls. Furthermore, the authors edited the *PDS* gene of several of these species and developed a system to detect mutations in single *N. tabacum* protoplasts by two consecutive PCRs, digestion of the product and regeneration of these edited protoplasts (179). Hsu *et al.* (2021) transformed *N. benthamiana* protoplasts with different CRISPR/Cas9 systems, such as *Sp*Cas9, *Sa*Cas9, *Francisella novicida* Cas12a and a cytosine base editor, introducing mutations into the *PDS*, *Ethylene Receptor 1* (*ETR1*), *RNA-Dependent RNA Polymerase 6* (*RDR6*), and *Suppressor of Gene Silencing 3* (*SGS3*) genes. The group successfully regenerated full plantlets from the edited protoplasts and developed a method to grow these regenerants to flower and produce seed *in vitro* (180).

Moreover, the delivery of preassembled Cas9/sgRNA ribonucleoprotein (RNP) complexes into protoplasts of *A. thaliana*, *N. attenuata* and rice to introduce targeted genome edits has been described (181). Additionally, protoplasts from lettuce (181) and bread wheat (182) were transfected with RNP complexes and genome edited cells were regenerated into fully-grown plants, as an effort to produce a DNA-free genome editing approach.

In conclusion, the evaluation of genome editing activity in protoplasts has become a standard procedure in a variety of crops. The use of protoplasts allows more control over the amount of template delivered for higher precision and efficiency (176). Moreover, the ability to regenerate them into fully-grown plants offers the possibility to obtain non-GMO plants, especially using non-integrative approaches, such as RNP complexes.

3.1.3 Self-cleavage ribozymes to deliver the sgRNA

Ribozymes are RNAs with catalytic activity. They were first described in 1982 and are divided in two main groups: large catalytic RNA (RNase P, group I and II introns) and the small catalytic RNA (hammerheads, hairpins, hepatitis delta and Varkud satellite RNA). They are considered metalloenzymes, since they require a divalent cation for activity, usually Mg^{2+} (183).

Traditionally, sgRNAs are produced *in vivo* using the RNA Pol III promoters from the *U6* and *U3 snRNA* genes, expressed constitutively and ubiquitously. Therefore, they cannot be used for spatiotemporal or specific cell/tissue genome editing (100). Also, in some non-model organisms RNA Pol III promoters have not been well identified and heterologous RNA Pol III promoters often work poorly (100, 101). As an alternative, ribozymes have been used to deliver sgRNA under an RNA Pol II promoter, producing several guides from the same transcript, without using multiple Pol III (100) or as part of a single transcript unit (STU) with the *Cas9* gene for a coordinate expression of Cas9 and the sgRNA (101). Gao *et al.* (2014) were the first to describe the use of ribozymes to deliver a mature guide and they called this construct Ribozyme-gRNA-Ribozyme (RGR) unit. A hammerhead type ribozyme was attached to the 5'-end and the hepatitis delta virus ribozyme was attached to the 3'-end of a guide targeting the *GFP* gene. This was produced by *in vitro* transcription from an SP6 promoter and successfully caused editing when they were incubated with the Cas9 protein and a PCR fragment containing the *GFP* target sequence (98). He *et al.* (2017) reported that RNA Pol II and RNA Pol III promoters can drive the expression of two tandem RGR units in rice with equal efficiencies, expanding the choices of promoters for gRNA transcription (100). Tang *et al.* (2016) expressed the *Sp*Cas9 and the sgRNA as a polycistronic unit separated by a hammerhead ribozyme under an RNA Pol II promoter. Using this system, two sites

in the rice *PDS* gene were targeted, obtaining genome editing frequencies of 38% and 29% assessed by band intensity in PCR/RE assays. Moreover, circular RT-PCR analysis show the *in vivo* self-cleavage activity of the HHRz indicating the correct separation of the sgRNA and the *Cas9* mRNA during transcription. Furthermore, genome edits were also generated in two additional rice genes and the dicots *A. thaliana* and tobacco, indicating that the STU system can be used in different species. The authors also studied the co-expression of two tandem RGRs from a single transcript unit to target two sites of the *PDS* gene in rice, detecting targeted mutations, while targeted chromosomal deletions and inversions were also recovered (101).

Hence, the use of ribozymes allows the expression of a single transcript unit of *Cas9* gene and guide RNA under the same RNA Pol II promoter, ensuring the delivery of the CRISPR/Cas9 system into the same cell to maximise the genome editing efficiency.

In this chapter, the expression and *in vivo* activity of re-assembled split Cas9 orthologs was tested. Targeted genome editing of exon 2 and exon 4 of the *NtFT4* gene was assessed using two guides (TobFT1 and TobFT4) in *N. tabacum* protoplasts. Potential off-targets in different *NtFTs* genes were investigated by *in silico* analysis. Finally, the *in vitro* self-cleavage activity of HHRz to produce mature *NtFT4* sgRNAs from a RGR unit was evaluated.

3.2 RESULTS

3.2.1 Split Cas9 proteins can re-assemble in plant cells and introduce targeted gene edits

Full-length *Sp*Cas9 was split at the amino acid positions Glu573/Cys574, as described by Zetsche *et al.* (2015) (82) and Truong *et al.* (2015) (83). Inteins from the catalytic subunit of DNA Pol II DnaE from *Nostoc punctiforme* were attached to each split Cas9 fragment. Then, the full-length and split *Sp*Cas9-intein split domains were cloned into the pCAMBIA1300 plant expression vector, under control of the CaMV 35S promoter. Split *Sa*Cas9 constructs (split at the positions Glu739/Ser740) were a gift from S. Toki, where the full-length *Sa*Cas9, and each of the split domains had been cloned into the plant expression vector pRI201N (143).

Figure 11 shows a representation of both *Sp*Cas9 and *Sa*Cas9 constructs. The maps of the final vectors and the transgene sequences are presented in Appendix A1.

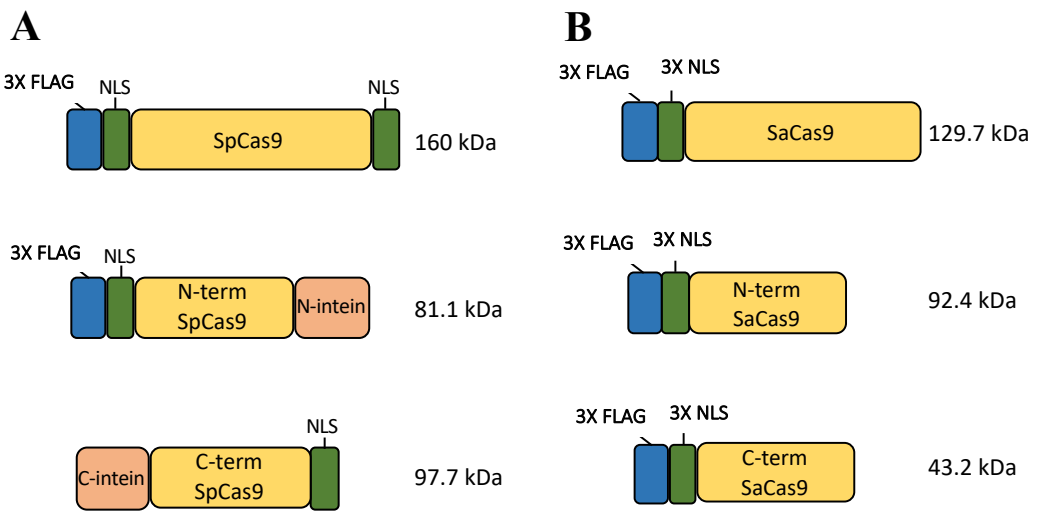


Figure 11. Representation of the different *Sp*Cas9 and *Sa*Cas9 constructs.

A. Representation of the full-length *Sp*Cas9, N-*Sp*Cas9 N-intein, and C-intein C-*Sp*Cas9 constructs. Full-length *Sp*Cas9 is a 160 kDa protein, with an N-terminal 3x FLAG-tag and two Nuclear Localization signals (NLS) at each end. N-*Sp*Cas9 N-intein construct is an 81.1 kDa protein with an N-terminal 3x FLAG-tag and an NLS signal terminating at Glu573, where an N-intein was fused at its C-terminus. C-intein C-*Sp*Cas9 is a protein of 97.7 kDa, starting from Cys574, which has a C-intein added at its N-terminal and an NLS at its C-terminus. **B. Representation of the full-length *Sa*Cas9, *Sa*Cas9 739N, and *Sa*Cas9 740C constructs**. The constructs at their N-terminal have a 3x FLAG-tag and three NLS signals. Full-length *Sa*Cas9 is a protein of 129.7 kDa. *Sa*Cas9 739N is the N-terminal portion of the protein terminating at Asn739 with a size of 92.4 kDa. *Sa*Cas9 740C is the C-terminal portion of the protein, starting at Cys740 with a size of 43.2 kDa.

Following cloning of the full-length, and split domains, of *Sp*Cas9 and *Sa*Cas9 in plant expression vectors, they were *Agro*infiltrated into *N. benthamiana* leaves. It has been reported that the co-delivery by *Agro*infiltration of the *tomato bushy stunt virus* p19 protein improves the expression of heterologous genes in plants, so this approach was tested (136, 184, 185). After 3 days post-infiltration, samples were harvested for Cas9 detection.

Figure 12 shows the detection by immunoblotting of the full-length *Sp*Cas9 (160 kDa) and N-*Sp*Cas9 N-intein (81.1 kDa) proteins in *N. benthamiana* leaf extracts using an anti-FLAG® M2 antibody (panel A), or a monoclonal antibody against the N-terminal region of *Sp*Cas9 (panel B), at different exposure times. The C-intein C-*Sp*Cas9 was not detected since this portion of the protein is not tagged (as shown in Figure 11, panel A) and is not recognised by the antibody for *Sp*Cas9. Coomassie blue stained membranes showing total transferred proteins are displayed in Appendix A2.

In general, the co-infiltration of p19 improved the expression of the transgenes as demonstrated by a brighter band, for example in lane 10 of Figure 12A, which corresponds to the co-expression of both split *Sp*Cas9 domains with p19, compared to lane 9 without p19. If the re-assembly of *Sp*Cas9 by the inteins is successful, full-length protein should be detected even under denaturing conditions due to the formation of a peptide bond between each split domain. When N-*Sp*Cas9 N-intein, C-intein C-*Sp*Cas9 and p19 were co-infiltrated into *N. benthamiana* leaves, it was possible to detect a band at the expected size of the full-length *Sp*Cas9 using both anti-FLAG® M2 and anti CRISPR/Cas9 antibodies (shown by an arrow in Figure 12 A and B), indicating the re-association of the protein

A non-specific band at 100 kDa is observed when anti-FLAG® M2 antibody is used for immunoblotting, but not when the Western Blot was carried out using the anti CRISPR/Cas9 antibody, indicating the specificity of this antibody against *Sp*Cas9 protein.

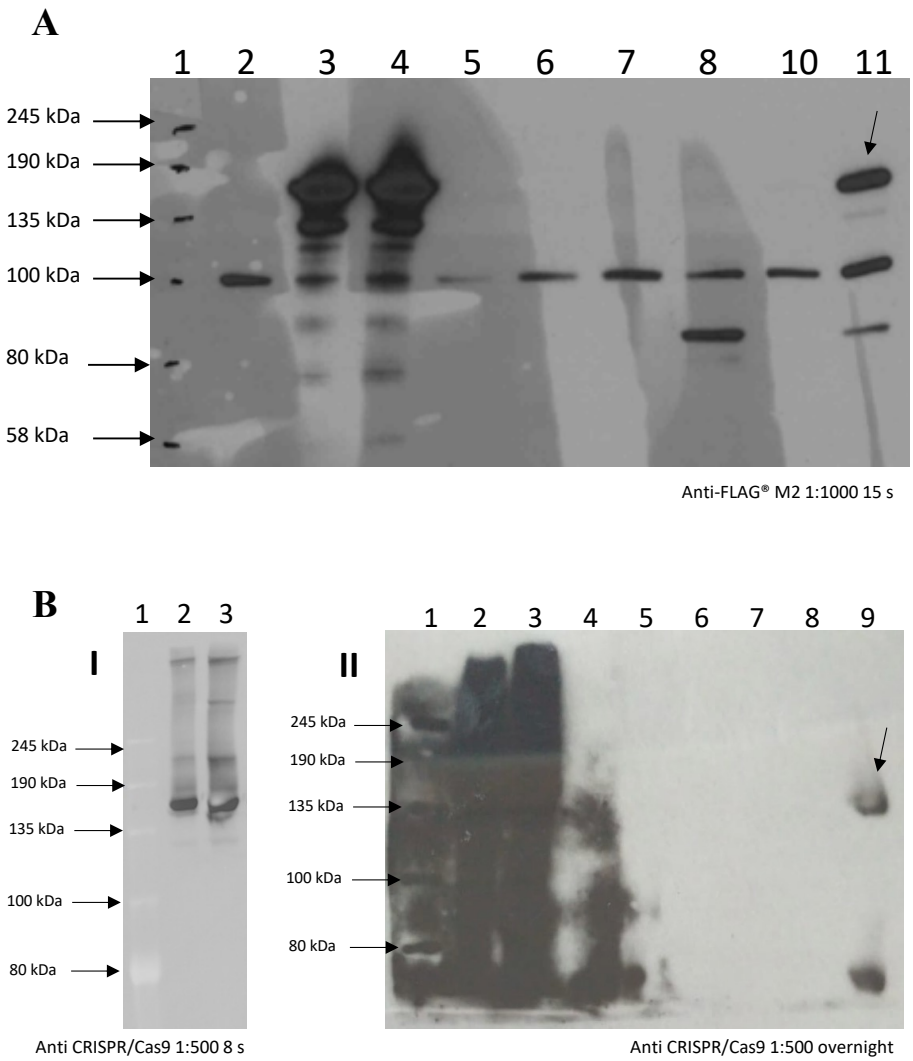


Figure 12. Expression of *Sp*Cas9 constructs in *N. benthamiana* leaves.

A. Immunoblot using anti-FLAG® M2 antibody 1:1000, exposure time 15 s. Lane 1. Protein Standard Ladder (NEB #P7712). Lane 2. Mock. Lane 3. Full-length *Sp*Cas9. Lane 4. Full-length *Sp*Cas9 + p19. Lane 5. C-intein C-*Sp*Cas9. Lane 6. C-intein C-*Sp*Cas9 + p19. Lane 7. N-*Sp*Cas9 N-intein. Lane 8. N-*Sp*Cas9 N-intein + p19. Lane 9. N-*Sp*Cas9 N-intein + C-intein C-*Sp*Cas9. Lane 10. N-*Sp*Cas9 N-intein + C-intein C-*Sp*Cas9 + p19.

B. Immunoblot using anti CRISPR/Cas9 antibody 1:500, exposure time of (i) 8 s, and (ii) overnight. Lane 1. Protein Standard Ladder (NEB #P7712). Lane 2. Full-length *Sp*Cas9. Lane 3. Full-length *Sp*Cas9 + p19. Lane 4. N-*Sp*Cas9 N-intein. Lane 5. N-*Sp*Cas9 N-intein + p19. Lane 6. C-intein C-*Sp*Cas9. Lane 7. C-intein C-*Sp*Cas9 + p19. Lane 8. N-*Sp*Cas9 N-intein + C-intein C-*Sp*Cas9. Lane 9. N-*Sp*Cas9 N-intein + C-intein C-*Sp*Cas9 + p19. Black arrows indicate fully re-assembled *Sp*Cas9.

On the other hand, Figure 13 shows the detection by immunoblotting of the full-length *Sa*Cas9 (129.7 kDa) protein, and the split *Sa*Cas9 739N (92.4 kDa) and *Sa*Cas9 740C (43.2 kDa) domains, using an anti-FLAG® M2 antibody. Samples were collected 3 days post-infiltration for *Sa*Cas9 protein detection. As with *Sp*Cas9, the co-infiltration of p19 improved the expression of the transgenes, exhibiting a brighter band, as *Sa*Cas9 full-length protein levels in lane 3 (+p19) and lane 7 (-p19). Coomassie blue stained membrane showing total transferred proteins is shown in Appendix A2

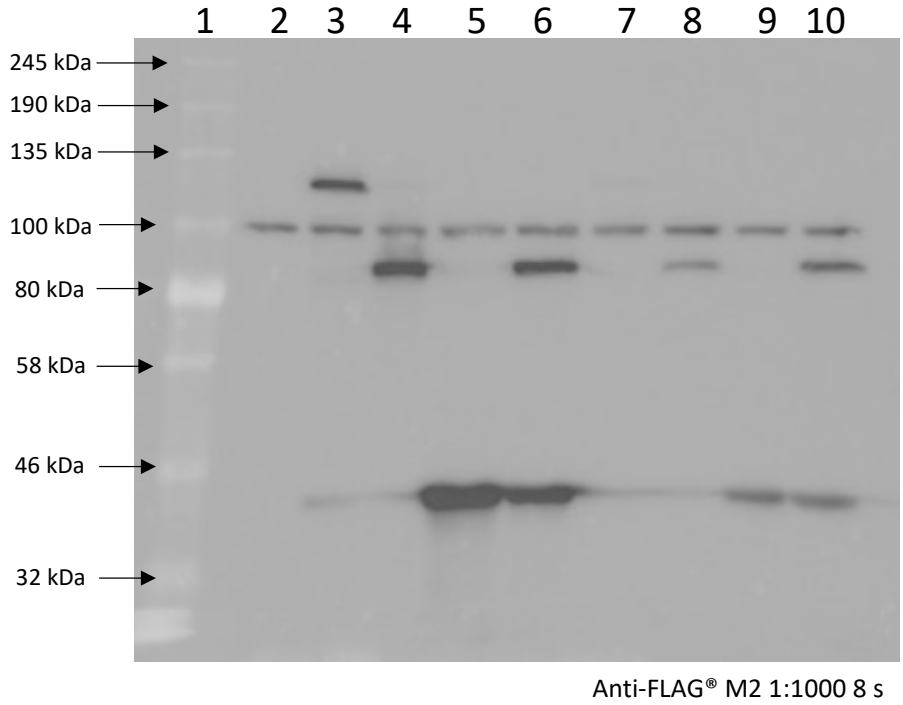


Figure 13. Expression of *Sa*Cas9 constructs in *N. benthamiana* leaves. Lane 1. Protein Standard Ladder (NEB #P7712). Lane 2. Mock. Lane 3 – 6. Expression with p19; Lane 3. Full-length *Sa*Cas9. Lane 4. *Sa*Cas9 739N. Lane 5. *Sa*Cas9 740C. Lane 6. *Sa*Cas9 739N + *Sa*Cas9 740C. Lane 7 – 10. Expression without p19; Lane 7. Full-length *Sa*Cas9. Lane 8. *Sa*Cas9 739N. Lane 9. *Sa*Cas9 740C. Lane 10. *Sa*Cas9 739N + *Sa*Cas9 740C. Immunoblot using anti-FLAG® M2 antibody 1:1000, exposure time 8 sec.

To determine if the split domains can reassemble *in vivo* and function as the full-length proteins, genome editing of the *N. tabacum PDS* gene was tested in tobacco protoplasts. Since *N. tabacum* genome has two homologs of the *PDS* gene, derived from its ancestors *N. tomentosiformis* (t) and *N. sylvestris* (s), gRNAs were designed to target both homologs (142).

Protoplasts were isolated from *N. tabacum* var. Maryland Mammoth leaves from 2–3-week-old plants, obtaining approximately 6 x 10^4 protoplasts/mL. The protoplasts were co-transformed using PEG4000 with the plant expression vectors expressing mCherry (as a positive control), the vectors expressing *Sp*Cas9 or *Sa*Cas9 and a plant vector expressing a sgRNA targeting the *PDS* gene. It was possible to detect red mCherry fluorescence 24 hours following transformation, as shown in Figure 14 (and Appendix A3). Protoplasts were collected to assess editing of the *PDS* gene 48 hours post transformation.

Protoplast DNA was extracted and the target region of the *PDS* gene was amplified. The PCR products were cloned, and individual colonies were screened for mutations.

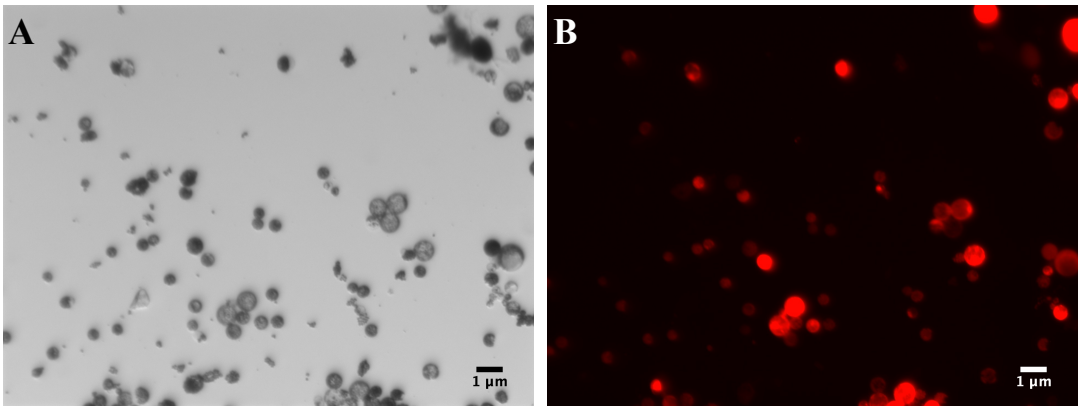


Figure 14. *N. tabacum* protoplasts transformed with split *Sa*Cas9. Isolated cells were transformed with plant vectors expressing mCherry, *Sa*Cas9 739N, *Sa*Cas9 740C and *PDS* sgRNA. Red fluorescence was checked after 24 hours post-transformation. **A.** bright field. **B.** Red fluorescence.

For *Sp*Cas9 full-length, eight colonies were sequenced and two showed insertions of 1 bp at the target site (Figure 15 A). For Split *Sp*Cas9 + inteins, 16 colonies were checked, with two exhibiting an insertion of 1 bp at the target position (Figure 15 B).

For *Sa*Cas9 full-length, six colonies were sequenced, identifying an insertion of 1 bp and two deletions of 1 bp (Figure 16 A). In the case of Split *Sa*Cas9, nine colonies were screened, identifying an insertion of 1 bp, a deletion of 1 bp and a transversion of C → A (Figure 16 B).

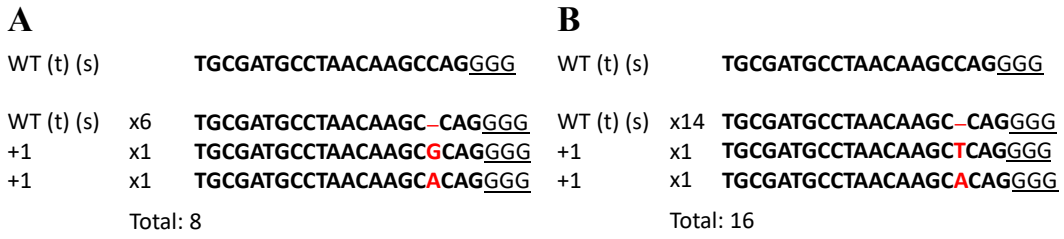


Figure 15. Targeted genome editing of *N. tabacum PDS* gene in protoplasts using *Sp*Cas9. A. Genome editing by *Sp*Cas9 full-length protein. Two insertions of 1 bp were detected. **B. Genome editing by split *Sp*Cas9 fused to inteins**. Two insertions of 1 bp were identified. (t). Genomic DNA sequence of *PDS* gene derived from *N. tomentosiformis*. (s) Genomic DNA sequence of *PDS* gene derived from *N. sylvestris.* The type of indel and the number of clones obtained for each sequence is indicated. Guide sequence is highlighted in bold, and the PAM site is underlined.

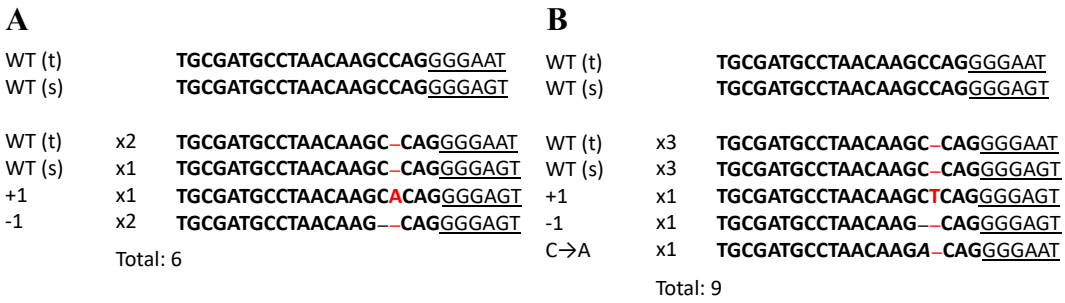


Figure 16. Targeted genome editing of *N. tabacum PDS* gene using *Sa*Cas9. A. Genome editing by *Sa*Cas9 full-length protein. An insertion and a deletion of 1 bp were detected. **B. Genome editing by split *Sa*Cas9 protein**. An insertion and two deletions of 1 bp were identified; additionally, a transversion C → A was found (shown in italic). (t). Genomic DNA sequence of *PDS* gene derived from *N. tomentosiformis*. (s) Genomic DNA sequence of *PDS* gene derived from *N. sylvestris.* The type of indel and the number of clones obtained for each sequence is indicated. Guide sequence is highlighted in bold, and the PAM site is underlined

Furthermore, two sgRNAs were designed to target exon 2 and exon 4 of the *N. tabacum FT4* gene (*NtFT4*), named TobFT1 and TobFT4.

Sequence alignments between TobFT1 and TobFT4 guides and *NtFT1-13* genes were carried out to determine possible off-targets sites (Figure 17). Mismatches with respect to the *NtFT4* sequence are shown in green and the target site is highlighted in orange.

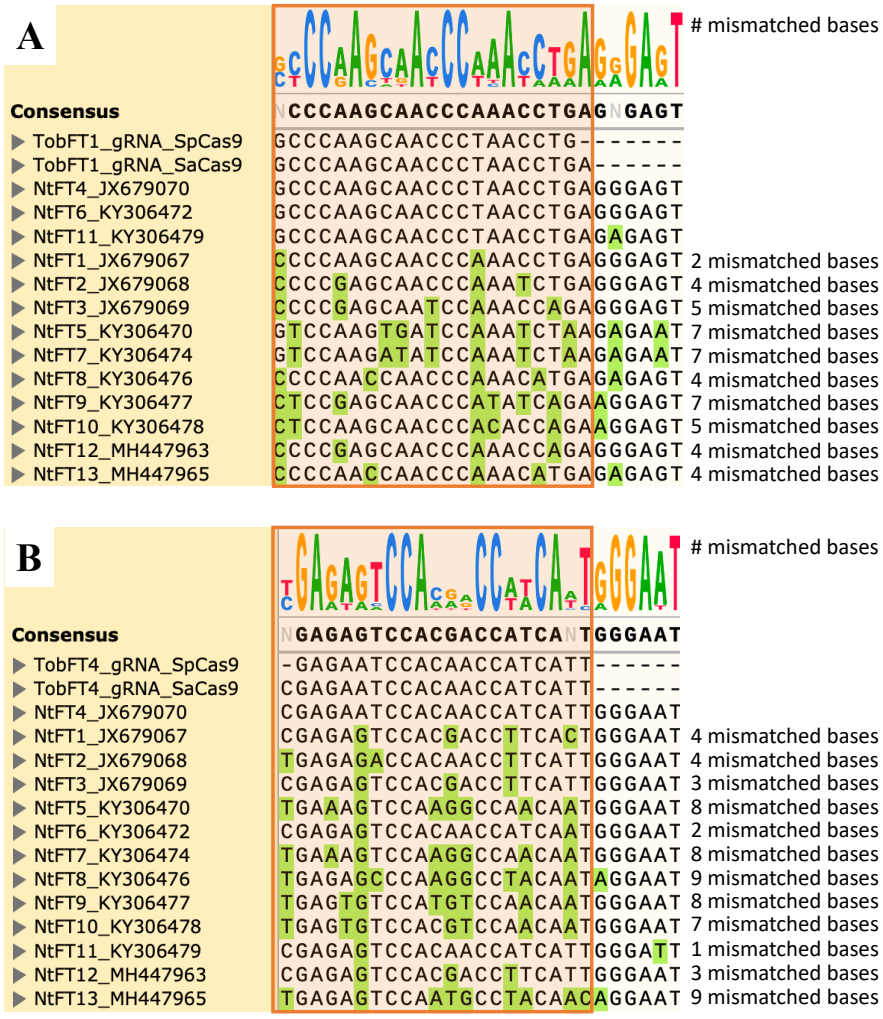


Figure 17. Determination of the off-target activity by sequence alignment between *NtFT4* guides RNA and *NtFT1-13* genes. GenBank accession number for each *NtFT* are included in the name of the sequence. Mismatches respect to the *NtFT4* gene sequence are shown in green. The target site in the exons 2 and 4 of *NtFT*s is highlighted in orange. **A. *NtFT* exon 2 alignment.** Sequence identity of the target site region between *NtFT4*, *NtFT6* and *NtFT11* is 100%, but *NtFT6* carries a nonsense mutation, hence is inactive and *NtFT11* has a mismatched base in the PAM site for both Cas9. Several mismatched bases were found between the target site in *NtFT4* gene and *NtFT1-3/5/7-10/13*. *B.* ***NtFT* exon 4 alignment.** TobFT4 guide RNAs are specific for *NtFT4* gene. Various mismatched bases between *NtFT4* and *NtFT1-3/5-13* were noticed in this target region.

Figure 17 A shown the alignment between TobFT1 *Sp*Cas9 and *Sa*Cas9 guides with the exon 2 of the *NtFT*s genes. Sequence identity of the target site region in exon 2 between *NtFT4*, *NtFT6* and *NtFT11* is 100%, however *NtFT6* carries a nonsense mutation, hence is inactive (163), while *NtFT11* exhibited a mismatched base in the PAM site for both Cas9. In comparison with *NtFT4*, *NtFT1* showed two mismatched bases, *NtFT2/8/12/13* four mismatched bases, *NtFT3/10* five mismatched bases and *NtFT5/7/9* seven mismatched bases.

On the other hand, Figure 17 B shows the alignment between TobFT4 *Sp*Cas9 and *Sa*Cas9 guides with the exon 4 of the *NtFT*s genes. In comparison with *NtFT4*, several mismatched bases are found in the Cas9 target site in the other *NtFTs*, indicating the specificity of both guides towards *NtFT4. NtFT11* exhibited one mismatched base in the target site*, NtFT6* two mismatched bases*, NtFT3/12* three mismatched bases*, NtFT1/2* four mismatched bases*, NtFT10* seven mismatched bases*, NtFT5/7/9* eight mismatched bases *and NtFT8/13* nine mismatched bases.

Targeted genome editing using TobFT1 and TobFT4 guide RNAs was tested using both full-length Cas9 enzymes. Protoplasts were transformed using the plant expression vectors described previously for *Sp*Cas9, *Sa*Cas9 and the *NtFT4* sgRNAs, DNA was isolated, the target gene was amplified and cloned to screen for mutations.

Genome editing of the *NtFT4* gene at the exon 2 target site was detected for both *Sp*Cas9 and *Sa*Cas9 enzymes using the TobFT1 guide. As shown in Figure 18, for *Sp*Cas9 full-length, 16 colonies were screened, of which one exhibited an insertion of 1 bp (panel A). For *Sa*Cas9 full-length, 15 colonies were analysed, with a deletion of 3 bp detected in one of the colonies (panel B).

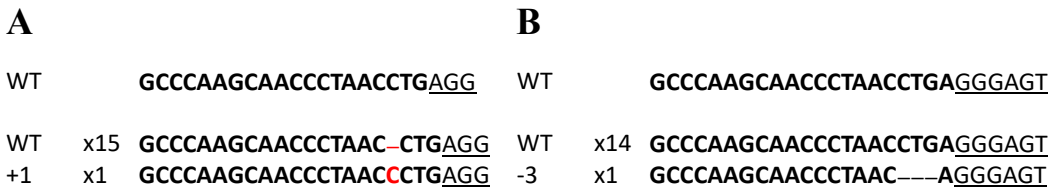


Figure 18. Targeted genome editing of exon 2 of *N. tabacum FT4* gene. A. Genome editing by *Sp*Cas9 full-length TobFT1 sgRNA. An insertion of 1 bp was found. **B. Genome editing by *Sa*Cas9 full-length TobFT1 sgRNA**. A deletion of 3 bp was detected. The type of indel and the number of clones obtained for each sequence is indicated. Guide sequence is highlighted in bold, and the PAM site is underlined.

In the case of the exon 4 target site of the *NtFT4* gene, for *Sp*Cas9 full-length 16 colonies were screened from which one showed an insertion of 1 bp (Figure 19 A). For *Sa*Cas9 full-length, 16 colonies were analysed, all of them being WT sequences (Figure 19 B).

In summary, split Cas9 proteins can be delivered separately into plant cells where they are able to re-assemble and introduce mutations into a desired target gene.

3.2.2 Assessment of the self-cleavage activity of ribozymes in the RGR unit

Hammerhead ribozymes were attached to each end of the *NtFT4* sgRNAs, creating the RGR unit (Figure 9), and their self-cleavage activity was tested. RGR units were transcribed *in vitro* by a T7 RNA Polymerase, the generated RNA was analysed by cRT-PCR and the products were cloned and sequenced (Figure 10). Using this method, it is possible to check the sequences of both ends of the sgRNAs after the excision of the HHRzs. The 5'-HHRz is excised completely from both sgRNAs, while the 3'-HHRz leaves a scar of six nucleotides (CCTGTC). Sequencing results showed that it was possible to detect both mature sgRNAs. For TobFT1 sgRNA (Figure 20 A) one of the sequences exhibited the full-length RNA scaffold (76 bp) plus the complete 3'- HHRz scar and additional five non-matching nucleotides (CCCGG). A second sequence displayed the full-length RNA scaffold and an incomplete 3'- HHRz scar of 4 bp (CCTG). Nine sequences displayed a shorter RNA scaffold of 32, 33, 60 or 61 bp long.

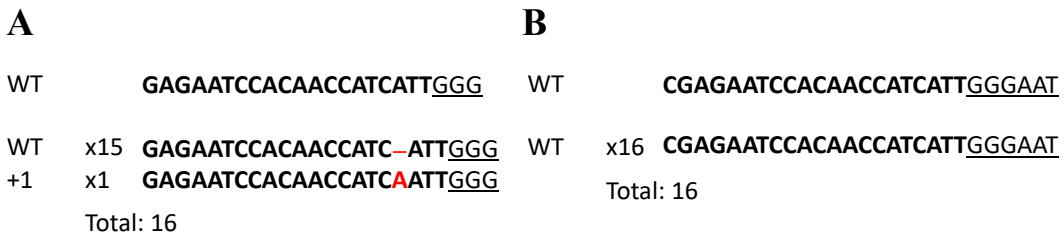


Figure 19. Targeted genome editing of exon 4 of *N. tabacum FT4* gene. A. Genome editing by *Sp*Cas9 full-length TobFT4 sgRNA. An insertion of 1 bp was found. **B. Genome editing by *Sa*Cas9 full-length TobFT4 sgRNA**. No gene editing was detected. The type of indel and the number of clones obtained for each sequence is indicated. Guide sequence is highlighted in bold, and the PAM site is underlined.

Similar results were obtained for TobFT4 sgRNA (Figure 20 B), where one sequence showed the full-length RNA scaffold, the complete 3'-HHRz scar and additional 15 bp. A second sequence also exhibited the full-length RNA scaffold and an incomplete 3'-HHRz scar of four nucleotides (CCTG). Incomplete RNA scaffolds of 33 and 73 bp long were also found in two sequences. Trimmed RNA scaffold sequences might be due to RNA degradation during sample preparation (101).

In summary, hammerhead ribozymes attached to the sgRNA are able to self-excise enabling the formation of a mature guide RNA for functional genome editing.

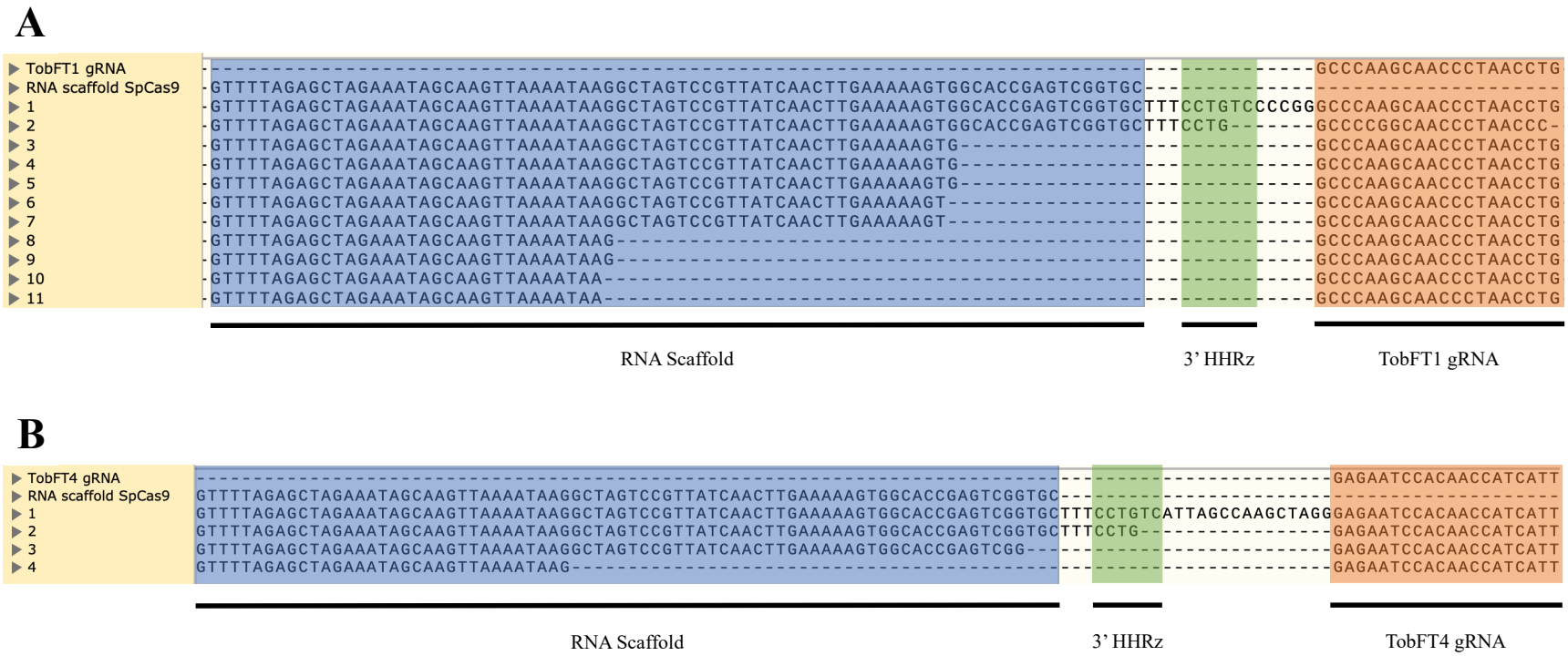


Figure 20. ***In vitro*** **self-cleavage activity of the ribozymes in the RGR unit.** Mature TobFT1 and TobFT4 gRNAs are detected. No 5'-HHRz was found, while the characteristic scar of 6 bp of the 3'-HHRz (in green) was noticed in both analyses. Trimmed RNA scaffold sequences were also distinguished. **A. TobFT1 sgRNA**. **B. TobFT4 sgRNA**. The sequences of the RNA scaffold, 3'HHRz scar and gRNA are shown in blue, green, and orange, respectively.

3.3 DISCUSSION

In this part of the work, each of the two domains of the split *Sp*Cas and *Sa*Cas9 proteins were cloned into plant expression vectors and delivered by *Agro*infiltration into *N. benthamiana* plants, or PEG transformation into *N. tabacum* protoplasts. Immunoblot results of the split *Sp*Cas9 showed reconstitution of the full-size Cas9 protein mediated by the fused inteins, while expression of both split domains of the *Sa*Cas9 protein were detected. Wright *et al.* reported splitting the *Sp*Cas9 enzyme into the nuclease and α-helical recognition lobes, where the functional re-association is through the formation of a ternary complex with the sgRNA (81). However, improvements to the system have been made to control the re-assembly of both halves. One of such improvements is the fusion of inteins to each half of Cas9 protein, which facilitates the functional reconstitution of the protein by the formation of a peptide bond between them (83).

The effect on the co-expression of *Agro*-infiltrated *Sp*Cas9 and *Sa*Cas9 proteins with the *Tomato bushy stunt virus* p19 protein was tested. p19 is a small protein of 19 kDa produced by the virus to suppress the PTGS machinery which protects plants from virus infection by degrading the viral RNA. The p19 protein binds to small interfering RNAs (siRNA) involved in the PTGS response to viruses, sequestering them and allowing viral proliferation (184, 186). In this work, the co-expression of p19 with *Sp*Cas9 and *Sa*Cas9 in *N. benthamiana* plants enhanced the levels of expression of the Cas9 proteins (Figures 12 and 13). In the case of split *Sp*Cas9, the re-assembled protein was observed on immunoblots only when p19 was co-infiltrated with both N- and C- domains (Figure 12). Zhang *et al.* (2019) found that genome edits were induced only when FoMV-*Sp*Cas9 was co-delivered with p19 fused to the sgRNA or in plants constitutively expressing p19, indicating that this peptide enhanced the levels of Cas9 protein in young leaves and efficiently introduced mutations in systemic leaves (136). However, in a different study using FoMV, the co-infiltration of p19 or HcPro, a different viral silencing suppressor, did not improve the systemic targeted mutagenesis of the *PDS* gene, due to the localized method of delivery of the PTGS suppressors. When TuMV (that encodes HcPro naturally) was co-infected with FoMV, an increase of the gene editing in systemic leaves was detected, however it was also found that the plants die about 10 dpi because the virus was highly virulent (185). Nonetheless, it has been reported that in

certain varieties of *N. tabacum* p19 induces a hypersensitive response, developing necrosis at the site of infection and resulting in a decrease of transgene expression (186). This effect is reported for the varieties I-64, TI-95, Petite Havana H4, NC95 and Xanthi, while Little Crittenden was unaffected (186-188). As an alternative to p19, the use of the TMV 126 kDa suppressor gene also enhanced TRV-mediated VIGS thus increasing silencing phenotypes (189). It has been also described that the 16 kDa protein encoded by TRV is a weak suppressor of RNA silencing compared to p19, thus for TRV no extra silencing suppressor is needed (108).

Immunoblot analysis using the anti-FLAG® M2 monoclonal antibody, shown an unspecific band at ~100 kDa in both infiltrated and mock *N. benthamiana* leaf samples. FLAG (sequence: DYKDDDDK) is a small amino acid hydrophilic epitope tag and it is one of the most popular tags in molecular biology. In other plant species a similar unspecific band has been reported (190-192). Nonetheless, when a monoclonal antibody against the N-terminal portion of *Sp*Cas9 was used, a single band was detected. So, the use of antibodies which recognize specifically either *Sp*Cas9 or *Sa*Cas9 should be considered.

Protoplasts allow a quick platform for *in vivo* screening and validation of genome editing events. Protocols for high yield of protoplasts isolation have been described for several plant species (176, 179). The absence of cell wall facilitates their transformation with exogenous components, such as transiently express vectors or RNP complexes and these transformed and edited protoplasts can be regenerated into non-GMO plants (175, 177). The co-deliver in protoplasts of fluorescent markers, such as GFP or mCherry, in the same plasmid with the CRISPR/Cas9 reagents facilitates the direct comparison of transformation efficiency. Moreover, protoplasts with fluorescent signal may also be enriched by cell-sorting, separating transformed from un-transformed populations for further analyses, such as sequencing (178). In this work, *N. tabacum* protoplasts were used to assess the *in vivo* activity of the re-assembled Cas9 proteins by targeted mutagenesis of the *PDS* gene and to validate the targeted genome editing of the *NtFT4* gene by the full-length Cas9 orthologs and designed gRNAs TobFT1 and TobFT4. As observed with the full-length proteins, re-assembled *Sp*Cas9 and *Sa*Cas9 successfully introduced targeted genome edits, suggesting that a functional Cas9 was formed *in planta.* Mutations were observed in exon 2 of *NtFT4* gene using TobFT1 gRNA and both Cas9 enzymes, however for the

guide RNA targeting the *NtFT4* exon 4 (TobFT4), an insertion was only detected using *Sp*Cas9 enzyme in the limited number of clones analysed.

Similar genome editing efficiencies between *Sa*Cas9 and *Sp*Cas9 have been reported by Kaya *et al.* (2016) in rice and *N. tabacum*, determined by the screening of eight to 24 colonies (142). In contrast, Raitskin et al. (2019) indicated a higher genome editing efficiency of *Sa*Cas9 compared to *Sp*Cas9, *As*Cas12a or *Lb*Cas12a. In their study, *N. benthamiana* protoplasts were transformed with a fused construct of yellow fluorescent protein (YFP), *Cas9* gene and sgRNA. Transformation efficiency was assessed by quantification of protoplasts with nuclei YFP signal. Targeted genome edits were identified by NGS and then the number of mutations at the target site was quantified using bioinformatic tools. To compare the targeted genome editing efficiency among the different Cas9 proteins, the quantity of mutations in each sample was normalised to the quantified transfection efficiency (193). A similar approach could had been used in this thesis work, to compare targeted genome editing efficiencies between Cas9 orthologs or full-length or split Cas9. However, this escapes the scope and aims of using protoplasts in this study, which were to prove *in vivo* activity of the re-assembled Cas9 and validate the introduction of targeted mutations in the *NtFT4* gene by the Cas9 orthologs and designed gRNAs.

Genome editing efficiency in protoplasts depend on the number of viable isolated cells and transformation efficiency. Factors such as plant genotype, age, source of tissue, plant growth conditions, enzyme mixture, incubation time and digestion or pH affect the isolation of plant protoplasts (176). Since each plant species needs a specific optimal condition, several protocols have been published suggesting improvements for efficient protoplasts isolation (176, 179). In this thesis work, cell density was assessed using a haemocytometer, with a yield of approximately 6 x 10^4 protoplasts/mL. Similar yields had been reported for other plant species (179). Protoplasts transformation can be by electroporation, microinjection, or PEG transformation. Among these, PEG-mediated method is mostly used because it offers high transformation efficiency and protoplasts viability, it is low cost and it doesn't require additional equipment (180, 194). Factors that affect protoplasts transformation mediated by PEG are plant species, source material, isolation methods, PEG concentration, transformation time, DNA (or RNP) concentration and cell number (179, 195). Transformation efficiency is determined by comparing the

number of fluorescing cells to non-fluorescing cells using a fluorescent microscope and a haemocytometer (196). However, in this thesis work, transformation efficiency was not estimated.

A great disadvantage of the CRISPR/Cas9 system is the possibility of off-target cleavage. Modrzejewski *et al.* (2019) defined an off-target effect due to genome editing as an "Unintended cleavage and mutations at untargeted genomic sites with similar but not identical sequences compared to the target site" (197). A systematic review in plants studied five factors that had been reported affect the incident of off-targets effects: Number of mismatches, position of mismatches, GC-content of the targeting sequence, altered Cas9 variants and delivery methods (198). The results indicated that the rate of off-target effects varied from 59% when one mismatch between the on-target and off-target site exists to 0% when four or more mismatches are presented. Mismatches located within the first eight bp adjacent to the PAM highly decreased the occurrence of off-targets effects. It was previously reported that to increase the on-target effect, a perfect complementation between the target site and the gRNA in the seed sequence (7-12 bp adjacent to the PAM) was recommended (46). No evidence was found indicating that the GC-content of the targeting sequence significantly affects the incidence of off-target effects, while the database concerning the impact of the nuclease variant and the delivery method was very poor to prove a significant impact of these factors on the occurrence of off-target events (198).

Therefore, the main factor to decrease the possibility of off-target effects due to genome editing is a careful design of the gRNA. In this thesis work, two sgRNAs were designed to target the *N. tabacum FT4* gene, an inducer of flowering in tobacco (159). As *Sp*Cas9 and *Sa*Cas9 differ in their PAM sites (NGG and NGGRRT, respectively), the guides were thoroughly designed to target the same region of exon 2 and exon 4 of the *NtFT4* gene. Possible off-targets were assessed by sequence alignment of 13 *NtFTs* orthologs. For TobFT1, potential off-targets were found in the exon 2 of *NtFT6* and *NtFT11*, however the first carries a non-sense mutation making it inactive (163), while *NtFT11* has a mismatched base in the PAM site of both Cas9. Mismatches of two to seven bp were found between the gRNAs and the other *NtFTs* orthologs. On the other hand, no off-targets were found for TobFT4 guide since several mismatched bases (one to nine bp) were noticed in the exon 4

target site between *NtFT4* and the other *NtFT*s sequences. Hence, these observations imply a high specificity between the designed gRNAs and *NtFT4* gene and suggests a low probability of off-target events with other *NtFTs* orthologs.

In order to increase the likelihood of delivering the sgRNA into the same cell as the Cas9 protein, a single transcriptional unit was created by fusing self-cleavable ribozymes to each end of the sgRNA to create an RGR unit, which will be cloned downstream of the *Cas9* gene in the plant viral vectors. This artificial gene can be expressed *in vivo* from a single promoter. The autocatalytic activity of the ribozymes was analysed by cRT-PCR, where the 3'-HHRz scar was detected for TobFT1 and TobFT4 *NtFT4* gRNAs. As expected, no 5'-HHRz was detected. Shorter scaffold RNA sequences were observed in both cases, probably because of RNA degradation during the experimental procedures, as reported by Tang *et al.* (101).

It has been reported that genome editing activity can be achieved from STU regardless of the presence or absence of HHRz. The study compared a STU with and without HHRz between the 3' end of the *Cas9* gene and the guide RNA and found that high editing efficiency was achieved regardless of the presence or absence of the HHRz, concluding that the ribozyme sequence is not necessary to generate functional gRNAs. The authors reported that mature guide RNAs are processed in planta from the *Sp*Cas9-gRNA STU, by the binding of the Cas9 protein and further cleavage activity of endogenous RNases, such as RNase III and RNase T1 (102). In *archaea* and bacteria, a similar processing system has been proposed, where mature crRNAs are cut from a polycistronic unit called pre-crRNA. This pre-crRNA forms a duplex with the tracrRNA, which is stabilized by the Cas9 protein, recognized, and cleaved by RNase III to produce functional crRNAs (199).

In summary, split Cas9 proteins can re-associate in plants cells and introduce mutations into a desired target as the full-length proteins. Guide RNAs specific for the tobacco *FT4* gene were designed and shown to induce genome edits at the targeted site. The addition of self-cleaving ribozymes to these guides enables the design of a polycistronic unit with the split and full-length Cas9 endonucleases to increase the efficiency of delivery of the CRISPR RNA and Cas9 protein into the same plant cell.

CHAPTER 4: POTATO VIRUS X AND TOBACCO RATTLE VIRUS AS VIRAL EXPRESSION VECTORS FOR THE DELIVERY OF Cas9-RGR TO EDIT *N. tabacum FT4* GENE

4.1 INTRODUCTION

In the past 20 years, strategies for crop breeding have increasingly involved the insertion of a gene of interest into the plant genome, creating a stable, transgenic line which will produce the heterologous proteins over successive generations. This process is slow, laborious and time consuming. An alternative approach of transiently expressing the heterologous proteins has been proposed, particularly via the delivery of the coding sequences by plant virus vectors. The use of plant viruses as vectors presents many advantages. Viruses spread systemically within days, delivering the transgene into mature leaves in less time compared with the development of stable, transgenic plants (107). High expression levels of heterologous proteins can be obtained since the virus replicates within the infected plant cells (139). Specifically, RNA viruses (apart from retroviruses) don't incorporate themselves into the host genome, then its use allows the development of non-integrative (non-GMO) strategies for crop improvement. The main disadvantage of using plant viruses as vectors is their limited cargo capacity, which can interfere with their stability. In some cases, this effect is related with the restricted amount of space inside the capsid, as the dsDNA *Cauliflower mosaic virus* (111). The cargo capacity limitation in vectors where a duplicated sub-genomic promoter (sgP) was used for gene expression has been studied. For example, Avesani *et al.* (2007) tested five different inserts with sizes between 0.2 to 1.7 kb to be expressed by PVX. The authors reported that the ratio of vectors without insert increased when larger inserts were delivered, since the viral replicase changes templates at sequence repeats leading to mutations or loss of the transgene (110). To solve this problem, the use of a different sgP from a related virus, or the fusion of the heterologous protein with the coat protein using a 2A catalytic peptide have been suggested (111). A different approach is to replace the movement proteins and coat protein genes with exogenous sequences, hence becoming non-infectious replicons, eliminating the possibility of spreading through the plant (117). Non-infectious viral replicons from Geminivirus (ssDNA viruses) have been developed from bean yellow dwarf virus and wheat dwarf virus and induce genome editing in several crops (200-203).

Cody and Scholthof (2019) defined three stages of the use of plant virus vectors through the years (204). First, virus vectors were used to express reporter genes, such as GUS or GFP to track virus movements in inoculated tissue and systemically (107, 114). Next, with the discovery of post-transcriptional RNA silencing and the advance of sequencing tools, it was possible to develop the virus-induced gene-silencing (VIGS) strategy to reduce the expression of plant target genes (120, 121). Lastly, with the rise of gene editing technologies, particularly the CRISPR/Cas9 system, virus vectors have been used to deliver its components into plants for targeted genome editing. Some advantages of the use of plant virus vectors with the CRISPR system are less off-target activity due to its transient expression and the possibility to induce genome edits in meristematic cells due to the ability of some viruses to infect this tissue (54, 108).

4.1.1 Use of viruses to deliver single or multiple sgRNAs into plants

As mentioned in section 1.2.3 of CHAPTER 1, several viruses have been used as vectors to deliver CRISPR/Cas9 components into plants. (+) ssRNA viruses have been used to deliver single or multiple sgRNAs into plants that have been transformed to constitutively express the Cas9 protein. Ali *et al.* (2015) used TRV to deliver single or multiple sgRNAs into *N. benthamiana* plants, finding genome edits in the target site(s) in inoculated leaves and systemic leaves. Mutations were found in seed progeny of these plants, indicating that TRV successfully crossed into meristematic tissue to deliver the sgRNA (128). Next, they tested the delivery of a sgRNA targeting the *PDS* gene in tobacco plants using TRV and PEBV, which can be used in important agronomical plants such as legumes. Genome edits were assessed by the T7 endonuclease I (T7EI) assay and frequencies were estimated by band intensity using ImageJ software, indicating that the PEBV system (57-63%) is more efficient than the TRV system (27-35%) (131).

Ellison *et al.* (2020) reported the use of mobile RNA elements, such as a segment of *FT* RNA, or tRNAs, that fused to sgRNAs and delivered by TRV improved the uniformity of the viral delivery in systemic leaves and the rate of successful edits (higher than 80%) in target genes in somatic and germ line cells of *N. benthamiana*. A previous report described that the *FT* RNA can mediate systemic trafficking of heterologous RNAs (including virus RNA) into the apical meristem (205). Heritable

mutations were found in progeny, 65% of seedlings derived from infected parents contained mutations in at least one *PDS* allele, while 100% of the seedlings derived from plants where the *AGAMOUS* gene was targeted showed mutations in the target site (97).

Uranga *et al*. (2021) reported the use of PVX to deliver single and multiple sgRNAs into *N. benthamiana*, where no statistical difference in the genome editing frequency was observed when tRNAs were fused at the 5'-, 3'-, or both ends of sgRNAs. Edited progeny obtained by tissue culture showed higher frequency of mutations compared with its parental tissue. As Ellison *et al.* (2020) reported for TRV, the use of *FT* allowed the movement of PVX into germ line cells, achieving biallelic genome edits in 22% of the analysed seedlings collected from PVX infected plants without the need for tissue culture (134).

Hu *et al.* (2019) engineered the barley stripe mosaic virus to deliver sgRNAs into wheat and maize plants constitutively expressing Cas9. Genome edits rates up to 78% were determined for the *TaGASR7* target gene in wheat, involved in the control of grain length and weight, whilst for maize, genome edits frequencies in the *ZmTMS5* gene target, involved in pollen fertility, were 48% (132).

Jiang *et al.* (2019) developed the Beet necrotic yellow vein virus as a vector to deliver heterologous proteins and tested it use for genome editing by expressing a sgRNA targeting the *PDS* gene in tobacco, finding that 78% (26/30) of the infected plants developed a photobleaching phenotype (133).

Mei *et al.* (2019) used FoMV to express sgRNAs to introduce genome edits into the *PDS* gene in tobacco, finding that at the 7 dpi frequencies of genome editing in the infiltrated leaf were between 74% to 91%, while in the top systemic leaf this percentage was between 0 to 8% and in flower tissues between 6.9% to 14.8%, however no mutations were inherited. Localized expression of the *Tomato bushy stunt virus* p19, an RNA silencing suppressor, did not improve the genome editing efficiency in infiltrated and systemic leaves. However, an increased rate of mutations (>70%) was found in systemic leaves when a TuMV virus carrying the RNA silencing suppressor HcPro was co-delivered, indicating a synergistic interaction with FoMV (185).

In the above studies, sgRNAs were delivered by viruses into plants constitutively expressing Cas9. In contrast, Cody *et al.* (2017) tested the use of a second generation

TMV vector (TRBO from TMV RNA-based overexpression), a vector for high efficiency overexpression, where the coat protein is deleted preventing its systemic movement, but still allowing localized cell-to-cell movement. *N. benthamiana* 16c plants (constitutively expressing GFP) were co-infiltrated with *Sp*Cas9 and TRBO-sgRNA targeting GFP and indels frequencies were increasing from less than 2% to 48% at 3 dpi and then to 70% between 6 and 7 dpi. Next, genome edits were successfully induced into endogenous tobacco genes using single or multiple sgRNAs, and particularly in these case, higher genome edits frequencies were found when multiple sites were targeted (130).

4.1.2 Use of virus to deliver full-length or split Cas9 into plants

Even though a negative correlation between the insert size and the stability of the vector has been reported, studies where the full-length Cas9 with the sgRNA were delivered using plant RNA viral vectors have been published. Zhang *et al.* (2019) report successful genome edits in the *PDS* target gene when the *Sp*Cas9 and sgRNA were co-delivered by the FoMV vector. They also show that the co-expression with *p19* greatly enhanced expression of the Cas9 protein, leading to efficient genome editing in systemic leaves. It is noteworthy that the authors found mostly substitutions, and only one small deletion (136). Ariga *et al.* (2020) used PVX to deliver a *Sp*Cas9-sgRNA fusion into *N. benthamiana* plants to target the *PDS* and *TOM1* genes. *Agrobacterium* harbouring the PVX vector was infiltrated into tobacco leaves along with a vector expressing *p19,* and genome editing analyses were carried out on these infiltrated leaves. CAPS analysis identified a higher efficiency of editing compared with *Agro*-infiltrated leaves where *Sp*Cas9 and the sgRNAs were transiently expressed using non-viral plant expression vectors. The authors also tried to obtain marker-free genome edited plants by regeneration of shoots without using antibiotic selection. From 50 regenerated shoots from PVX-*Sp*Cas9-*TOM1* sgRNA infiltrated leaves, 31/50 shoots carried mutations in one or both *TOM1* alleles, while the presence of PVX RNA was confirmed by RT-PCR in 15/17 shoots (88%), indicating that PVX infects most of the plant cells in infiltrated leaves and it is able to express sufficient amounts of Cas9 protein. Edited progeny from these regenerated plants were obtained, whilst no PVX was detected in these progenies, showing that it is possible to obtain genome editing plants without viral RNA or exogenous DNA. Finally, to expand on this idea, mechanical passage of PVX virions

isolated from infected leaves (sap infection) and further shoot regeneration from these leaves was tested. Genome edits were found in both *TOM1* alleles, but with a lower efficiency compared with shoots regenerated directly from *Agro*-inoculated leaves with PVX-*Sp*Cas9-sgRNA (139).

In contrast to (+) ssRNA viruses, such as PVX or FoMV, it has been reported that (-) ssRNA rhabdoviruses can stably express genes up to 6 kb (206). Ma *et al.* (2020) used the *Sonchus yellow net rhabdovirus* to deliver a *Sp*Cas9-tRNA-sgRNA-tRNA construct into *N. benthamiana* 16c plants stably expressing GFP and assessed genome editing frequencies by PCR/RE analysis and band intensity using ImageJ software. The rate of mutation of four target genes varied between 40% to 91%, depending on the gRNA used and the plant and similar frequencies were found when multiple sgRNAs were delivered simultaneously. Shoot regeneration of plants from infected leaves with SYNV-*Sp*Cas9-*PDS* sgRNA (M0 plants) showed that 17/30 (~57%) regenerated plants were homozygous/bi-allelic, while 28/30 (~93%) shoots carried targeted mutations of any type. Similar frequencies were found when two more target sites were analysed. Finally, even though SYNV doesn't infect germ line cells, the authors demonstrated stable inheritance of mutations in M1 and M2 progenies (137). Gao *et al.* (2019) engineered the *Barley yellow striate mosaic virus* as vector. *Sp*Cas9 and a sgRNA targeting GFP where cloned into an *Agrobacterium* BYSMV vector and delivered into *N. benthamiana* 16c plants, showing successful targeted genome edits (138)

Additionally, the use of split *Sa*Cas9 has been reported to meet the viral cargo capacity criteria. Kaya *et al.* (2017) tested two version of split *Sa*Cas9 (430N/431C and 739N/740C), where one fragment of split *Sa*Cas9 was delivered into tobacco plants by ToMV and the other end by a plant expression vector to prevent viral cross protection (140). CAPS analysis and further sequencing of clones derived from undigested bands showed that the construct split at position 739N/740C was more efficient inducing mutations into the *PDS* target gene than the 430N/431C split version. Moreover, no genome edits were obtained when full-length *Sa*Cas9 was delivered by ToMV (143).

In summary, the use of plant virus vectors offers many advantages in the transient expression of heterologous proteins in plants. CRISPR/Cas9 allows the introduction of targeted genome edits to manipulate important traits to improve crop varieties.

In this chapter, the delivery by TRV and PVX of full-length and split *Sp*Cas9 or *Sa*Cas9 fused to an RGR unit into *N. tabacum* var. Maryland Mammoth plants was tested. In the case of split Cas9, the delivery of the fragments by the same, or different virus vectors was analysed. The presence of Cas9 protein and mRNA, sgRNAs processing by self-cleavage ribozymes, and genome edits in the *NtFT4* target gene that could lead to a delay in the flowering time, were all assessed.

4.2 RESULTS

4.2.1 Virus vectors to systemically deliver CRISPR/Cas9 components into tobacco plants.

Using plant plasmid expression vectors, it was possible to successfully deliver the CRISPR/Cas9 components into plant cells and detect targeted genome editing in both *PDS* and *NtFT4* genes (as shown in CHAPTER 3.). However, such expression is transient and localized just to the area of *Agro*infiltration. Plant viruses, such as PVX and TRV, have been developed as vectors for *in planta* gene expression, allowing the delivery of transgenes systemically throughout inoculated plants.

N. benthamiana plants infected with a PVX vector expressing GFP, or an empty TRV virus expression vector, developed systemic symptoms between 7-10 dpi. Compared with a non-infected plant (Figure 21A), PVX-GFP infected plants exhibited chlorotic mottled leaves (Figure 21 B) and GFP signal was detected under UV light in inoculated leaves and in systemic leaves (Figure 21 C), spreading through the vascular tissue and petiole into the proximal part of the systemic leaf initially (white arrow) and then eventually through the whole leaf.

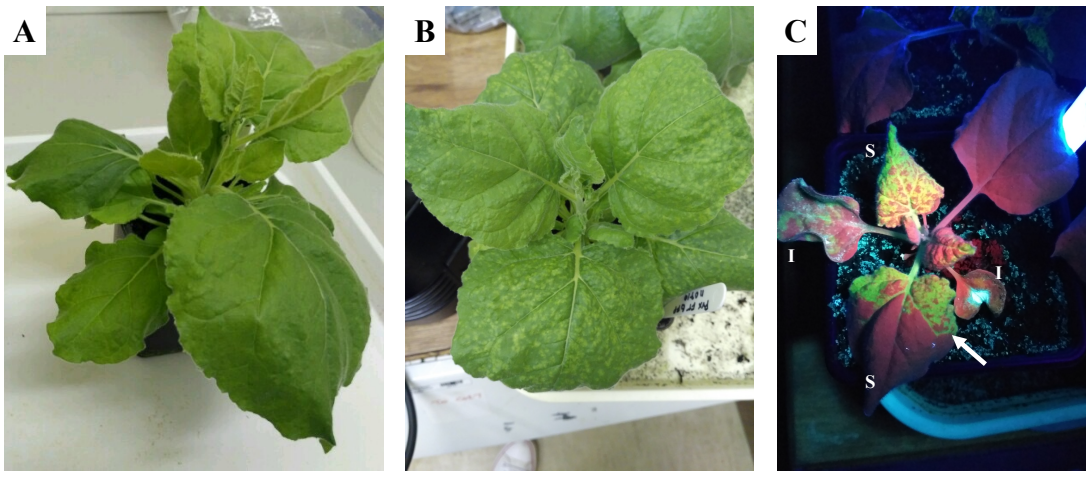


Figure 21. *N. benthamiana* plants infected with PVX-GFP virus vector at 18 dpi. A. Non-infected plant. **B.** Chlorotic mottled leaves indicate virus infection. **C.** Under UV light, GFP signal was detected in the infected (I) and systemic (S) leaves. The white arrow shows the initial spread of virus in the proximal part of a systemic leaf.

TRV infected plants are smaller in size compared to the mock control (Figure 22 A). As seen in Figure 22 B, TRV infected plants display chlorotic curled leaves (white arrow), and necrosis has developed through the main vein (black arrow).

To create the viral vectors for genome editing, different Cas9-RGR transcription units were constructed, cloning the RGR downstream of either *SpCas9, SaCas9,* or the split versions of the *Cas9* genes. This transcriptional unit was cloned into PVX and TRV2 vectors under the control of the duplicated coat protein promoter for PVX, or the PEBV coat protein promoter for TRV2. The final vector maps and sequences are presented in Appendix A1.

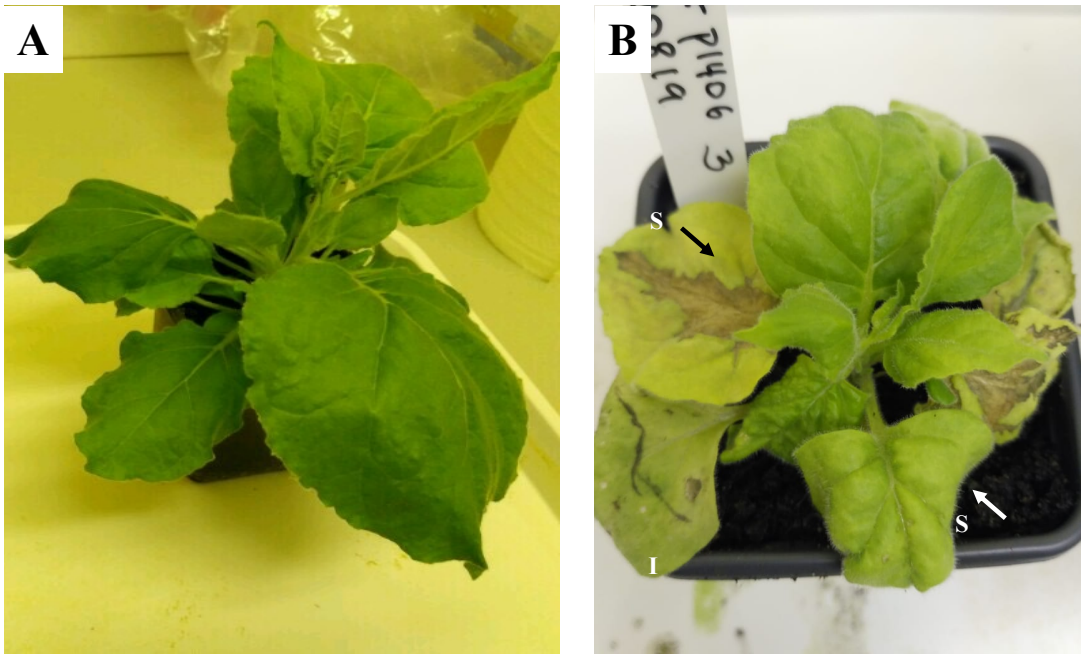


Figure 22. *N. benthamiana* plants infected with TRV virus vector at 18 dpi. Compared to the mock control **(A)** infected plants **(B)** are smaller in size, exhibiting chlorotic curled leaves (white arrow), with necrosis through the main vein (black arrow) of infected (I) and systemic (S) leaves.

4.2.2 Delivery of the *Sp*Cas9 - RGR constructs into *N. tabacum* plants by TRV and PVX viral vectors

3–4-week-old *N. tabacum* plants were infected with the TRV and PVX vectors expressing *Sp*Cas9 full-length TobFT1, or TobFT4 sgRNAs, or a combination of N-*Sp*Cas9 N-intein TobFT1 and C-intein C-*Sp*Cas9 TobFT4. Three independent plants (biological replicates) were infected per virus construct. Three independent plants were inoculated with water only, corresponding to mock controls (WT). Symptoms of infection in the infiltrated leaves were visible after approximately 7 dpi, while systemic symptoms were noticeable 14 dpi.

When TRV was used as vector to infect *N. tabacum* plants, clear systemic symptoms of TRV infection such as curled leaves and necrosis through the main vein were observed, indicating successful viral infection (Figure 23 A-C).

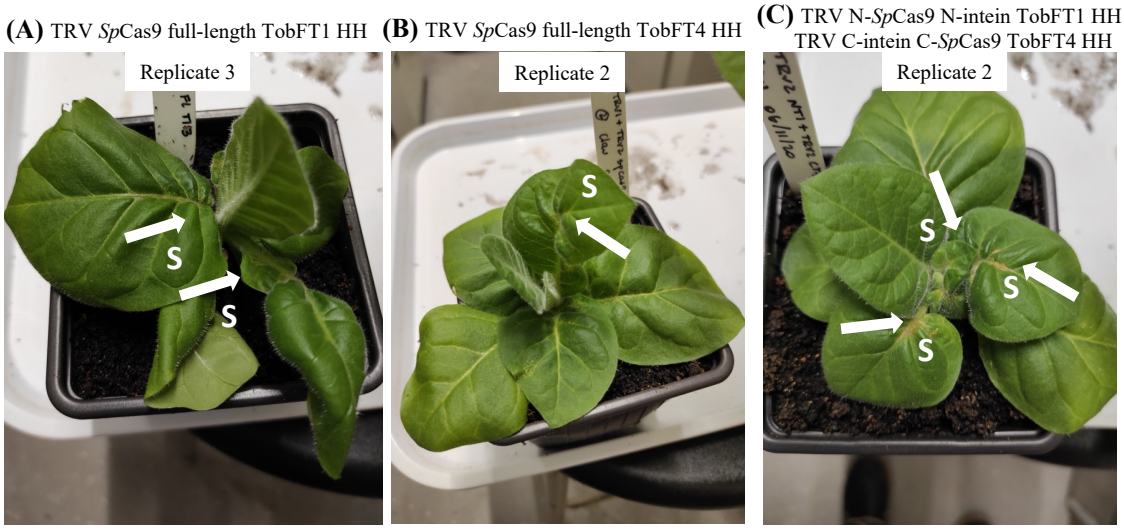


Figure 23. Tobacco plants infected with TRV *Sp*Cas9-NtFT4 HH sgRNAs constructs showing clear symptoms of systemic viral spread. A. TRV *Sp*Cas9 full-length TobFT1 HH. **B.** TRV *Sp*Cas9 full-length TobFT4 HH. **C.** TRV N-*Sp*Cas9 N-intein + TRV C-intein C-*Sp*Cas9. White arrows indicate visible viral infection symptoms in systemic leaves.

On the other hand, no evident systemic symptoms were detected in tobacco plants infected with PVX-*Sp*Cas9-RGR, in contrast plants infected with PVX-GFP exhibited chlorosis in systemic leaves (Figure 24 A-C).

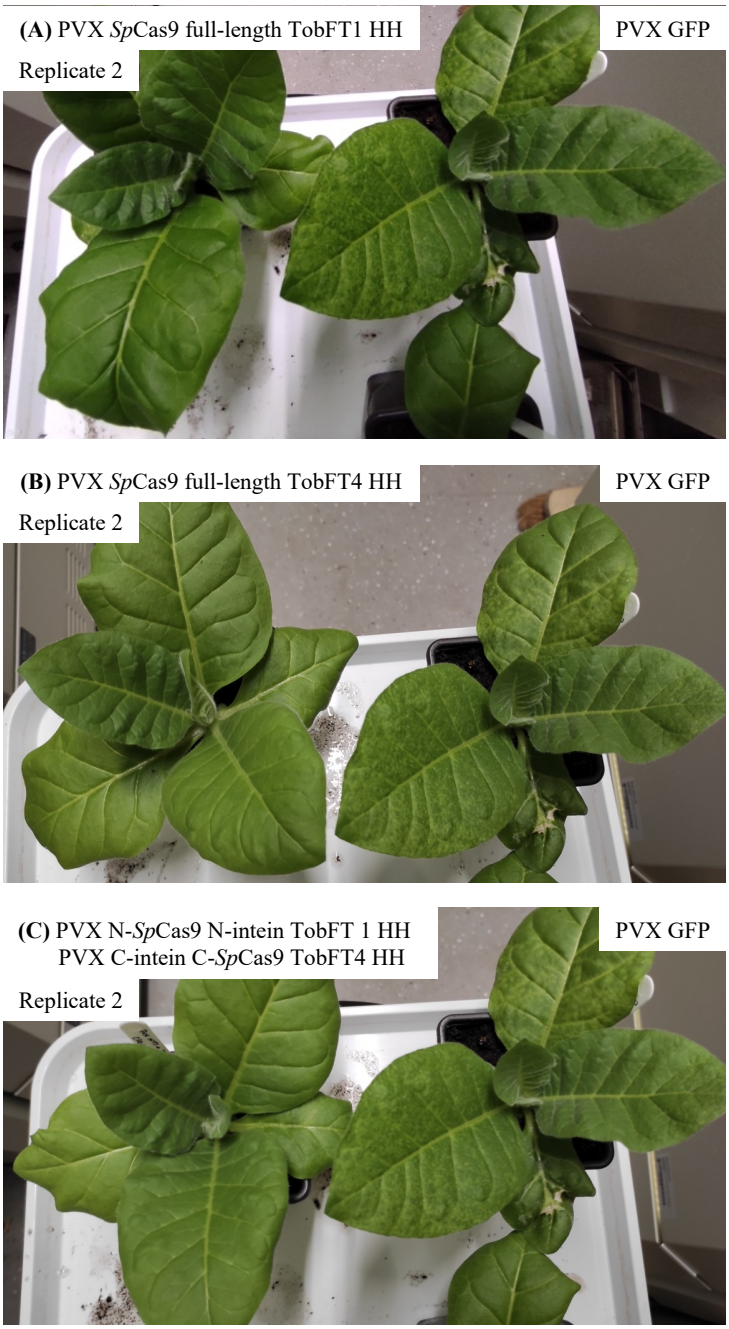


Figure 24. Representative images of tobacco plants infected with PVX *Sp*Cas9-NtFT4 HH sgRNAs constructs. No clear systemic symptoms of viral spread were observed, unlike plants infected with PVX-GFP**. A.** Left TRV *Sp*Cas9 full-length TobFT1 HH, right PVX-GFP. **B.** Left TRV *Sp*Cas9 full-length TobFT4 HH, right PVX GFP. **C.** Left TRV N-*Sp*Cas9 N-intein + TRV C-intein C-*Sp*Cas9, right PVX-GFP.

A combination of TRV and PVX vectors was tested for the delivery of the split *Sp*Cas9 domains. Figure 25 shows three biological replicates of plants co-infected with PVX N-*Sp*Cas9 N-intein TobFT1 HH and TRV C-intein-C-*Sp*Cas9 TobFT4 HH in panel A, while three biological replicates of plants co-infected with TRV N-*Sp*Cas9 N-intein TobFT1 HH and PVX C-intein-C-*Sp*Cas9 TobFT4 HH are shown in panel B. Clear symptoms of systemic TRV infection can be seen in some of these plants (white arrows), however, no PVX symptoms were evident.

Infected and systemic leaf samples from all the infected plants were collected approximately 14 dpi for protein, RNA, and genome editing analysis

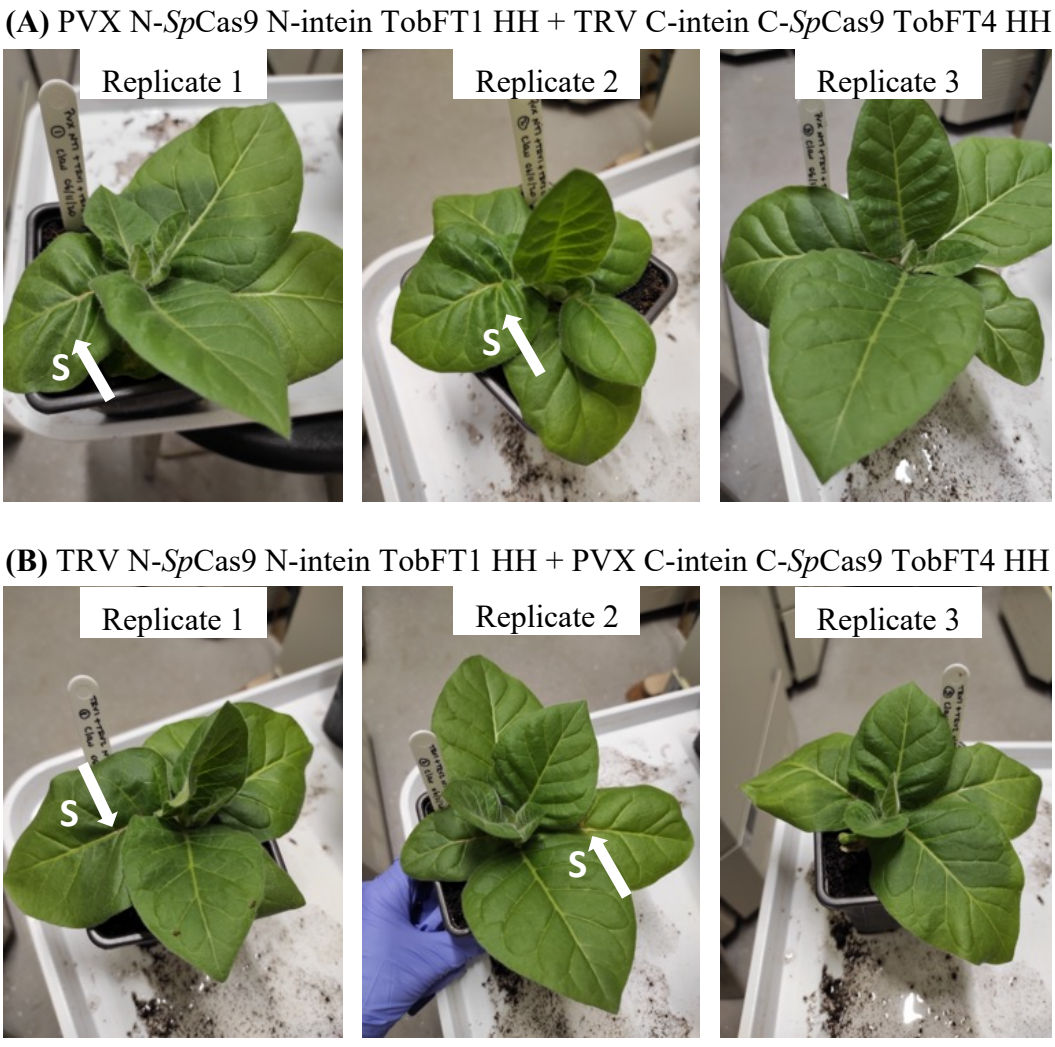


Figure 25. Representative pictures of the co-delivery of split *Sp*Cas9 into *N. tabacum* plants using PVX and TRV. Three replicates for each combination are shown. Clear symptoms of systemic TRV infection are seen in some plants (white arrows), whilst no PVX symptoms are observed. **A.** PVX N-*Sp*Cas9 N-intein TobFT1 HH and TRV C-intein-C-*Sp*Cas9 TobFT4 HH replicates 1, 2 and 3. **B.** TRV N-*Sp*Cas9 N-intein TobFT1 HH and PVX C-intein-C-*Sp*Cas9 TobFT4 HH replicates 1, 2 and 3.

4.2.2.1 Detection of *SpCas9* mRNA in TRV/PVX infected and systemic tobacco leaves

The detection of transcripts of the different genes expressed from the PVX and TRV2 vectors in infected and systemic leaves was analysed. Total RNA was extracted, cDNA was synthesized from these samples and RT-PCR analysis was conducted. Appropriate PCR controls were included. Positive controls of amplification correspond to plasmids harbouring the different constructs or empty viral plasmids. Negative controls of amplification correspond to no RNA for cDNA synthesis (RT-PCR negative control) or no cDNA (PCR negative control) to discard the presence of contaminants. Raw images of the agarose gels are shown in Appendix A4, supplementary figures 9, 10 and 13.

As validation of the RNA quality, amplification of a region of 372 bp of the housekeeping gene *ELONGATION FACTOR 1-ALPHA* was carried out. As shown in panels A and B of Figures 26, 27 and 28, strong amplification of *EF1α* was obtained in all infected, systemic and WT mock samples, indicating that the extracted RNA was of good quality for further analysis.

Transcripts of the N-terminal and C-terminal domains of the *Sp*Cas9 protein were detected in leaves infected with TRV or PVX constructs using specific primers for each *Sp*Cas9 domain (Figure 26 A and 27 A, respectively).

In TRV or PVX infected leaves the presence of TRV1/TRV2 (Figure 26 A) and PVX (Figure 27 A) viruses was confirmed. Bands for TRV1 and TRV2 can be seen in mock controls (lanes 13 – 15), but as the cDNA synthesis negative control (Lane 16) and PCR negative control (Lane 17) are both clean, it is likely that this is most probably due to cross-contamination in earlier steps of the protocol.

When TRV or PVX systemic leaf samples were analysed using the same conditions as for infected leaves, no amplification of *Sp*Cas9 N-terminal, C-terminal, or PVX virus was detected (Figures 26 B and 27 B). Some bands for *TRV1* (Figure 26 B, lanes 8, 9 and 11) and *TRV2* RNAs (Figure 26 B, lane 11) are observed. Even though, systemic symptoms were observed in infected tobacco plants with TRV, no *SpCas9* mRNA was detected. In the case of PVX systemic leaves, the absence of

PVX and *SpCas9* mRNA correlates with what was shown in Figure 24, where no symptoms of PVX infection were observed in these leaves.

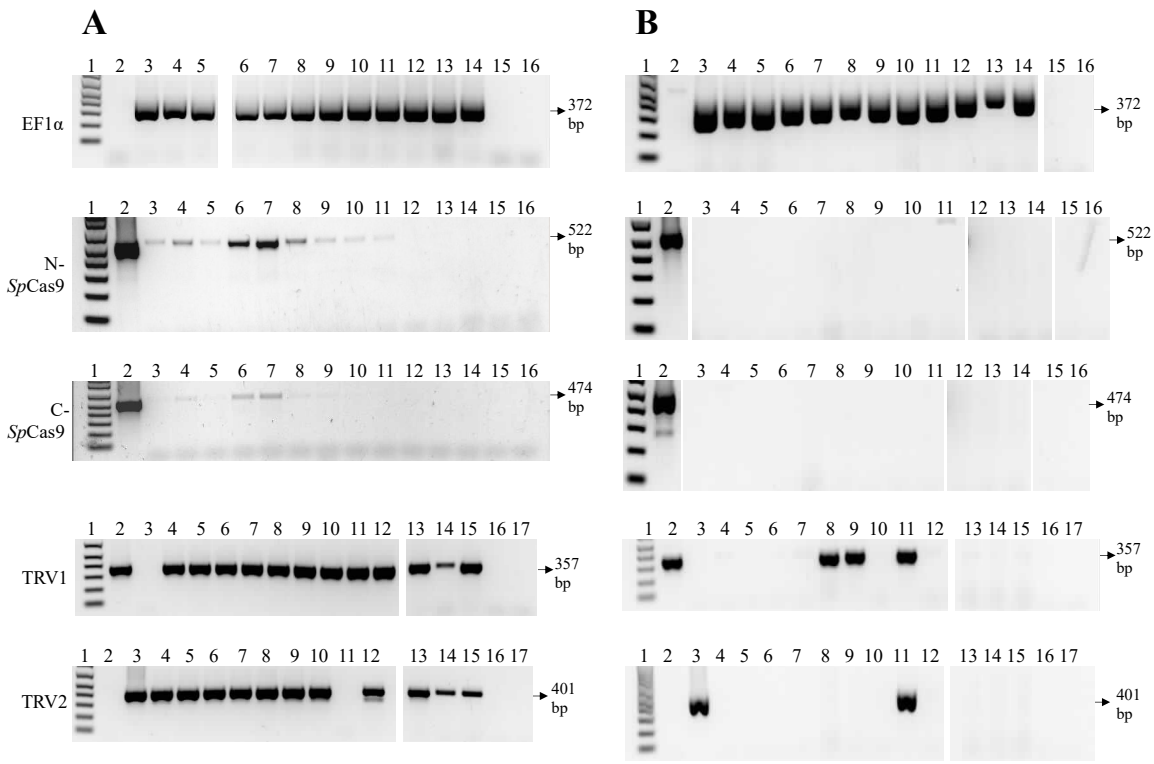


Figure 26. Detection of *SpCas9* mRNA in TRV infected and systemic *N. tabacum* leaves. A. TRV infected leaves. B. TRV systemic leaves. Three biological replicates per construct were analysed. PCR products size is indicated in each image.

EF1α samples were loaded as following. Lane 1. 1 kb plus DNA ladder. Lane 2. Empty. Lanes 3 – 5. Mock controls. Lanes 6 -8. TRV *Sp*Cas9 full-length TobFT1 HH. Lanes 9 - 11. TRV *Sp*Cas9 full-length TobFT4 HH. Lanes 12 – 14. TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4. Lane 15. RT-PCR negative control. Lane 16. PCR negative control.

*Sp*Cas9 N- and C- terminal infected and systemic leaves samples are as following: Lane 1. 1 kb plus DNA ladder. Lane 2. Positive control. Lanes 3 – 5. TRV *Sp*Cas9 full-length TobFT1 HH. Lanes 6 - 8. TRV *Sp*Cas9 full-length TobFT4 HH. Lanes 9 – 11. TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4. Lanes 12 – 14. Mock controls. Lane 15. RT-PCR negative control. Lane 16. PCR negative control.

TRV1 and TRV2 loading order is Lane 1. 1 kb plus DNA ladder. Lane 2. TRV1 positive control. Lane 3. TRV2 positive control. Lanes 4 – 6. TRV *Sp*Cas9 full-length TobFT1 HH. Lanes 7 - 9. TRV *Sp*Cas9 full-length TobFT4 HH. Lanes 10 – 12. TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4. Lane 13 - 15. Mock controls. Lane 16. RT-PCR negative control. Lane 17. PCR negative control.

The expression of *SpCas9* mRNA in infected and systemic leaves of plants where the split *Sp*Cas9-RGR constructs were delivered by co-infection of PVX and TRV was analysed. As shown in Figure 28 A and B, RNA of good quality was obtained from all samples, as indicated by the strong bands for the housekeeping gene *EF1α*. Expression of *Sp*Cas9 N- and C- terminal domains can be detected in all the infected leaves (Panel A), but not in systemic samples (Panel B). *PVX* RNA virus is also only detected in infected leaves, while *TRV1* and *TRV2* RNA viruses can be detected in both infected and systemic leaves of co-infected plants.

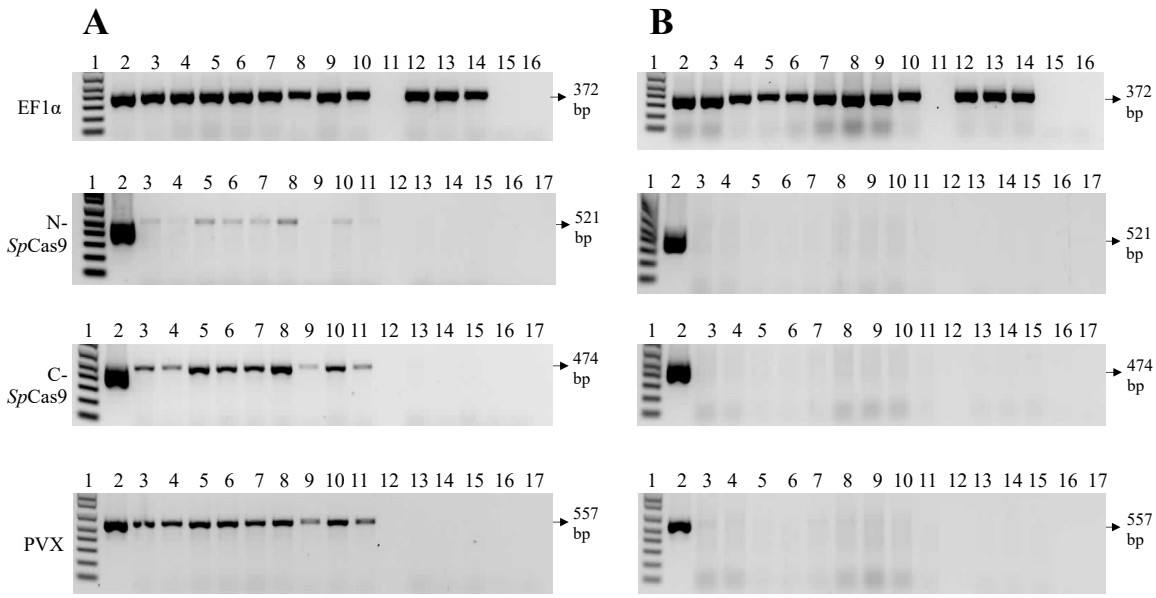


Figure 27. Detection of *SpCas9* mRNA in PVX infected and systemic *N. tabacum* leaves.

A. PVX infected leaves. **B.** PVX Systemic leaves. The order of the samples is the same for both panels. Three biological replicates per construct were analysed. PCR products size is indicated in each image.
In the case of EF1α samples were loaded as following. Lane 1. 1 kb plus DNA ladder. Lanes 2 -4. PVX *Sp*Cas9 full-length TobFT1 HH. Lanes 5 -7. PVX *Sp*Cas9 full-length TobFT4 HH. Lanes 8 – 10. PVX N-*Sp*Cas9 N-intein TobFT1 HH + PVX C-intein C-*Sp*Cas9 TobFT4. Lane 11. Empty. Lanes 12 – 14. Mock controls. Lane 15. RT-PCR negative control. Lane 16. PCR negative control.

For SpCas9 N- and C- terminal and PVX the order of the samples is: Lane 1. 1 kb plus DNA ladder. Lane 2: Positive control. Lanes 3 – 5. PVX *Sp*Cas9 full-length TobFT1 HH. Lanes 6 -8. PVX *Sp*Cas9 full-length TobFT4 HH. Lanes 9 – 11. PVX N-*Sp*Cas9 N-intein TobFT1 HH + PVX C-intein C-*Sp*Cas9 TobFT4. Lane 12. Empty. Lanes 13 – 15. Mock controls. Lane 16. RT-PCR negative control. Lane 17. PCR negative control.

This is consistent with what was observed above for the PVX and TRV viruses alone (Figure 25 A and B).

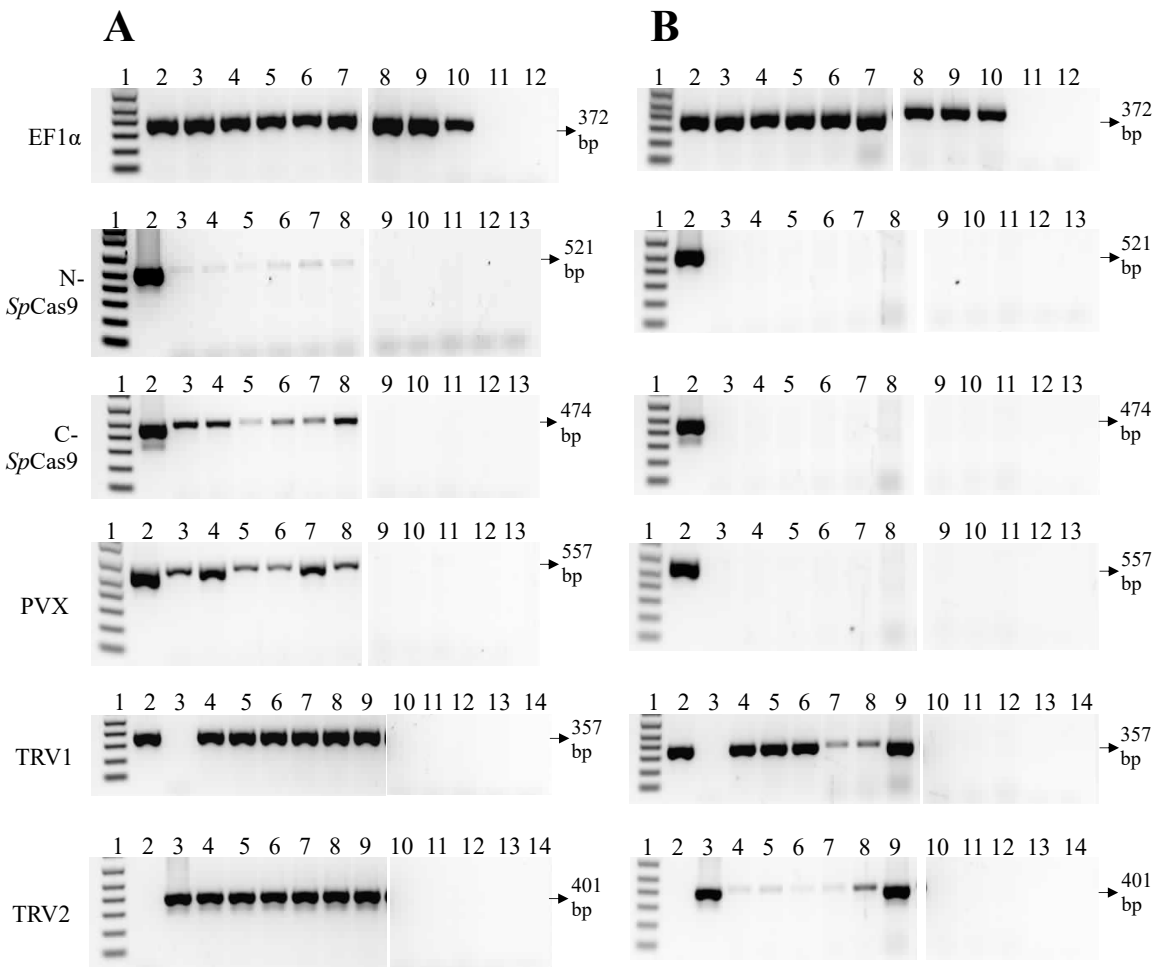


Figure 28. Detection of *SpCas9* mRNA in samples co-infected with TRV and PVX to deliver the split *Sp*Cas9-RGR unit. A. Infected leaves. B. Systemic leaves. Three biological replicates were analysed. PCR products size is indicated in each image.

EF1α samples were loaded as following. Lane 1. 1 kb plus DNA ladder. Lanes 2 – 4. TRV N-*Sp*Cas9 N-intein TobFT1 HH + PVX C-intein-C-*Sp*Cas9 TobFT4 HH. Lanes 5 – 7 PVX N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein-C-*Sp*Cas9 TobFT4 HH. Lanes 8 – 10. Mock controls. Lane 11. RT-PCR negative control. Lane 12. PCR negative control.

*Sp*Cas9 N- and C- terminal and PVX are as following: Lane 1. 1 kb plus DNA ladder. Lane 2. Positive control. Lanes 3 – 5. TRV N-*Sp*Cas9 N-intein TobFT1 HH + PVX C-intein-C-*Sp*Cas9 TobFT4 HH. Lanes 6 – 8 PVX N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein-C-*Sp*Cas9 TobFT4 HH. Lanes 9 – 11. Mock controls. Lane 12. RT-PCR negative control. Lane 13. PCR negative control.

TRV1 and TRV2 loading order is Lane 1. 1 kb plus DNA ladder. Lane 2. TRV1 positive control. Lane 3. TRV2 positive control. Lanes 4 – 6. TRV N-*Sp*Cas9 N-intein TobFT1 HH + PVX C-intein-C-*Sp*Cas9 TobFT4 HH. Lanes 7 – 9 PVX N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein-C-*Sp*Cas9 TobFT4 HH. Lanes 10 – 12. Mock controls. Lane 13. RT-PCR negative control. Lane 14. PCR negative control.

4.2.2.2 Immunoblot analysis to detect *Sp*Cas9 protein in TRV/PVX infected and systemic tobacco leaves

Infected and systemic leaf samples were analysed to verify *Sp*Cas9 protein expression. Western Blot analyses using a monoclonal antibody specific for *Sp*Cas9 were carried out (Figure 29). Coomassie or Ponceau red stained membranes showing the amount of transferred proteins are shown in Appendix A5. In leaf samples infected with TRV *Sp*Cas9 full-length and TobFT1 sgRNA (Figure 29 A, lanes 3 – 5) or TobFT4 sgRNA (Figure 29 A, lanes 6 to 8), an expected band of around ~160 kDa was detected, indicating that the full-size protein was successfully delivered by TRV and expressed in plant cells. In samples where TRV N-*Sp*Cas9 N-intein TobFT1 HH and TRV C-intein C-*Sp*Cas9 TobFT4 were co-infiltrated no re-assembled Cas9 protein or N-terminal domain (the portion recognized by the antibody) was detected (Figure 29 A, lanes 9 – 11).

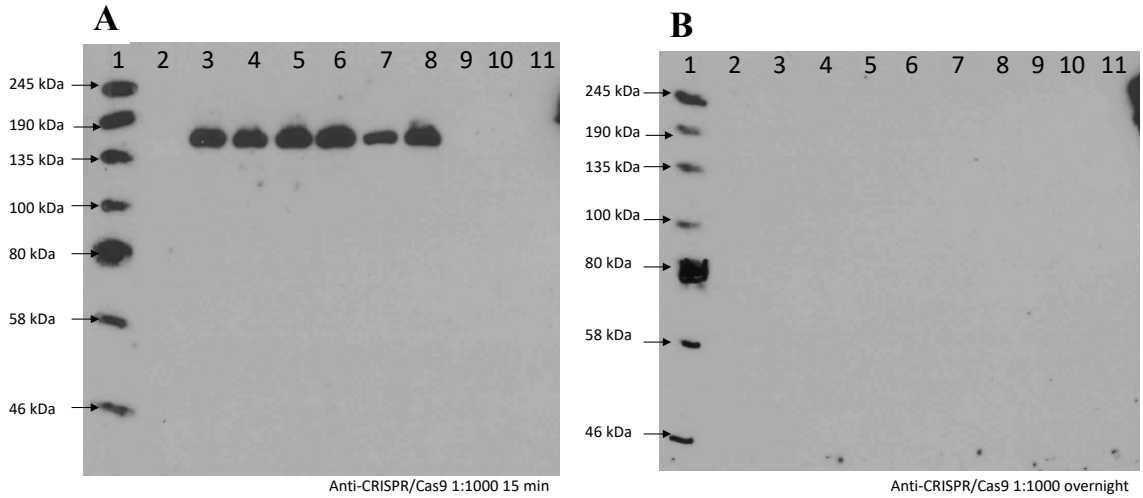


Figure 29. TRV-mediated expression of *Sp*Cas9 constructs in *N. tabacum* leaves. Total leaf protein extracts from three biological replicates per construct were analysed using a monoclonal antibody against *Sp*Cas9 (1:1000). The order of both immunoblots is as follow: Lane 1. Protein Standard Ladder (NEB #P7712). Lane 2. Mock. Lanes 3 - 5. TRV2 *Sp*Cas9 full-length TobFT1 HH. Lanes 6 - 8. TRV2 *Sp*Cas9 full-length TobFT4 HH. Lanes 9 - 11.TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4. **A. Infected leaves**. A band of ~160 kDa was identified in samples where full-length *Sp*Cas9 was delivered. The exposure time was 15 minutes. **B. Systemic leaves**. No *Sp*Cas9 protein was detected in these samples. Exposure time was overnight

It was not possible to detect the presence of the *Sp*Cas9 protein in any systemic leaves of TRV infected plants (Figure 29 B), suggesting that in these biological replicates the viruses spread systemically as shown by the viral symptoms in Figure 23, but the Cas9 protein is not being expressed in those systemic leaves.

Total protein extracts from infected PVX-*Sp*Cas9-RGR leaves were also analysed by Western Blot using a monoclonal antibody specific for *Sp*Cas9. Systemic leaves weren't analysed since no evident systemic symptoms of infection were seen (Figure 24). As presented in Figure 30 A, in contrast with TRV, no band representing the full-length 160kDa *Sp*Cas9 protein was detected in infected leaf samples (Lanes 2 – 4 for *Sp*Cas9 full-length TobFT1 HH, and lanes 5 – 7 for *Sp*Cas9 full-length TobFT4 HH). Moreover, no re-assembled Cas9 protein or N-terminal domain were detected in any samples infected with PVX expressing the split *Sp*Cas9 protein (Lanes 8-10). However, non-specific bands at ~55 kDa and ~34 kDa are observed in most of the samples.

The presence of re-assembled *Sp*Cas9 was assessed in infected leaves where the split domains were co-delivered by PVX and TRV vectors. Consistent with what was previously observed with TRV and PVX vectors alone, no full-length protein or N-terminal domain were detected in any of the analysed samples, even if the membrane is exposed overnight (Data not shown). This result is also independent of the combination tested (Figure 30 B).

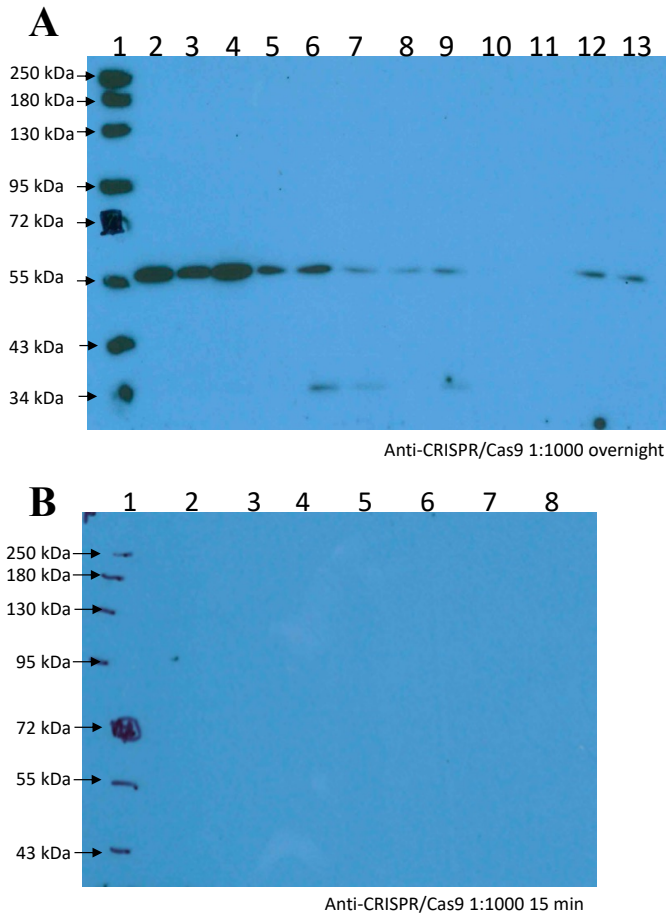


Figure 30. Expression of *Sp*Cas9 constructs in infected *N. tabacum* leaves with PVX or a combination of TRV and PVX. Total leaf protein extracts from three biological replicates per construct were analysed using a monoclonal antibody against *Sp*Cas9 (1:1000), but no protein is detected.

A. PVX-mediated expression of *Sp*Cas9 constructs. Non-specific bands of ~55 kDa and ~34 kDa are seen in some samples. Lane 1. Protein Standard Ladder (NEB #P7719). Lanes 2 - 4. PVX *Sp*Cas9 full-length TobFT1 HH. Lanes 5 - 7. PVX *Sp*Cas9 full-length TobFT4 HH. Lanes 8 - 10. PVX N-*Sp*Cas9 N-intein TobFT1 HH + PVX C-intein C-*Sp*Cas9 TobFT4. Lane 11 - 13. Mock controls. Exposure time was overnight.

B. Co-delivery of split *Sp*Cas9 constructs by PVX and TRV viral vectors. Lane 1. Protein Standard Ladder (NEB #P7719). Lanes 2 - 4. TRV N-*Sp*Cas9 N-intein TobFT1 HH and PVX C-intein-C-*Sp*Cas9 TobFT4 HH replicates. Lanes 5 - 7. PVX N-*Sp*Cas9 N-intein TobFT1 HH and TRV C-intein-C-*Sp*Cas9 TobFT4 HH replicates. Lanes 8. Mock control. Exposure time 15 min.

In summary, TRV viral systemic symptoms and expression of CRISPR/Cas9 components as a result of TRV infection can be seen in some *N. tabacum* plants where the CRISPR/Cas9 components were delivered by this virus, or a combination of TRV and PVX. *SpCas9* RNA is detectable in most of the leaves directly infected with PVX or TRV, while Cas9 protein was found only in leaves where the full-length protein was delivered by TRV. However, no Cas9 protein or Cas9 mRNA can be detected in systemic samples, even when a split version of *Sp*Cas9 was tested.

4.2.2.3 Evaluation of *in vivo* activity of self-cleavage ribozymes fused to *NtFT4* sgRNAs in infected *N. tabacum* leaves

To assess the self-cleavage activity of the HHRz from the Cas9-ribozyme-sgRNA-ribozyme unit, cRT-PCR was performed. Total RNA was extracted from three biological replicates of plants infected with TRV *Sp*Cas9 full-length TobFT1 HH and TRV *Sp*Cas9 full-length TobFT4 HH, and then this was circularized using T4 RNA ligase. cDNA was synthesized and RT-PCR was carried out using abutting primers for the gRNA and the scaffold RNA. Figure 31 shows that it is possible to detect both mature sgRNAs designed to target the *NtFT4* gene. The RNA scaffolds are depicted as blue blocks, HHRz are illustrated as green blocks, while guide gRNAs are highlighted as orange blocks.

In the case of TobFT1 (Figure 31 A), up to 3 bp of the 3'-HHRz scar were found. In the other hand, it was possible to identify up to 16 bp of the 43 bp 5'HHRz. Mostly complete RNA scaffold sequences of 76 bp were identified, however four trimmed RNA scaffold sequences of 58 bp and 73 bp were also found.

For TobFT4 (Figure 31 B), most of the sequences show 1 bp of the 3'HHRz scar, except one sample where all 6 bp can be detected. Moreover, up to 29 bp from the 5'HHRz were identified. Full-length and trimmed RNA scaffold of 61 bp, 63 bp, 71 bp and 72 bp were found. One sequence (P3.1) exhibited an insertion between the 3'- and 5'-HHRz. BLAST analysis indicates that it corresponds to 76 bp of the *coat protein* gene in TRV2.

Complete scaffold sequences were obtained in most of the samples analysed for both sgRNAs, but as mentioned in the section 3.2.2 of CHAPTER 3, trimmed sequences might be due to RNA degradation during sample preparation (101).

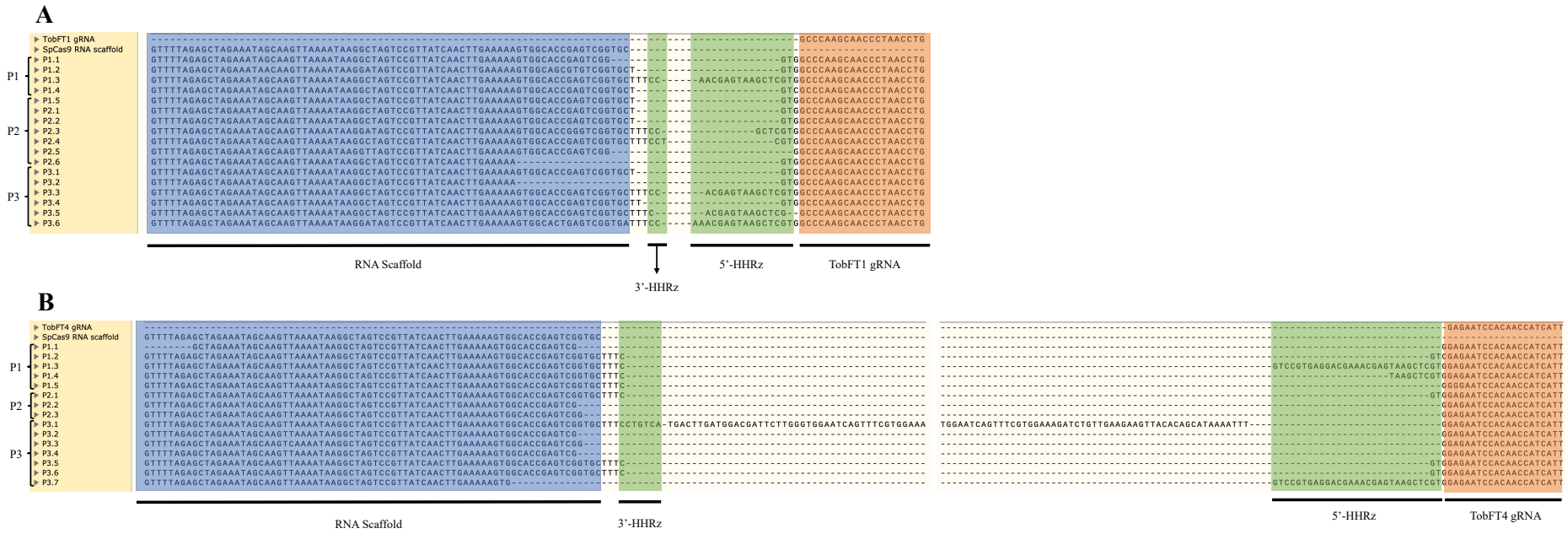


Figure 31. Evaluation of the *in vivo* self-cleavage activity of the HHRz. cRT-PCR analysis was conducted on samples from leaves infected with TRV *Sp*Cas9 full-length HHRz-TobFT1 or TobFT4-HHRz. PCR products were cloned, and sequences aligned with the scaffold and appropriate guide RNA. An extra G between the 5'HHRz and guide RNA was added during PCR step. Blue block: Scaffold RNA. Green blocks: scars from the excision of the 3'-HHRz or 5'-HHRz. Orange block: guide RNA. P1, P2 or P3 correspond to each biological replicate analysed. **A.** TRV *Sp*Cas9 full-length TobFT1 HH. **B.** TRV *Sp*Cas9 full-length TobFT4 HH. The alignment was divided due to its length.

In summary, mature sgRNAs are released from the polycistronic unit *Sp*Cas9 full-length-ribozyme-sgRNA-ribozyme delivered by TRV, indicating biological activity of the self-cleavage ribozymes.

4.2.2.4 Assessment of genome editing activity in tobacco plants infected with TRV full-length or split *Sp*Cas9 -RGR constructs

As presented previously, Cas9 protein and mRNA plus sgRNA were detected in three biological replicates of infected leaves with TRV *Sp*Cas9 full-length TobFT1 or TobFT4. Cas9 RNA was identified in infected leaves with PXV and the same constructs and split *Sp*Cas9-RGR delivered either by PVX, TRV, or a combination of both. Evidence for genome editing was thus evaluated in all samples where both *Cas9* mRNA and protein were found, while in the case of infected leaves where only *Cas9* RNA was observed, only a few samples were selected for analysis.

DNA was extracted from infected leaves and the target region in *NtFT4* gene (exon 2, or exon 4) was amplified. Initially mutations were screened by cloning the PCR product of the target site and subsequent sequencing of individual colonies. Figure 32 illustrates that for TRV *Sp*Cas9 full-length TobFT1 HH replicate 1 (Panel A), no mutations were detected in 73 colonies analysed. On the other hand, for TRV *Sp*Cas9 full-length TobFT4 HH replicate 1 (Panel B), 48 colonies were screened, of which two sequences exhibited an insertion of 1 bp (depicted in red) and four sequences displayed SNPs within the target site, but not at the expected cut site of Cas9 (3 bp upstream of the PAM site, SNPs are illustrated in italic red).

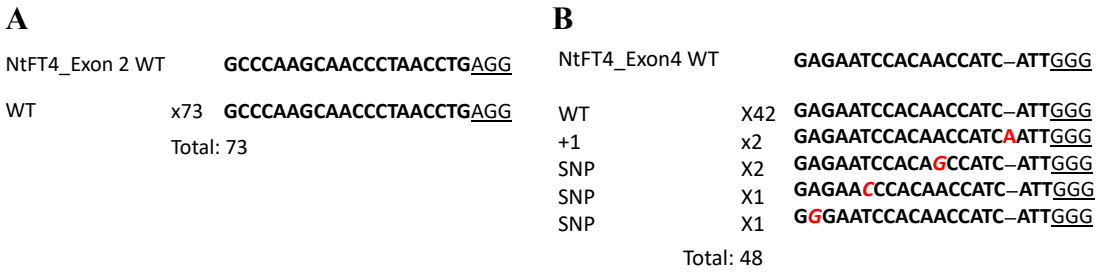


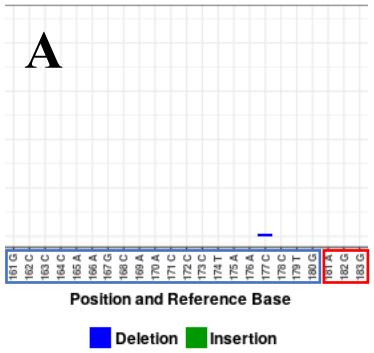
Figure 32. Evaluation of mutations in TRV infected leaves by colony PCR. Alignment of sequences from individual colonies. **A.** TRV *Sp*Cas9 full-length TobFT1 HH, replicate 1. 73 colonies were screened where only WT sequences were found. **B.** TRV *Sp*Cas9 full-length TobFT4 HH, replicate 1. 48 colonies were studied, two sequences showed an insertion of 1 bp, while 4 sequences had SNPs within the target sequence.

Since the screening of genome editing by colony PCR is laborious and its sensitivity isn't high enough to find low levels of mutations, a new approach was adopted which was next generation sequencing (NGS) of pooled products from PCR of the targeted sites. Purified PCR samples were sent for analysis and the service was performed by GENEWIZ. By this method it is possible to detect the introduction of insertions, deletions, and SNPs, however as Cas9 usually breaks the dsDNA 3-4 bp upstream of the PAM site and induces the NHEJ repairing pathway, only indels were considered as evidence of genome editing activity.

Figure 33 shows the sequencing results of infected leaves with TRV *Sp*Cas9 full-length TobFT1 HH. In the three biological replicates genome editing events were detectable, at rates of 0.42% (plant 1), 2.75% (plant 2) and 1.27% (plant 3). As explained previously SNPs weren't considered, which explains why the sum of WT and indels events isn't 100%.

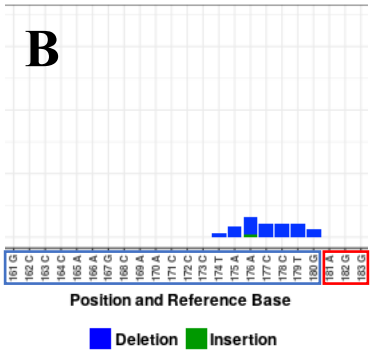
Figure 34 illustrates the sequencing results from leaves infected with TRV *Sp*Cas9 full-length TobFT4 HH. Analysis of genome editing in *NtFT4* exon 4 shows that the three biological replicates display mutations in the target site, at frequencies of 3.03% for replicate 1, 1.27% for replicate 2, and 1.13% for replicate 3. Hence, by this analysis it was possible to detect mutations in both target sites of *NtFT4* gene, in most cases at rates >1%.

Due to the high sensitivity of the sequencing technique, infected leaves with both halves of split *Sp*Cas9 delivered by TRV virus were subsequently analysed. In this samples, only *Cas9* mRNA was detected. DNA was extracted from three biological replicates of tobacco plants co-infected with TRV N-*Sp*Cas9 N-intein and TRV C-intein C-*Sp*Cas9, the *NTFT4* target sites were amplified, and samples were sent for NGS analysis. As shown in Figure 35, mutations were introduced successfully in the exon 4 of replicate 1, and exon 2 of replicate 3, indicating biological activity of the split *Sp*Cas9 system delivered by TRV vectors.



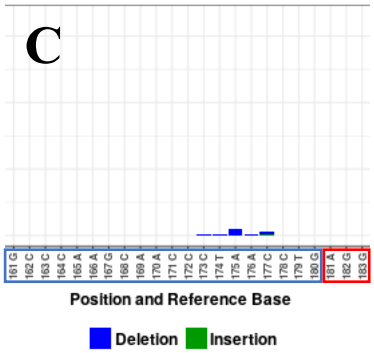
	#Reads	%	
WT	98056	45.28%	GCCCAAGCAACCCTAACCTGAGG
-1	893	0.41%	GCCCAAGCAACCCTAA–CTGAGG
-4	11	0.01%	GCCCAAGCAACCCTA----GAGG

0.42% gene editing



	#Reads	%	
WT	137414	55.45%	GCCCAAGCAACCCTAAC–CTGAGG
-6	2402	0.97%	GCCCAAGCAACCCT------AGG
-4	2005	0.81%	GCCCAAGCAACCCTA---GAGG
-3	1302	0.53%	GCCCAAGCAACCC---C–CTGAGG
+1	1096	0.44%	GCCCAAGCAACCCTAACCTGAGG

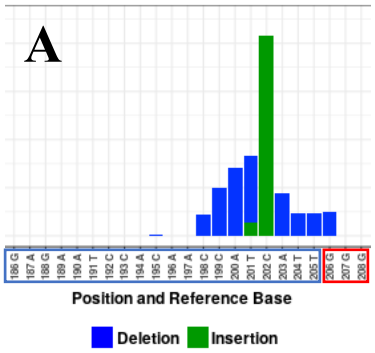
2.75% gene editing



	#Reads	%	
WT	133245	55.07%	GCCCAAGCAACCCTAAC–CTGAGG
-1	1736	0.72%	GCCCAAGCAACCCTA–C–CTGAGG
-1	911	0.38%	GCCCAAGCAACCCTAA--CTGAGG
+1	212	0.09%	GCCCAAGCAACCCTAACACTGAGG
-4	205	0.08%	GCCCAAGCAACCC-----CTGAGG

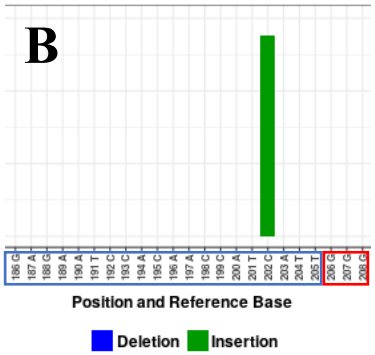
1.27% gene editing

Figure 33. Assessment of genome editing activity in leaves infected with TRV *Sp*Cas9 full-length TobFT1 HH. The editing in the target site was analysed by NGS. Insertions are depicted in red, while deletions are shown as dashes. The number of reads for each type of mutation and its frequency is shown. **A. Biological replicate 1**. Deletions were found, corresponding to the 0.42% of total reads. **B. Biological replicate 2**. Indels are the 2.75% of total reads. **C. Biological replicate 3**. Indels are found in this sample, being the 1.27% of total reads.



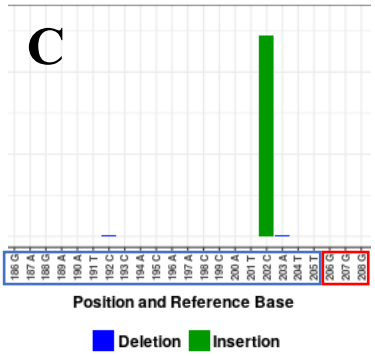
	#Reads	%	
WT	2273336	90.77%	GAGAATCCACAACCATC–ATTGGG
+1	41321	1.65%	GAGAATCCACAACCATCAATTGGG
+1	12062	0.48%	GAGAATCCACAACCATCTATTGGG
-4	5126	0.20%	GAGAATCCACAAC----–ATTGGG
-3	5078	0.20%	GAGAATCCACAACC---–ATTGGG
+1	4890	0.20%	GAGAATCCACAACCATCCATTGGG
-5	3898	0.16%	GAGAATCCACAACCAT-----–GG
-4	3549	0.14%	GAGAATCCACAACC----–TTGGG

3.03% gene editing



	#Reads	%	
WT	242534	90.59%	GAGAATCCACAACCATC–ATTGGG
+1	3409	1.27%	GAGAATCCACAACCATCAATTGGG

1.27% gene editing



	#Reads	%	
WT	287158	90.64%	GAGAATCCACAACCATC–ATTGGG
+1	3574	1.13%	GAGAATCCACAACCATCAATTGGG

1.13% gene editing

Figure 34. Assessment of genome editing activity in infected leaves with TRV *Sp*Cas9 full-length TobFT4 HH. The editing in the target site was analysed by NGS. Insertions are depicted in red, while deletions are shown as dashes. The amount of reads per type of mutation and its frequency is shown. **A. Biological replicate 1**. Indels were found, corresponding to the 3.03% of total reads. **B. Biological replicate 2**. Indels are the 1.27% of total reads. **C. Biological replicate 3**. Indels are found in this sample, being the 1.13% of total reads.

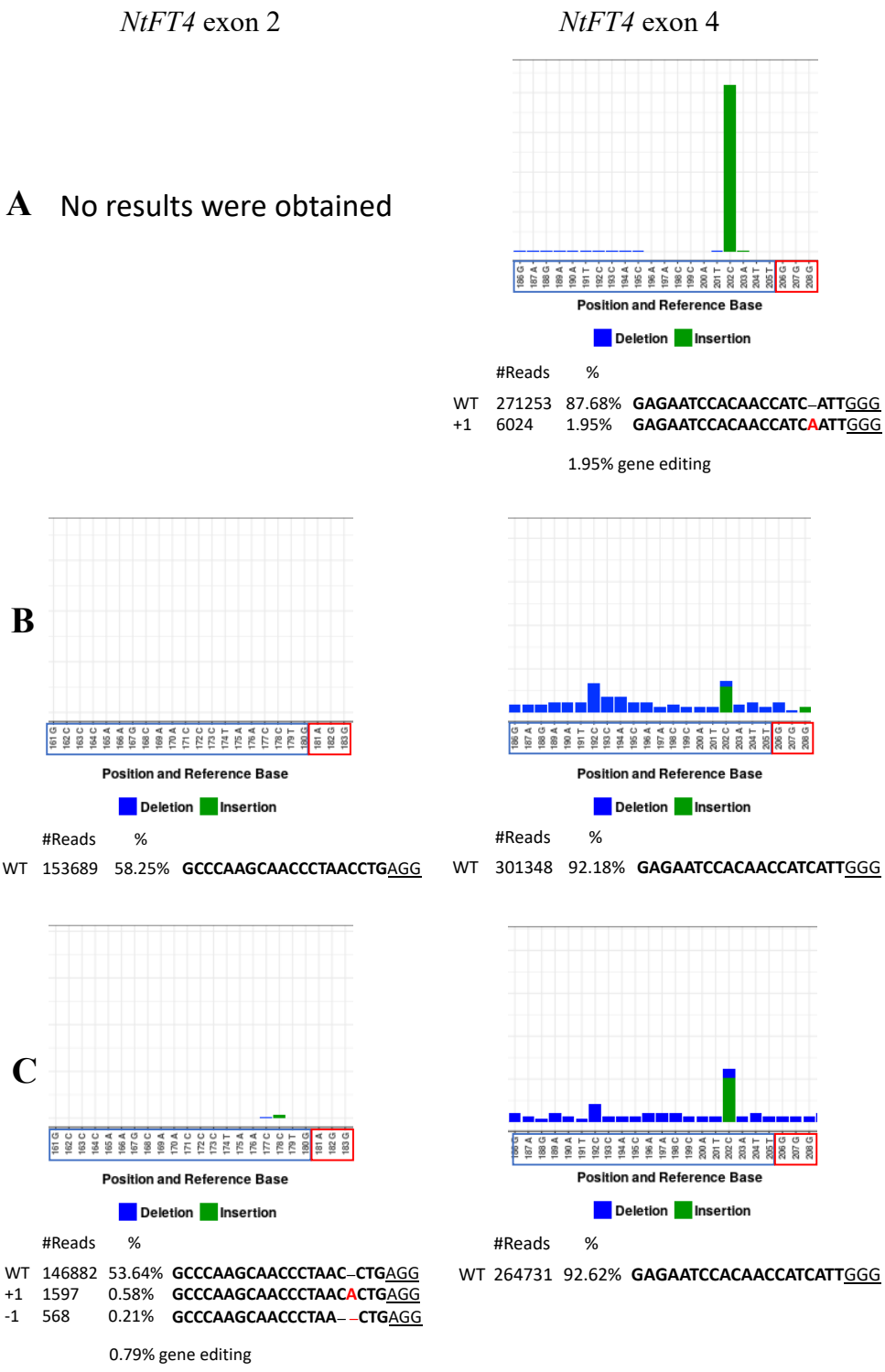


Figure 35. Genome editing events in leaves co-infected with TRV N-*Sp*Cas9 N-intein TobFT1 HH and TRV C-intein C-*Sp*Cas9 TobFT4 HH. *NtFT4* exon 2 target site is shown at the left, *NtFT4* exon 4 target site is illustrated at the right. The editing in the target site was analysed by NGS. Insertions are depicted in red, while deletions are shown as dashes. The number of reads per type of mutation and its frequency is shown. **A. Biological replicate 1**. No results were obtained for exon 2. For exon 4 a single 1 bp insertion was found at a rate of 1.95%. **B. Biological replicate 2**. No mutations were detected in either exon 2 or exon 4. **C. Biological replicate 3**. For exon 2, an insertion and a deletion were seen with a total frequency of 0.79%, for exon 4 no genome edits were found.

For replicate 1, no results were obtained for *NtFT4* exon 2 target site because the concentration of PCR product didn't meet the requirements for the analysis. For exon 4, an insertion of 1 bp was detected with a frequency of 1.95%. In the case of replicate 2, no genome editing events were found either in exon 2 or exon 4. On the other hand, when replicate 3 was analysed, an insertion and deletion was found in exon 2 at a total rate of 0.79%, while for exon 4 no genome editing was found.

In some cases, as for example in exon 4 of replicates 2 and 3, the percentage of mutated reads don't exceed 0.02% which is considered a basal noise signal and so these sequences are not shown.

Finally, the genome editing in a systemic sample from replicate 1 of TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4 HH plant was assessed by NGS. As shown in Figure 36, no mutations were detected in either of the target sites analysed. This result correlates with the absence of *Sp*Cas9 protein in these leaves, even though viral symptoms and *Cas9* mRNA in some samples were observed.

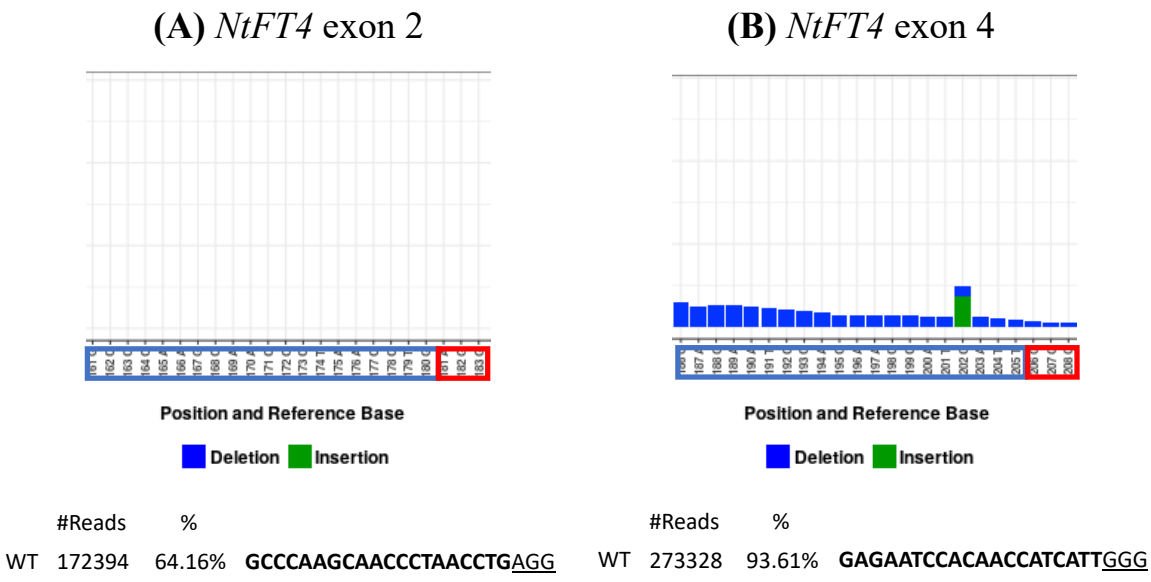


Figure 36. Genome editing events in systemic leaves of a plant co-infected with TRV N-*Sp*Cas9 N-intein TobFT1 HH and TRV C-intein C-*Sp*Cas9 TobFT4 HH. Genome editing was assessed in replicate 1. *NtFT4* exon 2 target site is shown at the left, while mutations in *NtFT4* exon 4 are illustrated at the right. The editing in the target site was analysed by NGS. Insertions are depicted in red, while deletions are shown as dashes. The number of reads per type of mutation and its frequency is shown. **A. *NtFT4* Exon 2 target site**. **B. *NtFT4* exon 4 target site**. No genome editing events were detected at either site.

4.2.2.5 Flowering time of plants infected with *Sp*Cas9-RGR constructs delivered by TRV and PVX virus vectors

As described by Harig *et al., NtFT4* gene is a floral inducer, thus mutation of this gene should create a delay in the flowering time (159). Since *N. tabacum* var. Maryland Mammoth flowers exclusively under SD conditions, infected plants were grown in SD conditions to promote flowering for approximately three months.

Compared with the WT mock control, a delay in flowering was observed in TRV *Sp*Cas9 full-length TobFT1 HH replicate 3 (Figure 37A) and TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4 HH replicate 2 (Figure 37B).

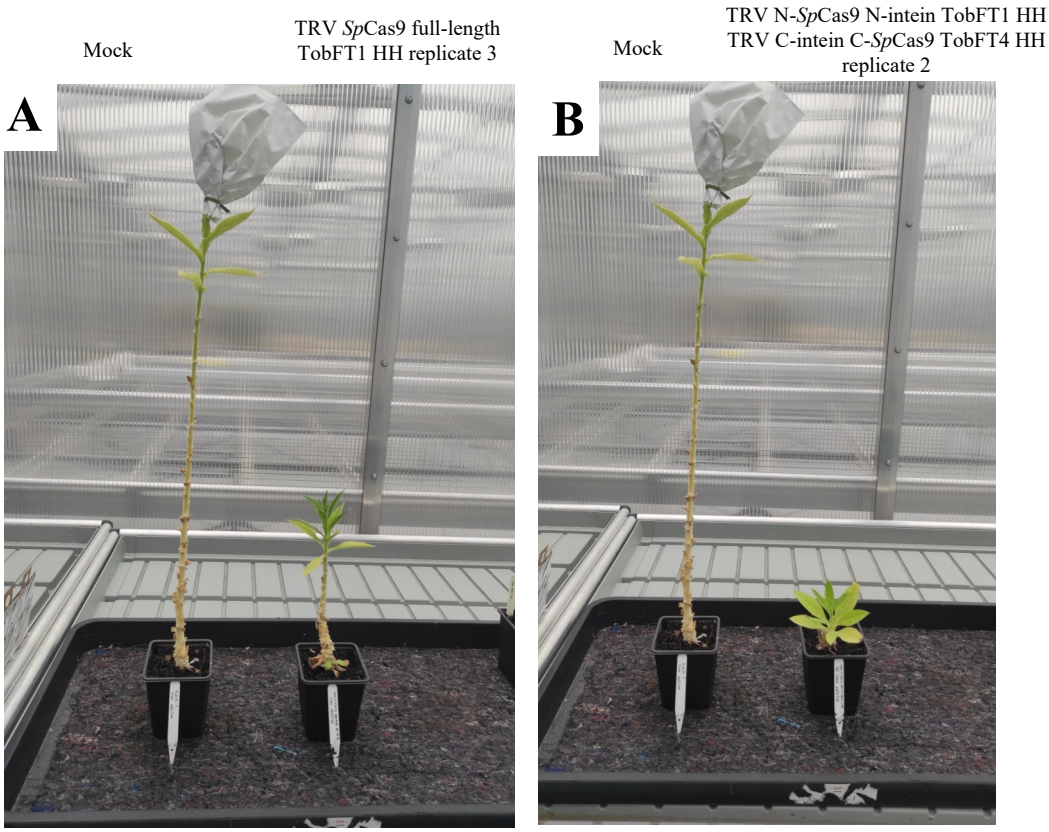


Figure 37. Flowering time of plants infected with TRV *Sp*Cas9-RGR constructs. Two plants displayed delayed flowering (right hand on each picture) compared with WT mock controls (Left hand in each picture) in SD conditions. **A**. TRV *Sp*Cas9 full-length TobFT1 HH replicate 3. **B.** TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4 HH replicate 2.

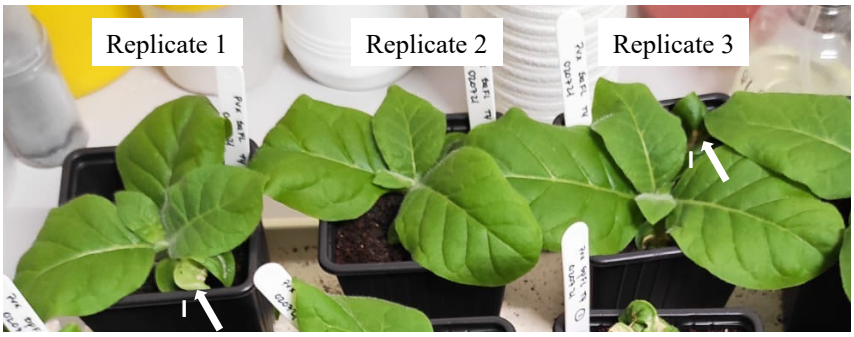
Systemic symptoms of viral infection were noticed in TRV *Sp*Cas9 full-length TobFT1 HH replicate 3 (Figure 23), whilst Cas9 protein and mRNA were detected in infected leaves, but not systemically (Figure 26). A genome editing rate of 1.27% was determined in infected leaves. In the case of TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4 HH replicate 2, RT-PCR analysis showed the presence of systemic TRV virus spread (Figure 26, lane 11), however genome editing analysis of infected leaves showed only WT sequences at both target sites. This plant in particular exhibited strong systemic symptoms of TRV infection which might themselves have caused the observed delay in flowering.

4.2.3 Delivery of the *Sa*Cas9 - RGR constructs into *N. tabacum* plants by TRV and PVX viral vectors

A smaller Cas9 (*Sa*Cas9) and its split version was delivered into *N. tabacum* plants by TRV and PVX vectors. 3-4 weeks old tobacco plants were infected with PVX and TRV vectors expressing *Sa*Cas9 full-length TobFT1 or TobFT4 sgRNAs, or a combination of *Sa*Cas9 739N TobFT1 and *Sa*Cas9 740C TobFT4. Three independent plants (biological replicates) were infected per virus construct. Three independent plants were inoculated with water only, corresponding to mock controls (WT). Leaf samples were taken at 14 dpi for RNA, protein, and genome editing analysis.

Figure 38 illustrates an example of tobacco plants infected with PVX *Sa*Cas9 RGR, where systemic symptoms of infection, such as chlorosis or mottled leaves, were not evident.

(A) PVX *Sa*Cas9 full-length TobFT1 HH



(B) PVX *Sa*Cas9 full-length TobFT4 HH

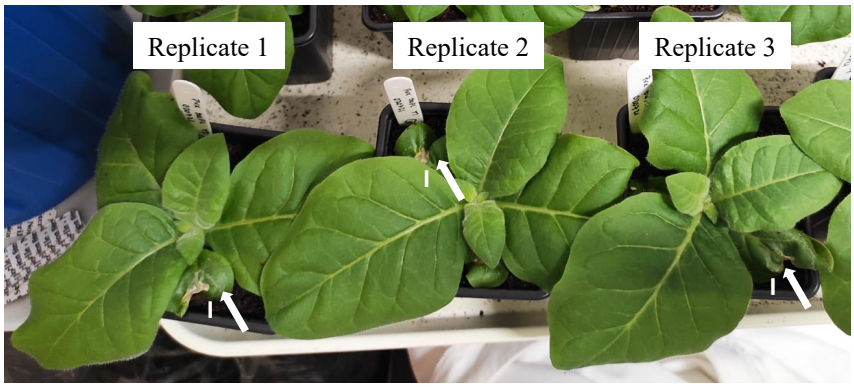


Figure 38. Tobacco plants infected with PVX *Sa*Cas9 constructs. No clear systemic symptoms of viral spread, as chlorosis or mottle leaves, were observed**. A.** PVX *Sa*Cas9 full-length TobFT1 HH replicates 1,2 and 3. **B.** PVX *Sa*Cas9 full-length TobFT4 HH replicates 1, 2 and 3. Infected leaves are indicated with white arrows.

On the other hand, Figure 39 shows tobacco plants infected with TRV *Sa*Cas9 full-length TobFT1 HH and TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH. Clear symptoms of systemic TRV infection can be seen in these plants, such as curled leaves and necrosis through the main vein. In some cases, the effect of the infection is so severe that it distorts the shape of the leaf (As for example TRV split *Sa*Cas9 replicate 2).

Moreover, tobacco plants were also infected with a mixture of PVX and TRV virus expressing the split *Sa*Cas9 constructs, however pictures of infection of these plants were not taken. As stated before, infected, and systemic samples were collected 14 dpi for further analysis.

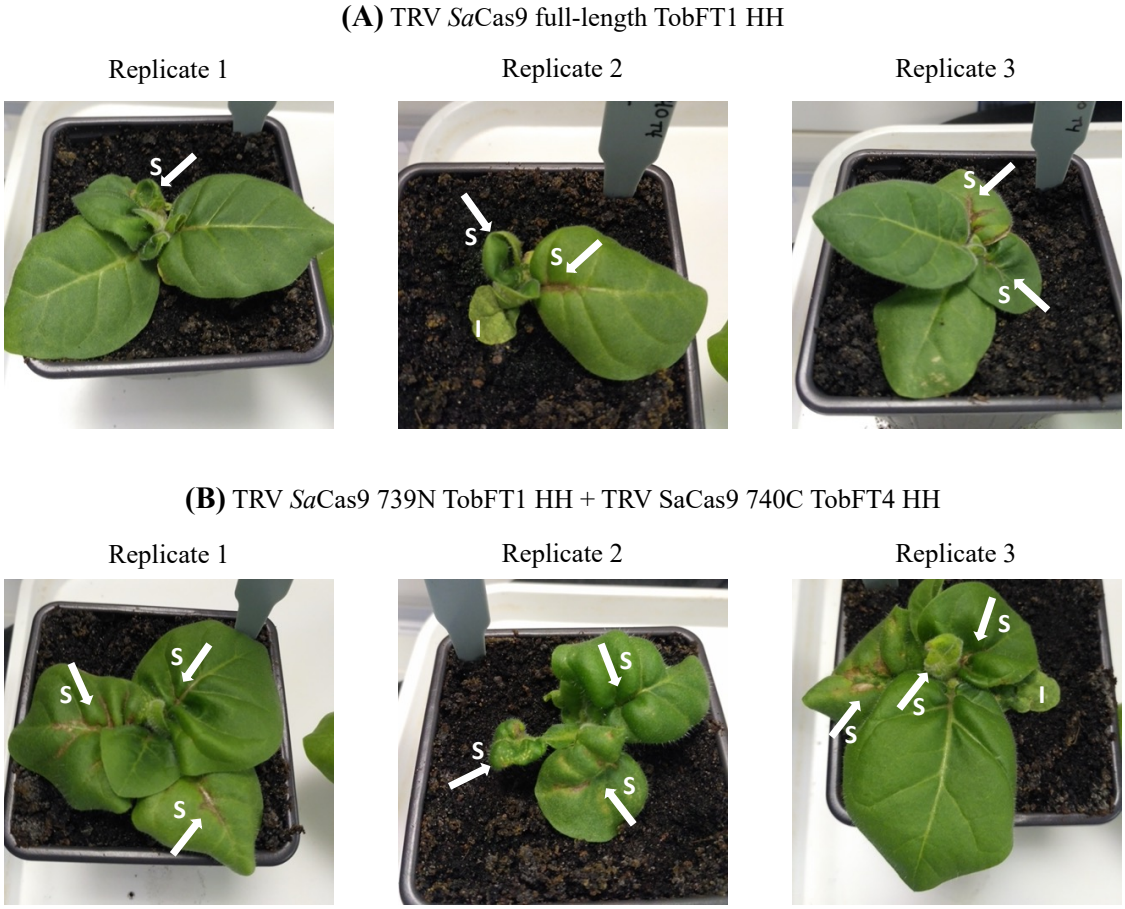


Figure 39. *N. tabacum* plants infected with TRV *Sa*Cas9 constructs. Clear systemic symptoms, as curled leaves and necrosis can be seen in these plants and are pointed out with white arrows. A. TRV *Sa*Cas9 full-length TobFT1 HH replicates 1, 2 and 3. B. TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH replicates 1, 2 and 3. (S) Systemic leaves. (I) Infected leaves.

4.2.3.1 Detection of *Sa*Cas9 mRNA in TRV/PVX infected and systemic *N. tabacum* leaves

The detection of *SaCas9* mRNA transcripts was carried out by RT-PCR in the infected and systemic leaves of plants where the transgene was delivered by either TRV, PVX, or a combination of those viruses. Total RNA was extracted and used as a template for cDNA synthesis. Specific primers for *SaCas9* N- and C- terminal sequences were used to test its expression. The presence of viral RNA in these samples was also tested. The amplification of the *EF1α* housekeeping gene was carried out to check the quality of the RNA obtained. Appropriate PCR controls were included. Positive controls of amplification correspond to plasmids harbouring the different constructs or empty viral plasmids. Negative controls of amplification correspond to no RNA for cDNA synthesis (RT-PCR negative control) or no cDNA (PCR negative control) to discard the presence of contaminants. Raw images of the agarose gels are shown in Appendix A4, supplementary figures 11, 12 and 13.

Using primers to amplify a region of 372 bp of the *EF1α* gene, strong and even expression of the *EF1α* gene was observed in all infected, systemic and WT mock leaf samples (Figures 40, 41 and 42), indicating that RNA of good quality had been obtained. The expression of *SaCas9* was evaluated using specific primers for its N- and C- terminal domain sequences. In plants infected with TRV *Sa*Cas9 constructs, strong bands can be perceived in samples from infected leaves (Figure 40, lanes 3-5 for full-length *Sa*Cas9, and lanes 9-11 for the split version). When systemic leaves were analysed, weak bands can be seen in samples where the *Sa*Cas9 full-length was delivered (Figure 40, lanes 6 to 8 for both termini), in contrast to what was observed for *Sp*Cas9 (Figure 26, panel B). In comparison to *SpCas9,* a strong band was observed in those samples where the C-terminal of the split *SaCas9* gene was delivered by a TRV vector, while a weakier band was detected for the N-terminal domain (in both cases lanes 12 to 14, Figure 40). Moreover, it was possible to detect high levels of *TRV1* RNA in infected and systemic samples whilst the levels of *TRV2* RNA was low in infected and systemic samples where the full-length *SaCas9* gene was delivered (Figure 40, lanes 4-6 infected leaves, lanes 7-9 systemic leaves), although strong TRV2 bands can be seen in infected and systemic samples where the split *Sa*Cas9 was delivered (Figure 40, lanes 10-12 infected leaves, lanes 13-15 systemic leaves).

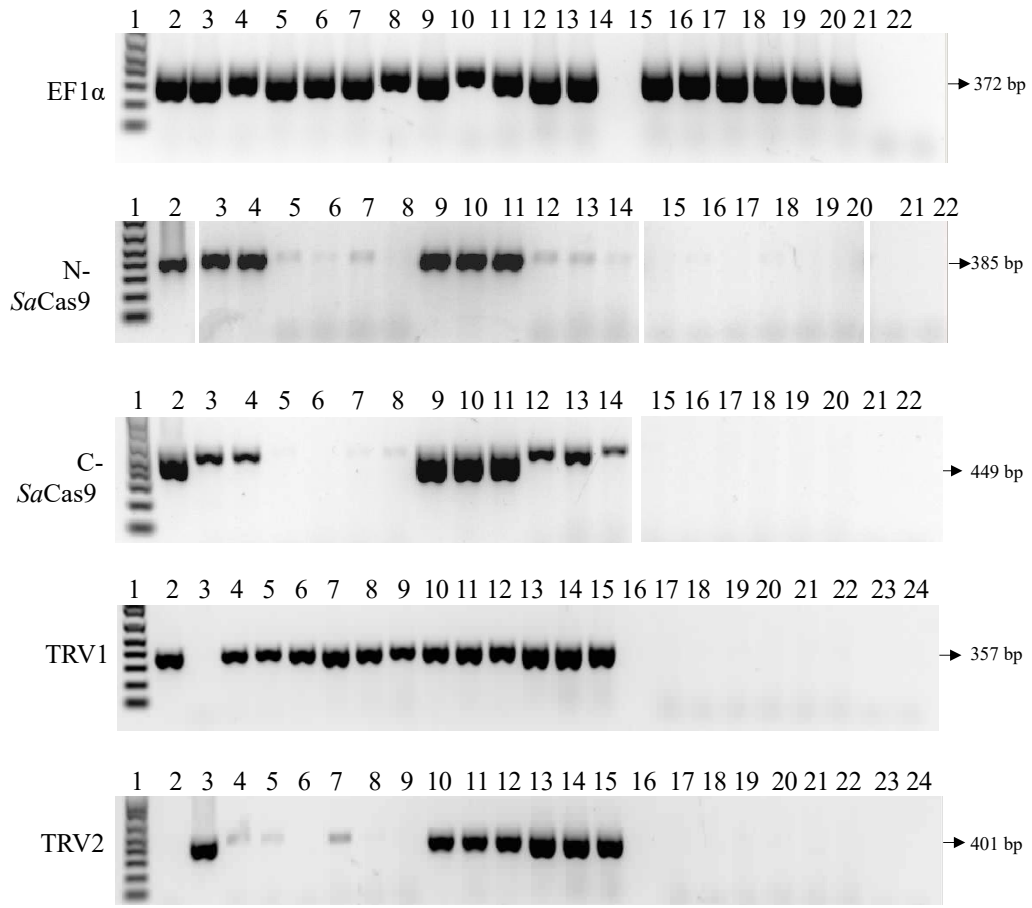


Figure 40. Detection of *SaCas9* mRNA in TRV infected and systemic *N. tabacum* leaves. Three biological replicates per construct were analysed. PCR product size is indicated in each image. EF1α samples were loaded as following. Lane 1. 1 kb plus DNA ladder. Lanes 2 – 4. TRV *Sa*Cas9 full-length TobFT1 HH infected leaves. Lanes 5 – 7. TRV *Sa*Cas9 full-length TobFT1 HH systemic leaves. Lanes 8 – 10. TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH infected leaves. Lanes 11 – 13. TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH systemic leaves. Lane 14. Empty. Lanes 15 – 17. Mock controls inoculated leaves. Lanes 18 – 20. Mock controls systemic leaves. Lane 21. RT-PCR negative control. Lane 22. PCR negative control. *Sa*Cas9 N- and C- terminal samples are as following: Lane 1. 1 kb plus DNA ladder. Lane 2. Positive control. Lanes 3 – 5. TRV *Sa*Cas9 full-length TobFT1 HH infected leaves. Lanes 6 – 8. TRV *Sa*Cas9 full-length TobFT1 HH systemic leaves. Lanes 9 – 11. TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH infected leaves. Lanes 12 – 14. TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH systemic leaves. Lanes 15 – 17. Mock controls inoculated leaves. Lanes 18 – 20. Mock controls systemic leaves. Lane 21. RT-PCR negative control. Lane 22. PCR negative control. TRV1 and TRV2 loading order is Lane 1. 1 kb plus DNA ladder. Lane 2. TRV1 positive control. Lane 3. TRV2 positive control. Lanes 4 – 6. TRV *Sa*Cas9 full-length TobFT1 HH infected leaves. Lanes 7 – 9. TRV *Sa*Cas9 full-length TobFT1 HH systemic leaves. Lanes 10 – 12. TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH infected leaves. Lanes 13 – 15. TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH systemic leaves. Lane 16. Empty. Lanes 17 – 19. Mock controls inoculated leaves. Lanes 20 – 22. Mock controls systemic leaves. Lane 23. RT-PCR negative control. Lane 24. PCR negative control.

In RT-PCR analysis of PVX delivered *Sa*Cas9-RGR constructs strong bands can be seen in infected leaves samples for both full-length *SaCas9* and split N- and C-termini of the gene (Figure 41 A). However, no PCR products were detected in systemic leaves (Figure 41 B). The same result is obtained when the presence of *PVX* virus RNA was assessed in these samples.

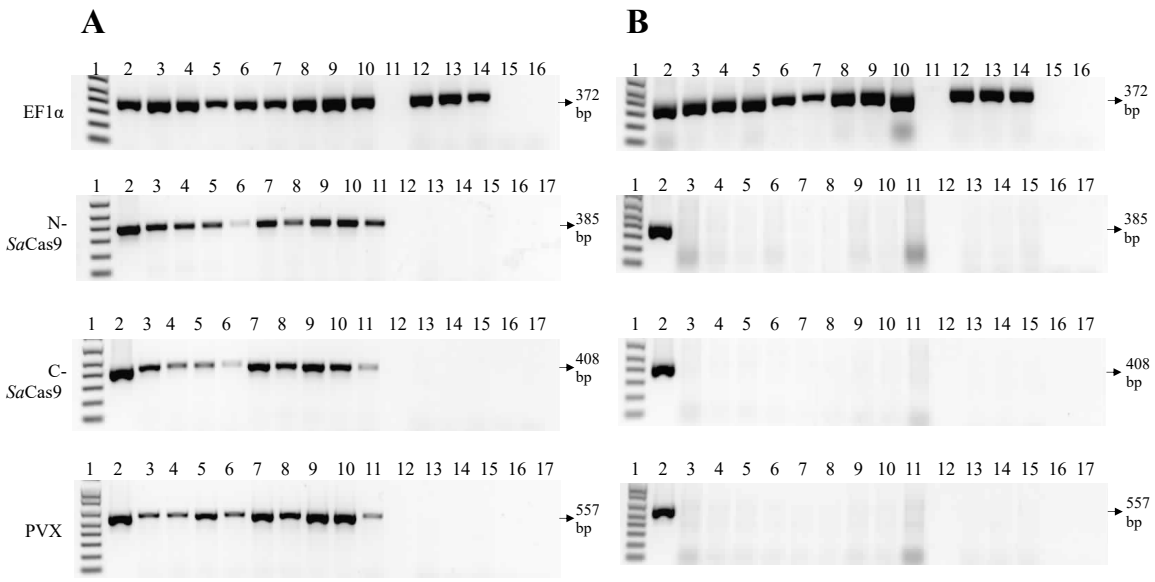


Figure 41. Detection of *SaCas9* mRNA in PVX infected and systemic tobacco leaves. Three biological replicates per construct were analysed. PCR products size is indicated in each image. The order of the samples is the same in both panels. **A.** PVX infected leaves. **B.** PVX systemic leaves.

EF1α samples were loaded as following. Lane 1. 1 kb plus DNA ladder. Lanes 2 – 4. PVX *Sa*Cas9 full-length TobFT1 HH. Lanes 5 – 7. PVX *Sa*Cas9 full-length TobFT4 HH. Lanes 8 – 10. PVX *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH. Lane 11. Empty. Lanes 12 – 13. Mock controls. Lane 15. RT-PCR negative control. Lane 16. PCR negative control.

*Sa*Cas9 N- and C- terminal and PVX samples are as following: Lane 1. 1 kb plus DNA ladder. Lane 2. Positive control. Lanes 3 – 5. PVX *Sa*Cas9 full-length TobFT1 HH. Lanes 6 – 8. PVX *Sa*Cas9 full-length TobFT4 HH. Lanes 9 – 11. PVX *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH. Lane 12. Empty. Lanes 13 – 15. Mock controls. Lane 16. RT-PCR negative control. Lane 17. PCR negative control.

Finally, the expression of the split *Sa*Cas9-RGR constructs was analysed in infected and systemic leaves of plants where these genes were co-delivered by TRV and PVX. Strong expression of the *SaCas9* N- and C- terminal domain was seen in infected leaves, regardless of which domain was delivered by PVX or TRV (Figure 42 A). When the presence of the viral RNA was assessed, strong bands can be observed for both *PVX* and *TRV* in infected leaves (Figure 42 A). In systemic leaves *PVX* can only be detected weakly in some leaves (Figure 42 B, lanes 5 to 7) whilst both *TRV1* and *TRV2* show strong expression in all the systemic samples analysed (Figure 42 B).

It is noteworthy that in systemic leaves of TRV *Sa*Cas9 739N TobFT1 + PVX *Sa*Cas9 740C TobFT4 replicate 3 both *SaCas9* N- and C- terminal domains as well as RNA of both viruses was detectable (Figure 42 B, lane 5 for N-, C- ends and PVX, lane 6 for TRV1 and TRV2). In another instance, expression of just the *SaCas9* C-terminal together with the *PVX*, *TRV1* and *TRV2* RNA was detected in the PVX *Sa*Cas9 739N TobFT1 + TRV *Sa*Cas9 740C TobFT4 replicate 1 (Figure 42 B, lane 6 for N-, C- ends and PVX, lane 7 for TRV1 and TRV2).

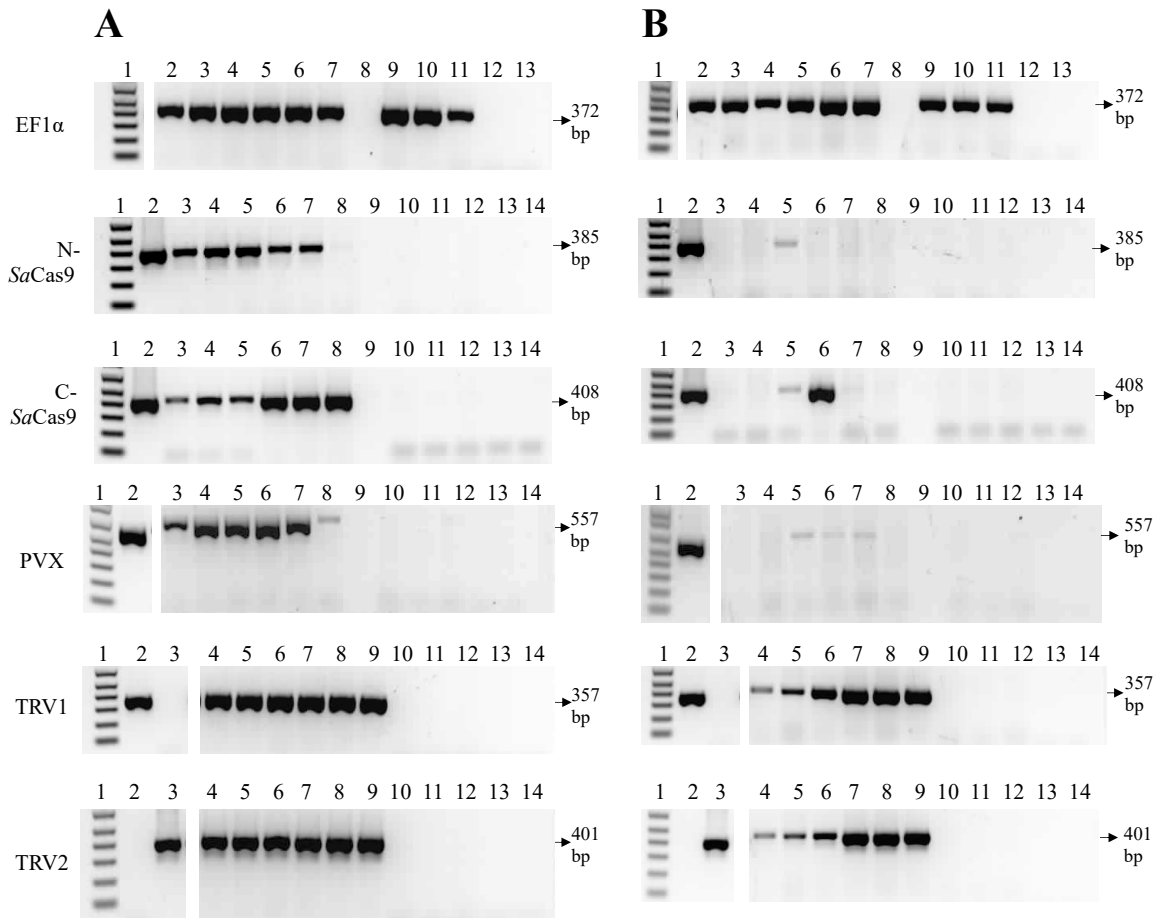


Figure 42. Detection of *SaCas9* mRNA in TRV and PVX co-infected and systemic tobacco leaves. Three biological replicates per construct were analysed. PCR products size is indicated in each image. The order of the samples is the same in both panels. **A.** Infected leaves. **B.** Systemic leaves.

EF1α samples were loaded as following. Lane 1. 1 kb plus DNA ladder. Lanes 2 – 4. TRV *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH. Lanes 5 – 7. PVX *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH. Lane 8. Empty. Lanes 9 - 11. Mock controls. Lane 12. RT-PCR negative control. Lane 13. PCR negative control.

*Sa*Cas9 N- and C- terminal and PVX samples are as following: Lane 1. 1 kb plus DNA ladder. Lane 2. Positive control. Lanes 3 – 5. TRV *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH. Lanes 6 – 8. PVX *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH. Lane 9. Empty. Lanes 10 - 12. Mock controls. Lane 13. RT-PCR negative control. Lane 14. PCR negative control.

TRV1 and TRV2 loading order is Lane 1. 1 kb plus DNA ladder. Lane 2. TRV1 positive control. Lane 3. TRV2 positive control. Lanes 4 – 6. TRV *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH. Lanes 7 – 9. PVX *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH. Lanes 10 - 12. Mock controls. Lane 13. RT-PCR negative control. Lane 14. PCR negative control.

4.2.3.2 Detection of *Sa*Cas9 protein in TRV/PVX infected *N. tabacum* leaves by immunoblotting

*Sa*Cas9 protein expression was tested in infected leaves by Western Blot. The analysis was carried out using two monoclonal antibodies that recognize the N- and C-terminal of *Sa*Cas9 protein. Ponceau red stained membranes showing total transferred proteins are displayed in Appendix A5.

The presence of *Sa*Cas9 protein was tested in leaves directly infected with TRV or PVX full-length or split *Sa*Cas9 constructs, however no protein was detected in any of these samples. In contrast, a strong band at ~130 kDa, corresponding to the size of the full-length protein, was detected in the positive control lane (*Sa*Cas9 delivered by the *Agrobacterium* vector pRI201N). These results are shown in Appendix A5, supplementary figures 17 (TRV) and 18 (PVX)

Furthermore, samples in which the split *Sa*Cas9 ends were delivered by co-infection of TRV and PVX were analysed by immunoblotting. Figure 43 A shows that *Sa*Cas9 N-terminal could not be detected in any of the replicates. However, a band at ~43 kDa, corresponding to *Sa*Cas9 740C domain, can be observed in the samples where this construct was delivered by TRV (lanes 2 -4) or PVX (lanes 5 – 7) (Figure 43 B). *Sa*Cas9 740C – RGR is 1340 bp long, fitting the predicted cargo capacity limit of both viruses. A non-specific band at ~55 kDa is also detected in these samples.

In summary, systemic symptoms of TRV viral infection can be seen in *N. tabacum* plants where this virus was used to deliver full-length or split *Sa*Cas9 fused to *NtFT4* sgRNAs. No systemic symptoms of PVX infection were evident, even when it was used in combination with TRV. *Sa*Cas9 RNA was detected in all infected leaf samples where the full-length gene was delivered either by PVX or TRV. Interestingly, weak systemic expression of *SaCas9* RNA can be seen when it is delivered by the TRV virus even though no *Sa*Cas9 protein was detected in these systemic leaf samples by immunoblot analysis. When split *Sa*Cas9 was delivered by TRV, PVX, or a combination of both, expression of both N- and C- domains is detectable in all infected leaves analysed. Moreover, western blot analysis of TRV and PVX co-infected leaves showed the presence of *Sa*Cas9 C-terminal protein in these samples. Clear expression of both N- and C- terminal sequences can be found in systemic samples from TRV infected plants, however no expression was found in

systemic samples of PVX infected plants. In systemic samples from plants where the split *Sa*Cas9-RGR was delivered by a combination of TRV and PVX, both N- and C- terminal *Sa*Cas9 domains were detected only in one plant replicate.

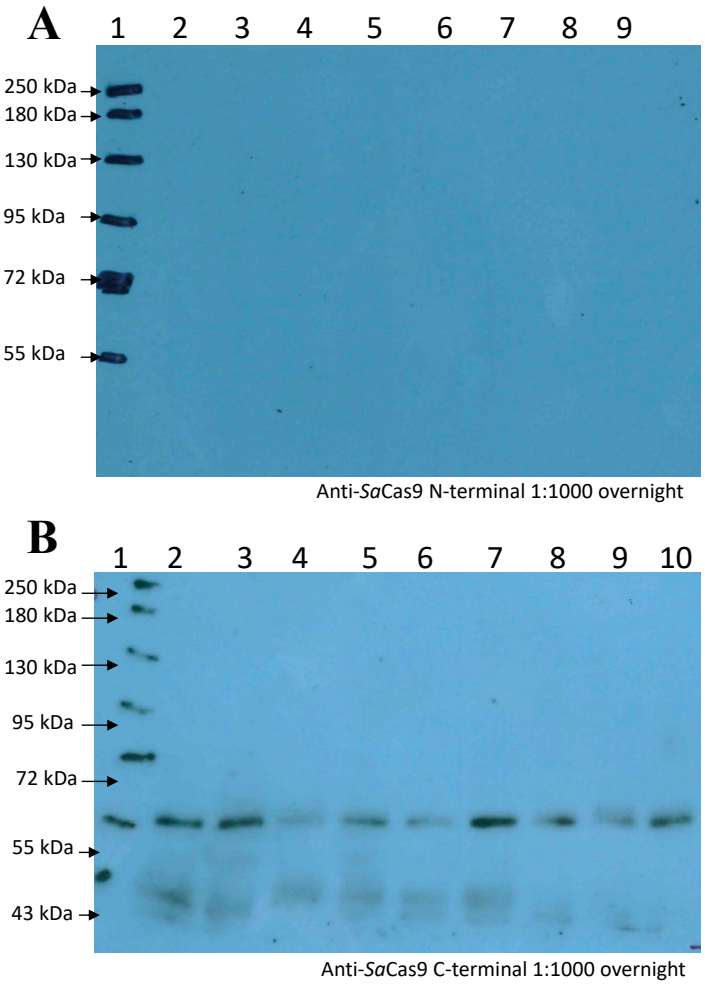


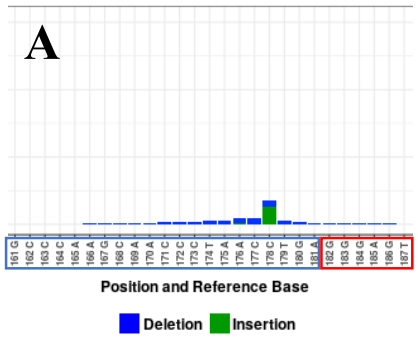
Figure 43. Expression of split *Sa*Cas9 protein in *N. tabacum* leaves co-infected with PVX and TRV virus vectors. Three biological replicates per constructs were tested. **A.** Immunoblot using anti-*Sa*Cas9 N-terminal antibody (1:1000), exposure time overnight. **B.** Immunoblot using anti-*Sa*Cas9 C-terminal antibody (1:1000), exposure time overnight. A band at ~43 kDa, corresponding to *Sa*Cas9 C-terminus is identified in samples where this construct was delivered by TRV (Lanes 2 – 4) or PVX (Lanes 5 – 7). A non-specific band at ~55 kDa can be seen in all tested samples. The order of the lanes is as following: Lane 1. Protein Standard Ladder (NEB #P7719). Lanes 2 - 4. PVX *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH. Lanes 5 - 7. TRV *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH. For panel A. Lane 8. Empty. Lane 9. Mock control, while panel B Lanes 8 – 10 correspond to mock inoculated control plants.

4.2.3.3 Assessment of genome editing activity in tobacco plants infected with TRV full-length or split *Sa*Cas9-RGR constructs

Genome editing was evaluated in tobacco plants where the full-length and split *Sa*Cas9, fused to *NtFT4* sgRNAs was delivered using the TRV vector. DNA was extracted from infected and selected systemic leaves and the target regions in the *NtFT4* gene (Exon 2 or Exon 4) was PCR amplified. Mutations in the target sites were detected by NGS.

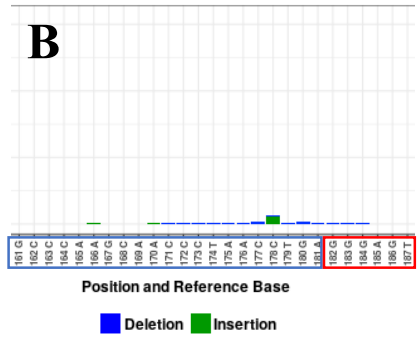
Successful genome editing events were found in leaves infected with TRV *Sa*Cas9 full-length TobFT1 HH. Sequencing results of exon 2 shown that both replicate 1 and replicate 2 exhibit several indels at the target site at a total frequency of 1.06% and 1.33%, respectively (Figure 44 A and B). Small insertions of 1 bp and deletions of up to 15 bp were found in both replicates. In contrast, no mutations were found in the target site of replicate 3. As stated before, Cas9 induces DSB into the target gene inducing the NHEJ repairing pathway where insertion or deletions are mostly generated. Hence SNPs weren't considered as evidence of genome editing, explaining why the sum of frequencies between WT and mutations is not 100%.

In infected leaves were the split *Sa*Cas9-RGR unit was co-delivered by two TRV vectors, mutations were found in both target sites in all of the three biological replicates (Figure 45). The left column of Figure 45 shows that for exon 2, genome editing events were found at total rates of 1.55% for replicate 1, 1.71% for replicate 2 and 1.16% for replicate 3. Interestingly, the right column of Figure 45 shows that for exon 4, an insertion of 1 bp was the only indel found, with total frequencies of 0.78% for replicate 1, 0.29% for replicate 2 and 0.46% for replicate 3.



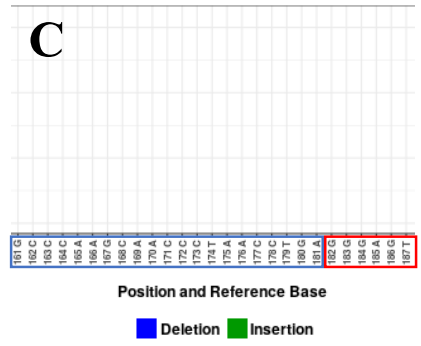
	#Reads	%	
WT	112593	58.60%	GCCCAAGCAACCCTAACC–TGAGGGAGT
+1	726	0.38%	GCCCAAGCAACCCTAACCTTGAGGGAGT
+1	256	0.13%	GCCCAAGCAACCCTAACCATGAGGGAGT
+1	230	0.12%	GCCCAAGCAACCCTAACCGTGAGGGAGT
+1	101	0.05%	GCCCAAGCAACCCTAACCCTGAGGGAGT
-1	89	0.05%	GCCCAAGCAACCCTAAC––TGAGGGAGT
-8	82	0.04%	GCCCAAGCAA–––––––––TGAGGGAGT
-15	81	0.04%	GCCCAAG–––––––––––––––GGAGT
-1	75	0.04%	GCCCAAGCAACCCTAACC––GAGGGAGT
-2	75	0.04%	GCCCAAGCAACCCTAA–––TGAGGGAGT
-4	74	0.04%	GCCCAAGCAACCCT––––TGAGGGAGT
-13	68	0.04%	GCCCAAGCAACCC––––––––––––––T
-2	67	0.03%	GCCCAAGCAACCCTAAC––––GAGGGAGT
-5	58	0.03%	GCCCAAGCAACCC––––––TGAGGGAGT
-5	57	0.03%	GCCCAAGCAACCCTAA––––––GGGAGT

1.06% gene editing



	#Reads	%	
WT	112130	59.84%	GCCCAAGCAACCCTAACC–TGAGGGAGT
+1	993	0.53%	GCCCAAGCAACCCTAACCTTGAGGGAGT
+1	711	0.38%	GCCCAAGCAACCCTAACCATGAGGGAGT
-7	180	0.09%	GCCCAAGCAAC––––––––TGAGGGAGT
-9	139	0.07%	GCCCAAGCAACCCTA–––––––––AGT
-2	135	0.07%	GCCCAAGCAACCCTAACC–––AGGGAGT
-4	127	0.06%	GCCCAAGCAACCCTAACC––––GGAGT
-2	112	0.06%	GCCCAAGCAACCCTAA–––TGAGGGAGT
+1	78	0.04%	GCCCAAGCAACCCTAACCCTGAGGGAGT
+1	47	0.03%	GCCCAAGCAACCCTAACCGTGAGGGAGT

1.33% gene editing



	#Reads	%	
WT	113133	60.0%	GCCCAAGCAACCCTAACCTGAGGGAGT

Figure 44. Assessment of genome editing events in TRV *Sa*Cas9 full-length TobFT1 HH infected leaves. The editing in the target site was analysed by NGS. Insertions are depicted in red, while deletions are shown as dashes. The number of reads per type of mutation and its frequency is shown **A. Biological replicate 1**. Several indels were found with a total frequency of 1.06%. **B. Biological replicate 2**. Indels with a total rate of 1.33% were noticed. **C. Biological replicate 3**. No mutations were observed in this sample.

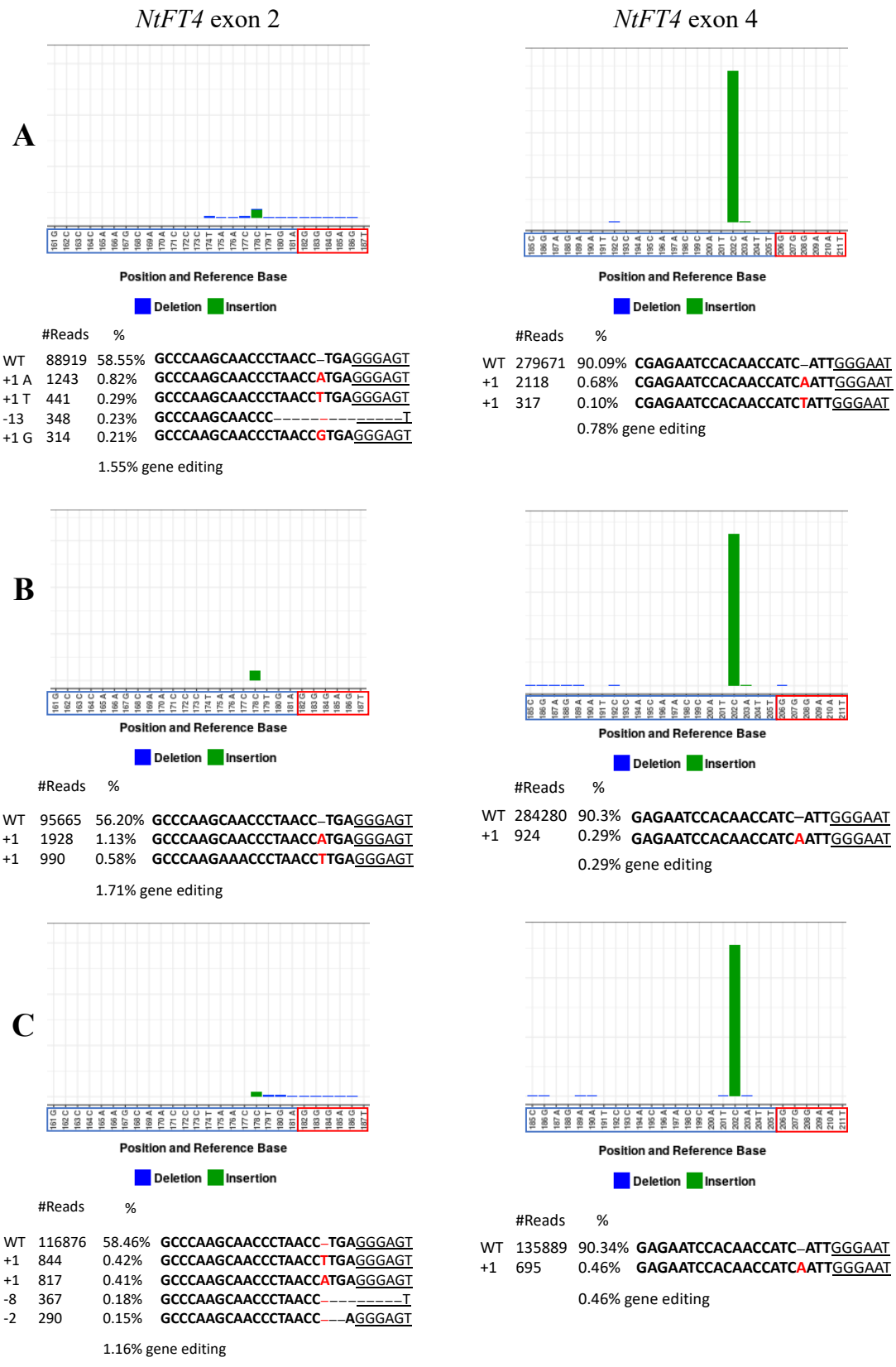


Figure 45. Genome editing events in co-infected leaves with TRV SaCas9 739N TobFT1 HH and TRV SaCas9 740C TobFT4 HH. *NtFT4* exon 2 target site is shown in the left column, while mutations in *NtFT4* exon 4 are illustrated in the right column. The editing in the target sites was analysed by NGS. *(Legend continue in next page)*

Finally, systemic samples were analysed from TRV *Sa*Cas9 full-length TobFT1 HH replicate 1, and TRV *Sa*Cas9 739N TobFT1 HH and TRV *Sa*Cas9 740C TobFT4 HH replicate 1. It is interesting to note that whilst *Cas9* RNA was detected in systemic leaves of plants where the split *Sa*Cas9 was delivered by TRV, no genome editing was found in any of the samples analysed as shown in Figure 46.

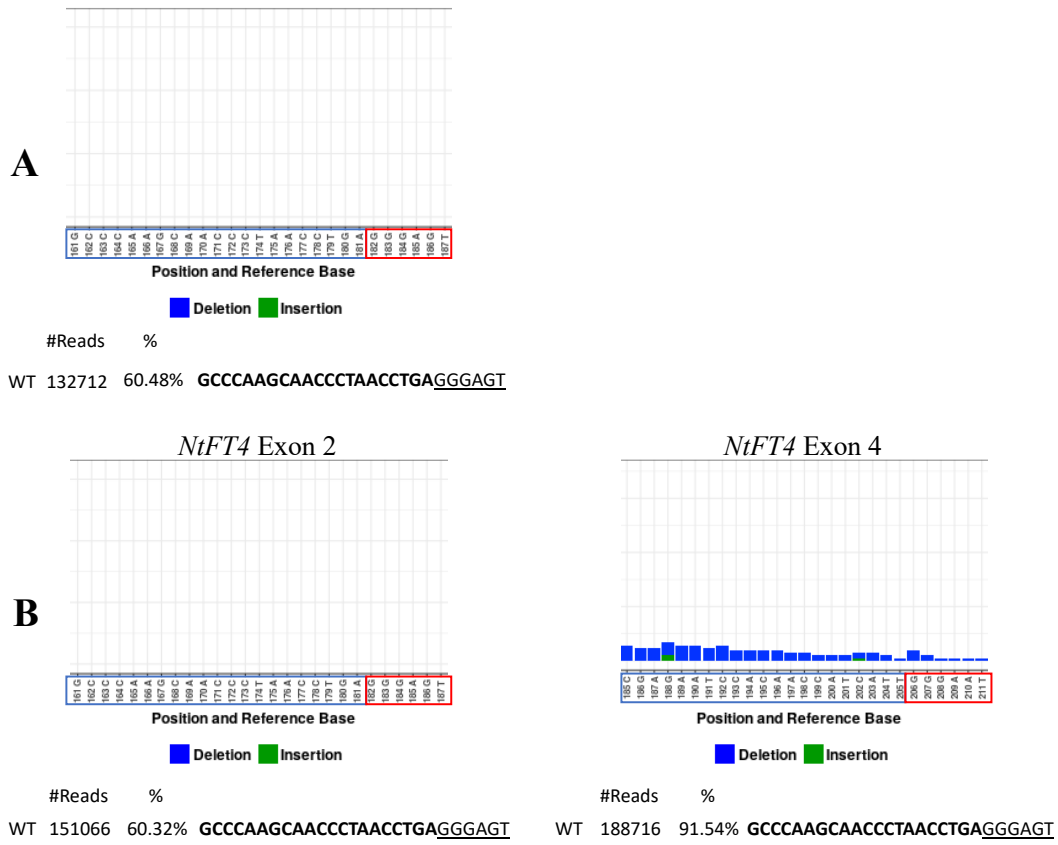


Figure 46. Genome editing events in systemic leaves of TRV-*Sa*Cas9-RGR infected plants. Genome editing was assessed in replicate 1. The editing in the target sites was analysed by NGS. Insertions are depicted in red, while deletions are shown as dashes. The number of reads and its frequency is shown. **A.** TRV *Sa*Cas9 full-length TobFT1 HH systemic leaf. No mutations were found at this target site. **B.** TRV *Sa*Cas9 739N TobFT1 HH and TRV *Sa*Cas9 740C TobFT4 HH systemic leaf. *NtFT4* exon 2 target site is shown at the left, while *NtFT4* exon 4 is illustrated at the right. No genome editing events were detected at these sites.

(*Continuation legend Figure 45).* Insertions are depicted in red, while deletions are shown as dashes. The number of reads per type of mutation and its frequency is shown. **A. Biological replicate 1.** For exon 2, indels were found at a rate of 1.55%. For exon 4 a 1 bp insertion was found at a rate of 0.78%. **B. Biological replicate 2.** An insertion of 1 bp was found in both target sites, with a frequency of 1.71% for exon 2 and 0.29% for exon 4. **C. Biological replicate 3.** For exon 2, indels were noticed, with a total frequency of 1.16%, for exon 4 a 1 bp insertion was found at a rate of 0.46%.

4.2.3.4 Flowering time of plants infected with *Sa*Cas9-RGR constructs delivered by TRV and PVX virus vectors

The introduction of mutations into *NtFT4* should create a delay in the flowering time of *N. tabacum* var. Maryland Mammoth plants (159). Infected plants were grown in SD conditions to promote flowering for approximately three months. Compared with the WT mock control, a delay in flowering was observed in two plants infected with TRV constructs (Figure 47). The first one is TRV *Sa*Cas9 full-length TobFT1 HH replicate 3 where no genome editing events were detected in infected leaf samples (Figure 44 C). No *Sa*Cas9 protein was detected in these samples by western blot analysis and only faint bands were seen in the *SaCas9* RNA expression analysis (Lane 5, figure 40). The second plant is TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH replicate 2. Insertions in both target sites were found at a total frequency of 1.71% for exon 2 and 0.29% for exon 4. Even though no *Sa*Cas9 protein was detected in this sample, strong bands were seen in RT-PCR analysis in both infected and systemic leaves samples (Lanes 10 and 13 respectively in Figure 40). Moreover, systemic symptoms of TRV infection were noticed (Figure 39). Thus, a delay in the flowering time of this specific plant could potentially be attributed to successful genome editing of one or both target sites in *NtFT4* gene.

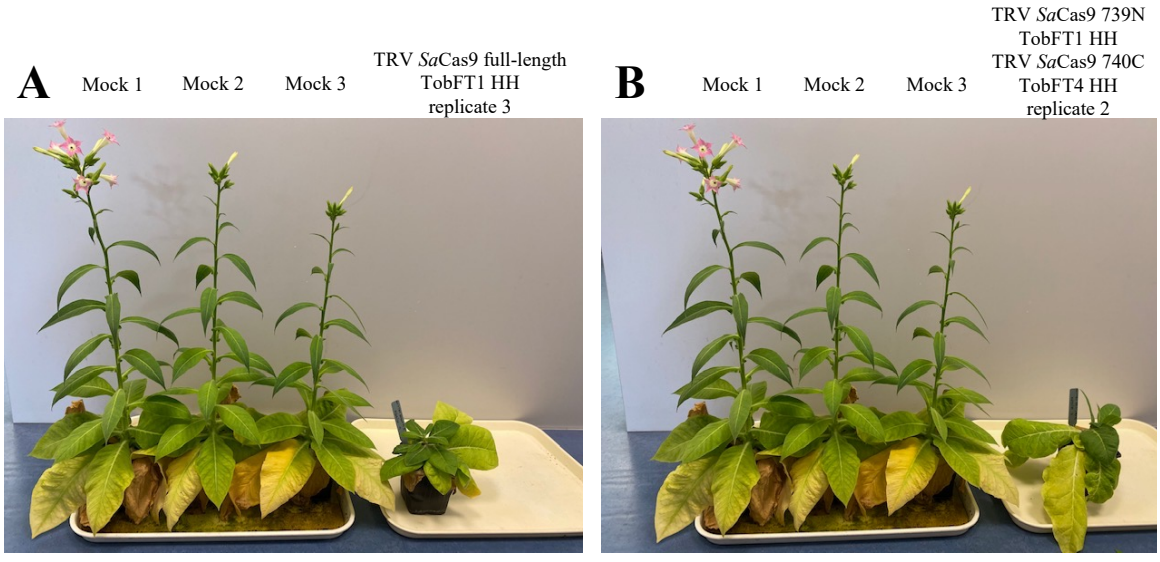


Figure 47. Flowering time in infected plants with TRV *Sa*Cas9-RGR constructs. Two plants shown a delay in its flowering time compared with WT mock controls. **A.** TRV *Sa*Cas9 full-length TobFT1 HH replicate 3. **B.** TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH replicate 2.

4.3 DISCUSSION

The use of plant virus as vectors to express heterologous proteins and RNA in plants represents an opportunity for the delivery of the CRISPR/Cas9 components with the aim of generating improvements in a broad range of crops. Plant viruses spread systemically throughout plants within days, hence genome edits can be achieved in a short period of time rather than weeks or months as in traditional stable transformation methods (107). Moreover, it had been described that some viruses, for example TRV, can enter meristematic cells and thus edited progeny can be obtained (108). The use of RNA viruses to deliver the genome editing system offers particular advantages such as less off-target activity, and non-integration of exogenous DNA which is a long-term goal in plant genetic manipulation.

In recent years, several studies have described the use of plant viruses as vectors for genome editing. Successful mutations were introduced by the expression of full-length Cas9 using the (+) ssRNA potexviruses FoMV and PVX, and (-) ssRNA rhabdoviruses BYSMV and SYNV (136-139). However, virus vectors usually have a limited cargo capacity (110, 111). To overcome this problem, the delivery of single or multiple sgRNAs into Cas9-expressing plants has been done using TRV (97, 127-129), TMV (130), PEBV (131), BSMV (132), BNYVV (133), FoMV (185) and PVX (134). Additionally, to reduce the size of the protein, the delivery of one of the halves of a split *Sa*Cas9 using ToMV and the other fragment by *Agrobacterium* plant expression vector successfully introduced mutations into the *PDS* gene in *N. benthamiana* plants (143).

Here the delivery, by PVX or TRV, of two full-length or split Cas9 orthologs in a transcriptional unit together with HHRz-sgRNA-HHRz targeting the *NtFT4* gene was tested in *N. tabacum* plants.

Viral systemic symptoms were seen in some plants, especially in plants infected with TRV. This was confirmed by RT-PCR analysis, where *TRV1* and *TRV2* RNA was detected in most of the systemic leaves (except in samples from plants infected with TRV full-length or split *Sp*Cas9-RGR, and TRV *Sa*Cas9 full-length RGR), while *PVX* RNA was only found in three systemic samples from split *Sa*Cas9 co-delivered by TRV (Figure 42 B). TRV was engineered as an *Agrobacterium*-based vector for

easy transformation, where TRV1 and TRV2 viral cDNA were cloned between the CaMV 35S promoter and a nopaline synthase terminator (120, 121). A PEBV CP promoter to express heterologous proteins in TRV2 was cloned in this thesis work. On the other hand, RNA viral copies of PVX vector are generated by *in vitro* transcription using a T7 RNA polymerase promoter and plants are infected by mechanical rubbing (114). This difference in the method of infection and subsequent production of viral copies in the cell could affect the efficiency of infection and could explain why strong TRV infections were obtained more consistently than for PVX. *Agro*-infiltrated vectors show greater and more equally dispersed infection of cells in inoculated leaves compared to traditional rub-inoculated vectors (204). Hence the use of an *Agrobacterium*-based PVX vector is strongly advised for further experiments.

As discussed, viruses are able to move systemically in *N. tabacum* plants. However, when these leaves were analysed by western blot or RT-PCR, in most of the samples no Cas9 protein or mRNA was found, and consequently no genome edits were detected. Full-length or split *SpCas9* mRNA was not detected systemically, indicating that the size of one or both halves might not meet the size requirement of cargo capacity of the viruses. For *Sa*Cas9, faint bands were detected for the full-length and N-terminal domain delivered by TRV. Strong bands were noticed for *Sa*Cas9 C-terminal domain when this construct was delivered by TRV, while weak expression was found in one PVX systemic leaf sample. Interestingly, the C-terminal domain was detected in infected leaves by immunoblot analysis when it was delivered by PVX or TRV in viral co-infection analyses. This result indicates that the size of the *Sa*Cas9 740C – RGR construct (1340 bp) is below the cargo capacity limit for TRV and PVX. The inability to detect full-length or split *SpCas9* RNA and the weak expression of full-length and N-terminal *SaCas9* RNA in systemic leaves, even when viral systemic symptoms can be seen, indicates a probable loss of all or part of the transgene during the replication cycles of the virus during its spread through the plant. A positive correlation between genetic instability of the viral vector and insert size has been reported, where large inserts reduce viral replication and expression levels and are more susceptible to the loss of the heterologous genes as early as the first infection passage (110, 111). For instance, plants infected with PVX-GFP progressively lost the transgene, and over time only the WT recombinants

prevailed (111). This has been attributed to the use of duplicated sgP to express exogenous genes (such as in PVX), since RNA homologous recombination occurs between both sub-genomic promoters, leading to a deletion of the transgene and one of the duplicated promoters. Thus, it is advised to use two different sgP to express heterologous genes (111). However, even when different sub-genomic promoters are used, as for example in the case of the TRV vector used in this thesis, the size of the transgene must still be considered. Considering the results obtained with *Sa*Cas9 740C-RGR, an ideal heterologous gene should not exceed 1.2 -1.3 kb for both PVX and TRV viruses.

Delivery of a split Cas9 protein was proposed as an alternative to meet the cargo capacity criteria of the virus vectors. For this, each end of the gene was delivered by TRV, PVX, or a combination of both. Viral cross protection has been reported in plants to prevent infection with the same virus twice (140, 141). Marton *et al.* (2010) reported that co-inoculation of two TRV2 viruses with different transgenes resulted in less efficient expression of both (125). In contrast, the co-infiltration of two FoMV viruses expressing the sgRNA and Cas9 gene separately did not show this cross-protection phenomenon (136). The use of complementary viruses to deliver protein fragments has been proposed (207). However, this strategy can lead to severe viral symptoms or cause the death of the plants (185). In this thesis, similar results as reported for FoMV were observed, where *Cas9* RNA expression was obtained in leaves co-infected with TRV, PVX, or a combination of them. Furthermore, genome edits (at frequencies ~1%) were detected in leaves where split Cas9 halves were co-delivered by TRV. Even though no sequencing analyses were done on leaves co-infected with TRV and PVX, it is possible that higher rates of mutation could be achieved in these samples due to the absence of viral cross protection as Marton *et al.* suggests. Therefore, further NGS analysis are recommended to compare frequencies of genome edits among these co-infected samples.

To increase the likelihood of delivering the sgRNAs into the same cell as the Cas9 protein, a single transcriptional unit was created by fusing self-cleavable ribozymes to each end of the sgRNA to create the RGR unit, which was positioned downstream of the *Cas9* gene. This artificial gene can be expressed *in vivo* from a single promoter. The autocatalytic activity of the ribozymes was analysed by cRT-PCR, and mature TobFT1 and TobFT4 sgRNAs were shown to be produced with intact

guide and scaffold RNAs. In some cases, the 3' end of the sgRNA was trimmed, which could be due to RNA degradation during sample preparation (101). The expected 3'-HHRz scar was found for both sgRNAs, while in some cases part of the 5'-HHRz was also detected, indicating its incomplete cleavage. It has been proposed that self-cleaving RNAs, such as HHRz or tRNAs, between multiple sgRNAs and/or the Cas9 is not needed (134, 139). It has been suggested that the plant's own RNA processing systems trim off the extra sequences resulting in a mature Cas9-sgRNA complex (102). Cody *et al.* (2017) used an engineered TMV virus vector to deliver three types of GFP sgRNAs into *N. benthamiana* 16c plants: one with a hepatitis delta virus (HDV) ribozyme at the 3' end (gHDV), a second sgRNA with a 5' hammerhead ribozyme and a 3' HDV ribozyme (RGR) and a third sgRNA without ribozyme (gGFP). The authors found that both gGFP and gHDV showed higher mutation rates at the target site (above 60%) than RGR (46%) by gel quantification analysis. Additionally, when gRGR was delivered without Cas9, RT-PCR and sequencing analyses exhibited a deletion of the sgRNA from the virus genome. Therefore, the authors concluded that the 5' HHRz had a negative effect on the virus replication and that the construct without ribozymes was sufficient and efficient for delivering biologically active sgRNAs (130). Considering these studies, the use of ribozymes seems to potentially hindrance the genome editing efficiency.

Mutations in the *NtFT4* target gene were successfully found when genomic DNA was analysed from directly infected leaves, indicating that viruses are able to deliver the CRISPR/Cas9 components and these are active in plant cells. *Sp*Cas9 showed slightly higher frequencies of genome editing than *Sa*Cas9 (3.03% versus 1.33% respectively). On the other hand, delivery of split *Sa*Cas9 was more successful introducing genome edits in both *NtFT4* target sites than split *Sp*Cas9. A recent report compared the targeted mutagenesis efficiency of four Cas9 orthologs in *N. benthamiana* protoplasts, finding that *Sa*Cas9 introduced more mutations at the target site than *Sp*Cas9, *As*Cas12a or *Lb*Cas12a (193). Nevertheless, Kaya *et al.* (2016) found that similar editing efficiencies between *Sp*Cas9 and *Sa*Cas9 were found in tobacco and rice callus (142). In this thesis, similar genome editing frequencies were obtained between *Sp*Cas and *Sa*Cas9, regardless of whether the full-length, or split protein approach was tested.

Genome editing frequencies in TRV infected leaves varied from 3.03% to 0.42% for *Sp*Cas9 full-length, 1.95% to 0.79% for split *Sp*Cas9, 1.33% to 1.06% for *Sa*Cas9 full-length, and 1.71% to 0.29% for split *Sa*Cas9. Ariga *et al.* (2020) indicated that in plants inoculated with a PVX-Cas9-sgRNA construct genome editing efficiency was higher than in *Agro*-infiltrated leaves expressing *Sp*Cas9 and sgRNA transiently. The authors also reported that their RT-PCR analysis suggests that *PXV-Cas9* RNA infected most of the cells of inoculated leaves, expressing large amount of Cas9 protein (139). Ma *et al.* (2020) using a SYNV-*Sp*Cas9-sgRNA vector, reported mutation frequencies between 40 to 91% depending on the particular sgRNA used (137).

The screening of mutations by colony PCR was a laborious and expensive method, in which genome edits at low frequency weren't identified. On the other hand, NGS analysis of the *NtFT4* target region proven to be a simple, reliable, and a sensitive tool for the detection of genome edits, finding mutations at rates as low as 0.29%. The frequency of genome edits depends on many factors, such as design of the sgRNA, genomic position of the target site, delivery technique and biology of the organism (4). Over the past few years, multiple techniques had been developed to detect targeted genome edits. These are:

1) Characterization of a phenotype, where the introduction of mutations into certain genes gives visible characteristics, as for example photo-bleach leaves when *PDS* gene is interrupted.

2) PCR techniques, which are low cost and easy to use. Some of them, such as the T7E1 or SURVEYOR® assays, rely on the detection of mismatches while others depend on the mutation of digestion sites, like RFLP or CAPS (PCR/RE) assays. Moreover, size comparison of the PCR products, for example by capillary electrophoresis, or qPCR methods using specific probes can also be used to distinguish between WT and mutants. However, further Sanger sequencing of the PCR product is needed to confirm the genome edit.

3) Digital PCR (dPCR) is a sensitive, high-throughput and precise tool, particularly useful in the detection of genome edits in polyploid organisms. Here, the DNA mixture is separated in small droplets or chambers and single parallel PCRs occurs

simultaneously. Then each reaction is evaluated by the presence or absence of fluorescent signal.

4) Sequencing methods, which can be separated into amplicon-based target sequencing or hybrid capture sequencing. The first is more efficient for smaller regions with a high percentage of on-target reads; in contrast, the second is used usually for larger target regions. NGS methods are a sensitive and high-throughput analysis, which allow the detection of low-frequency genome edits in heterogeneous samples. For PCR and dPCR methods a knowledge of the target genome is needed to design primers to amplify the edited region, however GC content or repetitive sequences can affect the development of these techniques. In these cases, NGS tools are recommended (4).

A delay in the flowering time was observed in four *N. tabacum* var. Maryland Mammoth plants infected with TRV, as depicted in Figures 37 and 47. Not all plants where genome editing was confirmed displayed a delay in the flowering time. Three independent plants were infected per virus constructs as biological replicates, defined as "parallel measurements of biologically distinct samples that capture random biological variation, which may itself be a subject of study or a noise source" (208). Even though the plant's age and infection conditions were the same as a start point, the viral replication and systemic movement will be independent (and maybe different) in each plant, causing a biological variation. Quantitative RT-PCR of the viral *cp* gene may contribute to account for this variability. Successful genome edits and a delay in flowering were obtained for two plants infected by TRV. In the case of full-length *Sp*Cas9-TobFT1 HH systemic symptoms of infection were noticed (Figure 23), but not the systemic presence of *Cas9* RNA (Figure 26). When each end of split *Sa*Cas9 was delivered, systemic symptoms (Figure 39) and systemic *Cas9* RNA (Figure 40) were observed. Hence, a delay in the flowering time of these specific plants could potentially be attributed to successful genome editing. TRV vector can infect meristematic cells, thus the seeds of this plant should be analysed to check for heritable genome edits. On the other two cases, the late flowering phenotype could be a result of strong infection by TRV, since the analysis of the target sites revealed only WT sequences.

Cas9 and viral RNA were detected in all of the RT-PCR analyses of infected leaves. Interestingly, full-length *Sp*Cas9 protein was also detected in TRV infected leaves, which supports the hypothesis that the transgene is lost during numerous rounds of viral replication required for systemic movement. Even though neither split *Sp*Cas9 domains, nor full-length and split *Sa*Cas9, were identified at the protein level, genome editing events were found in these samples, indicating that sufficient Cas9 protein was produced in the leaves to generate targeted genome edits, but the protein levels were below the detection limit of western blot analysis.

As screening for genome editing was not done in plants where systemic infection wasn't clear, such as in plants infected with PVX or a combination of PVX and TRV, *Cas9* and *PVX* RNA were detected in these samples. Since the NGS analysis is a very sensitive approach to screen for mutations, further analysis of this plants should be considered.

The use of virus vectors to deliver CRISPR/Cas9 components offers several advantages, however the virus cargo capacity limit is its main bottleneck. Interestingly, positive results have been reported by Uranga *et al.* (2021), Ma *et al.* (2020) and Ariga *et al.* (2020), where plants have been regenerated from virus inoculated leaves with these plants showing higher genome editing efficiency (134, 137, 139). Since in the present work, successful genome edits were detected in leaves infected with TRV full-length and split *Sp*Cas9/*Sa*Cas9 and *Cas9* RNA was detected in the PVX and TRV/PVX infected leaves with these constructs, regenerated shoots from these leaves will be analysed for genome editing.

CHAPTER 5: GENERATION OF *NtFT4* EDITED *N. tabacum* PLANTS BY TISSUE CULTURE OF LEAVES FOLLOWING INFECTION WITH TRV AND PVX Cas9-RGR CONSTRUCTS.

5.1 INTRODUCTION

The development of new cultivars by traditional plant breeding techniques is a long process that may take many years. The use of genome editing techniques, such as CRISPR/Cas9, could speed up the process of crop breeding and enable modification of a broad range of commercially beneficial genes. Even though this method is faster than traditional breeding, advances to improve it have been proposed. Among those, the regeneration of edited plants from protoplasts or leaves by the exposure of the tissue to various hormones has been successful in several species (209).

Plant regeneration relies on the ability to reprogramme somatic cells to generate new tissues, organs, or a complete plant without the need of sexual reproduction. This capacity to form new plantlets has been used by centuries for clonal propagation of crops, while *in vitro* regeneration of plants was developed during the early 1900s (210). Plant regeneration can be divided into organogenesis and somatic embryogenesis. The former is the development of organs from cultured tissue, while the second is the dedifferentiation of somatic cells to totipotent embryo cells, that can further generate a whole plant (211). Moreover, direct organogenesis refers to the development of *in vitro* organs directly from explants. On the other hand, indirect organogenesis requires an intermediate structure (callus) an undifferentiated tissue to form leaves, shoots, roots, and embryos (209). Three phases have been described for plant regeneration: 1) A “dedifferentiation” step, where somatic cells under certain hormonal stimulus obtain similar characteristics to meristematic cells. However, in the last few years evidence has suggested that this process is more similar to “transdifferentiation”, in which somatic cells resemble root tissue rather than meristematic cells (212). 2) Reprogramming step, where the cell fate of this new callus tissue is defined by hormonal balance. For example, high concentrations of auxin versus low concentration of cytokinin induce the development of roots, while the opposite stimulates shoot formation (213). 3) Morphogenesis phase, where the different tissues are formed (209).

The main hormones used for plant regeneration are auxin and cytokinin, however gibberellins, ethylene and abscisic acid also play roles in the regeneration of certain species (209). The most common auxins are naphthalene acetic acid (NAA), indole-3-butyric acid (IBA), 2,4-dichlorophenoxyacetic acid (2,4-D) and indole-3-acetic acid (IAA). The most common cytokinins are thidiazuron (TDZ), zeatin, kinetin, 6-benzylaminopurine (BAP), isopentenyl adenine (2iP) (214). Several protocols are available for plant regeneration of diverse species combining different types of auxins and cytokinins, however, conditions must be experimentally tested.

Another factor to consider for *in vitro* plant regeneration is media composition. Usually it contains macronutrients, micronutrients, vitamins, amino acids, nitrogen and organic supplements, carbon source, growth regulator (hormones) and solidifying agents. Commercial media available are Murashige and Skoog (MS) medium, Linsmaier and Skoog (LS) medium, Gamborg (B5) medium and Nitsch and Nitsch (NN) medium, and media selection will depend on plant nutritional requirements (214). Specifically, vitamins are required for correct growth and development since they participate as cofactors in several enzymatic reactions or as precursors for the synthesis of several biomolecules (215). The most common used are thiamine (B_1), nicotinic acid (B_3) and pyridoxine (B_6). Thiamin is necessary for all cells to growth, but nicotinic acid and pyridoxine are not essential for cell growth in some species, hence their addition to the media is optional (214). Since results are dependent on the plant species, among other factors, it's important to test plant regeneration protocols.

Considering that the development of new plants regenerated from somatic cells is a fairly quick process, this procedure its widely used for testing genome editing events by the CRISPR/Cas9 system. Interestingly, plants exhibiting homozygous or heterozygous mutations can be obtained in the first generation after regeneration, accelerating its use for crop breeding (216). A DNA-free genome editing approach has been developed by the delivery of an RNP complex of Cas9 protein and sgRNA into protoplasts, which were successfully regenerated to fully grown edited plants (181, 217). However, protoplast regeneration presents many technical difficulties, being inefficient or unattainable for some target crops (137, 177, 179). Another approach to obtain non-GMO plants is the regeneration of shoots from somatic tissue directly infected with RNA virus vectors, since these don’t integrate into the host

genome. Using this methodology, high genome editing efficiency can be achieved, due to the ability of the RNA virus to infect most of the plant cells of the infiltrated leaves (134, 137, 139). Uranga *et al.* (2021) described the regeneration of shoots from edited leaf tissue where single or multiple sgRNAs were delivered by PVX in plants constitutively expressing the Cas9 protein. This study reports genome editing frequencies between 46% to 95.8% in these regenerated plants where different genes were edited using single or multiple sgRNAs (134). Ariga *et al.* (2020) indicate that from 50 regenerated shoots from PVX-*Sp*Cas9-*TOM1* sgRNA infiltrated leaves, 31/50 shoots carried mutations in one or both *TOM1* alleles (139). Ma *et al.* (2020) reported that shoots regenerated from leaves infected with SYNV-*Sp*Cas9-*PDS* sgRNA shown ~57% (17/30) homozygous/bi-allelic mutations, while ~93% (28/30) carried targeted mutations of any type (137).

As presented in CHAPTER 4, genome editing events were detected in TRV-infected leaves with full-length or split Cas9 constructs. Even though no genome edits were assessed in the infected leaves with PVX or a combination of PVX and TRV, *Cas9* RNA was detected in these samples, which might be sufficient to introduce targeted genome edits. Considering the high percentage of mutations reported in shoots regenerated from virus infected leaves, this approach was tested. Shoots were regenerated from *N. tabacum* leaves directly infected with the PVX and TRV virus vectors expressing the full-length *Sp*Cas9, *Sa*Cas9 or split version of both fused to the *NtFT4*sgRNAs. Genome editing in the resulting shoots was assessed by CAPS assay, a PCR technique coupled with a digestion step.

5.2 RESULTS

5.2.1 Conditions for plant regeneration from *N. tabacum* leaf explants

To establish the protocol for regenerating tobacco leaves into fully-grown plants, wild type controls were used. *N. tabacum* leaves from 4-weeks old plants were cut in squares of approximately 1 cm^2 surface area and sterilized with 1% NaOCl for 30 min, followed by three washes with sterilized water. Each square was placed onto non-selective shoot induction media using 0.2 mg/L NAA and 1 mg/L BAP or kinetin to induce callus formation (Figure 48 A). The concentration of hormones used was as described by Ariga *et al*. (2020) (139).

In general, kinetin caused faster callus formation than BAP (data not shown). Moreover, the addition of Plant Preservative Mixture (PPM™) as biocide to prevent contamination of the tissue culture was examined, however a substantial delay in the production of calli was noticed, hence its use was discontinued. Sections were moved onto fresh shoot induction media if contamination was observed.

N. tabacum callus started to develop after a month of regeneration (Figure 48 B). Approximately two weeks later, shoots started to develop (Figure 48 C), and the plantlets were transferred from shoot induction media to non-selective regeneration media for root development, containing a vitamin mix of 1 mg/L thiamine, 0.5 mg/L nicotinic acid and 0.5 mg/L pyridoxine (Figure 48 D). Overall, the regeneration of leaves into grown plants took about two months and a half (Figure 48 E).

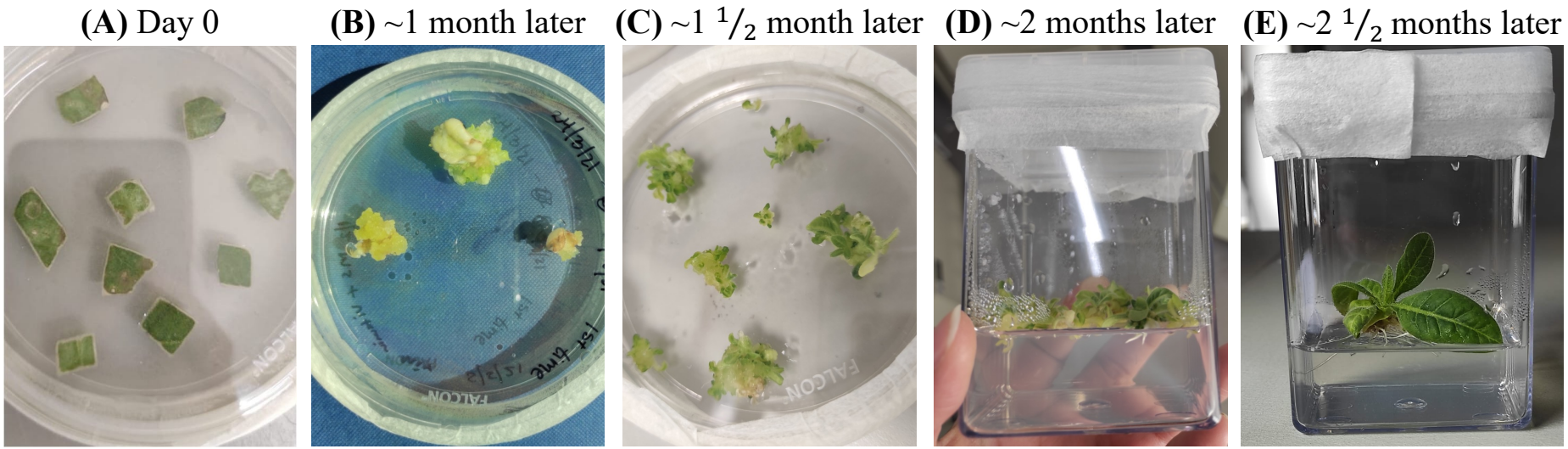


Figure 48. Shoot regeneration of plants from *N. tabacum* WT leaves. A. Square leaves sections were placed onto shoot induction media. **B.** Callus was formed after a month. **C.** Shoots appeared two weeks later, and plantlets were moved to regeneration media. **D.** Development of roots. **E.** Regenerated plantlet after 2 months and a half approximately.

5.2.2 Regeneration of shoots from virus infected *N. tabacum* leaves

N. tabacum leaves from 4-weeks old plants grown in soil or sterile conditions were infected with the different TRV and PVX viral constructs expressing the CRISPR/Cas9 system and shoots were regenerated from these leaves.

Regeneration from systemic leaves was not pursued, since no genome editing events, and inconsistent *Cas9* RNA expression or viral symptoms were seen in most of the plants (as presented in CHAPTER 4).

The PVX virus constructs carrying *Sp*Cas9 full-length or *Sa*Cas9 full-length with the guides TobFT1 or TobFT4 with ribozymes were *in vitro* transcribed into RNA which was rubbed onto *N. tabacum* leaves, while TRV virus expressing the same constructs was *Agro*-infiltrated. For the split Cas9 (N-*Sp*Cas9 N-intein TobFT1 HH and C-intein C-*Sp*Cas9 TobFT4 HH or *Sa*Cas9 739N TobFT1 HH and *Sa*Cas9 740C TobFT4 HH) a mixture of the same virus, or a combination of PVX and TRV, was used. Two rounds of inoculation were performed, one in plants grown in soil and one in plants grown in sterile conditions. Four infected leaves per round of infection were collected after 7 to 10 dpi, cut into sections, sterilized, and placed onto shoot induction media, as described in section 5.2.1. Between four to eight leaf sections were placed onto each plate. Carbenicillin was added as antibiotic to prevent *Agrobacterium* growth.

Calli were obtained approximately after a month of culture (Figure 49 A). When shoots started to develop, they were moved to regeneration media for root development, replacing the media every week (Figure 49 B). Multiple shoots were obtained from the same leaf section. Due to their high density, shoot samples were taken for genome editing analysis. Table 2 summarizes the viral constructs transformed, conditions where regenerated shoots were obtained, and number of shoots analysed. Shoot samples were retrieved from 10 conditions, marked as “yes” in Table 2, whilst for six conditions no samples were taken due to contamination (fungal or bacterial) Therefore, genome editing analysis for these samples was not pursued.

5.2.3 Assessment of genome editing in regenerated shoots from virus infected leaves

Genome editing was analysed in the regenerated shoots by the CAPS assay, also known as PCR/RE assay. For this, the cleavage site of the Cas9 in the guide RNA (3-4 bp upstream of the PAM site) needs to be within a restriction enzyme site. If a mutation is introduced into the targeted sequence, the interrupted site will not be recognized by the restriction enzyme.

Table 2. Regenerated shoots from viral infected leaves

Viral construct	Regenerated shoots	# Shoots for GE
TRV *Sp*Cas9 full-length TobFT1 HH	No	-
TRV *Sp*Cas9 full-length TobFT4 HH	No	-
PVX *Sp*Cas9 full-length TobFT1 HH	Yes	3
PVX *Sp*Cas9 full-length TobFT4 HH	Yes	8 (NA)
TRV N-*Sp*Cas9 N-intein TobFT1 HH TRV C-intein C-*Sp*Cas9 TobFT4 HH	Yes	39
PVX N-*Sp*Cas9 N-intein TobFT1 HH PVX C-intein C-*Sp*Cas9 TobFT4 HH	No	-
TRV N-*Sp*Cas9 N-intein TobFT1 HH PVX C-intein C-*Sp*Cas9 TobFT4 HH.	Yes	7
PVX N-*Sp*Cas9 N-intein TobFT1 HH TRV C-intein C-*Sp*Cas9 TobFT4 HH.	No	-
TRV *Sa*Cas9 full-length TobFT1 HH	Yes	21
TRV *Sa*Cas9 full-length TobFT4 HH	Yes	21 (NA)
PVX *Sa*Cas9 full-length TobFT1 HH	No	-
PVX *Sa*Cas9 full-length TobFT4 HH	Yes	8 (NA)
TRV *Sa*Cas9 739N TobFT1 HH TRV *Sa*Cas9 740C TobFT4 HH	Yes	29
PVX *Sa*Cas9 739N TobFT1 HH PVX *Sa*Cas9 740C TobFT4 HH	Yes	20
TRV *Sa*Cas9 739N TobFT1 HH PVX *Sa*Cas9 740C TobFT4 HH	No	-
PVX *Sa*Cas9 739N TobFT1 HH TRV *Sa*Cas9 740C TobFT4 HH	Yes	15

NA: Not analysed

To pursue this assay, the edited region is amplified and then the PCR product is digested. If a mutation is successfully induced, an undigested DNA band will be observed in agarose gels.

For *NtFT4* exon 2 the restriction enzyme *Dde*I has a cut site within the CRISPR/Cas9 target site. In contrast, no restriction enzyme could be used for *NtFT4* exon 4, hence 21 calli regenerated from TRV *Sa*Cas9 full-length TobFT4 HH infected leaves, eight calli regenerated from PVX *Sp*Cas9 full-length TobFT4 HH infected leaves and eight calli regenerated from PVX *Sa*Cas9 full-length TobFT4 HH infected leaves couldn't be examined for targeted genome editing by this method, shown as "NA" in Table 2 (due to time constraints other analyses were also not carried out).

Positive controls where genome editing was previously confirmed were used as a first approach to test the efficiency and sensitivity of the CAPS assay (Figure 50). These samples correspond to genomic DNA extracted from transformed protoplast with *Sp*Cas9 or *Sa*Cas9 TobFT1, in which WT sequences and indels were detected by Sanger sequencing of individual colonies (Figure 18). An additional sample was also tested, corresponding to a clone exhibiting a 4 bp deletion in the *NtFT4* exon 2 target region edited with *Sp*Cas9 full-length TobFT1. An un-edited gDNA sample from WT *N. tabacum* leaves was also tested. Moreover, the activity of *Dde*I in the Taq Polymerase buffer was evaluated since an incompatible buffer might affect restriction enzyme efficiency.

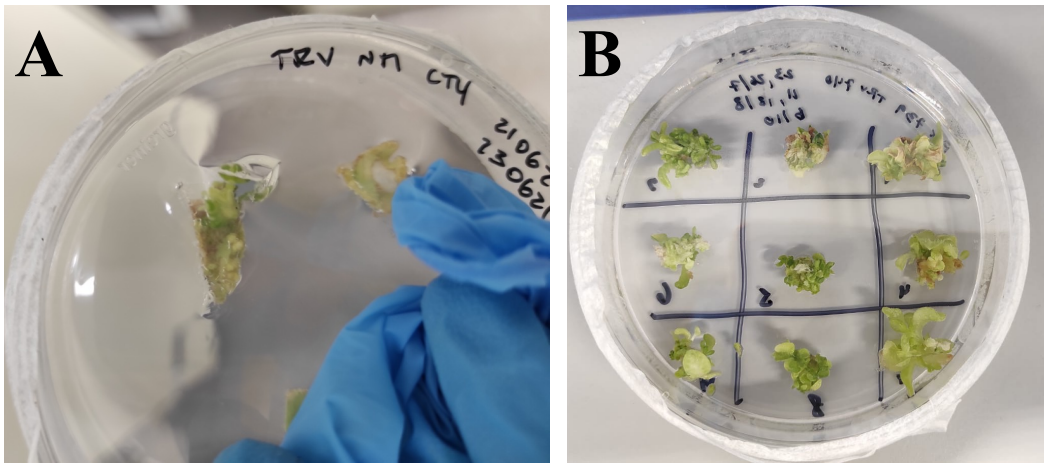


Figure 49. Shoot regeneration from virus infected tobacco leaves. A. Callus developed after a month on shoot induction media. **B.** Shoots were separated and placed on regeneration media, where shoot samples were taken for genome editing analysis.

DNA from these selected samples was amplified using primers for the *NtFT4* exon 2 target site. The digestion reaction using the *Dde*I enzyme was carried out immediately after the PCR reaction, without a purification step in between, and the digested PCR products were run on a 1.5% agarose gel, loading 20 μL of the mixture.

Genome edits were successfully assessed by the CAPS assay, as shown in Figure 50. PCR products from edited protoplast with both *Sp*Cas9 (lane 4) and *Sa*Cas9 (lane 5) displayed partial *Dde*I digestion, where an undigested band of 320 bp and two bands of 199 bp and 121 bp were found, indicating the presence of WT and edited DNA in the protoplast pool. In the case of the clone where a deletion of 4 bp was found, an undigested band is observed, along with other non-specific bands, indicating the presence of edited DNA (lane 3). Efficient cleavage activity of *Dde*I in the Taq Polymerase buffer was also confirmed, since in the WT control (mock, lane 2) two bands are seen, indicating complete cutting of an unmodified restriction site.

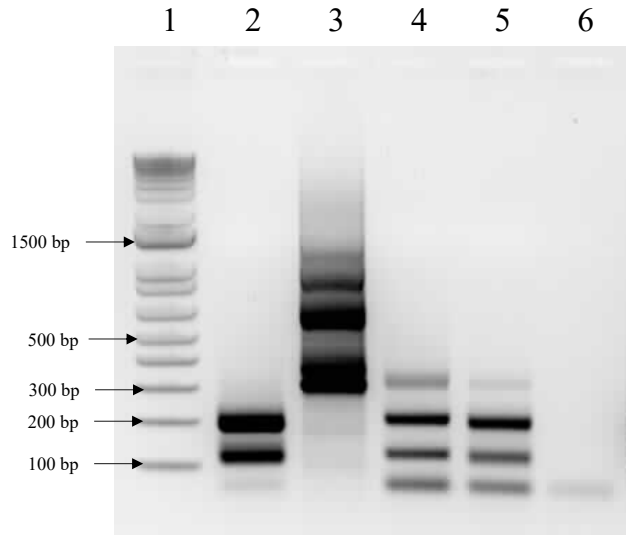


Figure 50. Assessment of genome editing in control samples using CAPS assay. *NtFT4* exon 2 targeted region was PCR amplified using Green GoTaq® Polymerase and then digested for 1 hr using the restriction enzyme *Dde*I. Reaction products were run on a 1.5% agarose gel, loading 20 μL of the mixture. Uncut bands of 320 bp are expected when genome editing occurs. On the other hand, two bands of 199 and 121 bp are expected in WT samples. Lane 1. 1 kb plus DNA ladder. Lane 2 DNA from WT *N. tabacum* leaf. Lane 3. *NtFT4* exon 2 edited with *Sp*Cas9 full-length TobFT1 isolated from protoplasts and cloned into pGEM®-T Easy vector. Lane 4. Protoplast DNA edited with *Sp*Cas9 full-length TobFT1. Lane 5. Protoplast DNA edited with *Sa*Cas9 full-length TobFT1. Lanes 6. Negative PCR control.

Since the CAPS assay is a fast and reliable assay to test genome editing events in plant material, it was used to assess the presence of mutations in the regenerated shoots. Samples and number of shoots analysed using this assay are shown in Table 2. DNA was extracted from a total of 134 shoots and the *NtFT4* exon 2 target site was amplified. The amplicons were digested with *Dde*I and the products were run on a 1.5% agarose gel. As positive controls, both undigested and digested samples were included. Negative controls of DNA extraction and amplification were also included. From the total of samples, 125 amplified correctly. A representative result of the CAPS assay to evaluate genome edits is shown in Figure 51, where all tested samples are wild type with no genome editing detected in the regenerated shoots from virus infected *N. tabacum* leaves. The remaining agarose gels showing the results obtained are in Appendix A6.

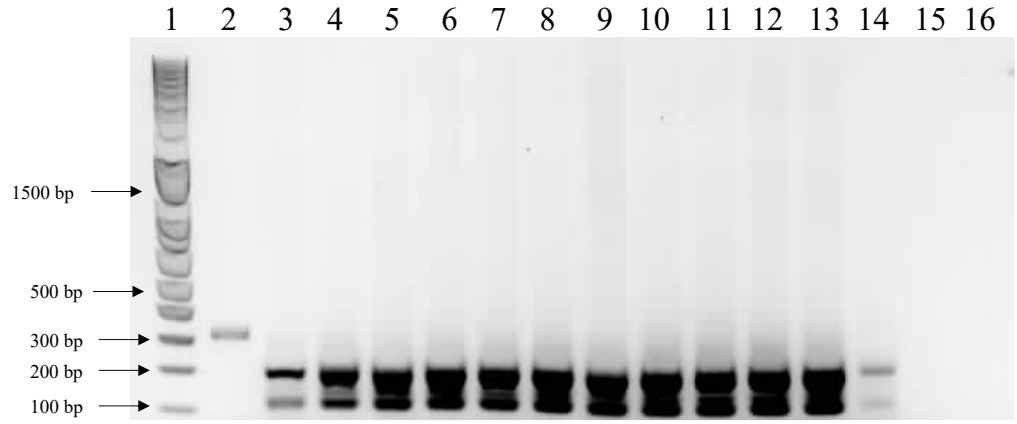


Figure 51. Assessment of genome editing in regenerated shoots from virus infected leaves. A representative figure of CAPS assay carried out in 18 samples is depicted. Lane 1. 1 kb plus DNA ladder. Lane 2. Undigested control sample showing the band size expected if a mutation is detected (320 bp). Lanes 3 – 13. Shoots regenerated from TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4 HH infected leaves. Lane 14. Digested WT control sample, where two bands at 199 and 121 are seen. Lane 15. Negative control of DNA extraction. Lane 16. Negative control of PCR amplification. All shoot regenerated samples exhibit two bands of 199 bp and 121 bp, indicating no genome editing events.

5.3 DISCUSSION

The ability of plants to regenerate has been used for years for *in vitro* propagation and the study of plant development. In the last years, several studies have reported the regeneration of plant cells infected with viral vectors delivering the CRISPR/Cas9 components for targeted genome editing.

Uranga *et al.* (2021) report genome editing frequencies between 46% to 95.8% in regenerated plants where different genes were edited using single or multiple sgRNAs delivered by PVX into Cas9 overexpressing *N. benthamiana* plants. This estimation was done using an algorithm to analyse the presence of mutations based on Sanger sequencing results (134). Ariga *et al.* (2020) report that from 50 regenerated shoots from PVX-*Sp*Cas9-*TOM1* sgRNA infiltrated leaves, 31 shoots carried mutations in one or both *TOM1* alleles. On the other hand, 1 out of 64 shoots regenerated from *Agrobacterium* harbouring the CRISPR/Cas9 system exhibited genome edits in the target site (139). In both studies, the authors stated that the editing efficiency in these regenerated plants was higher than in the parental tissue, indicating that the PVX vector infected most of the cells of the infiltrated leaves (134, 139). Ma *et al.* (2020) delivered the *Sp*Cas9-sgRNA flanked by two tRNAs using a SYNV vector. As previously, regenerated edited plants from infected leaves shown a high genome editing frequency. 17/30 of the regenerated plants were homozygous/bi-allelic for both *PDS* homologs, while 28/30 regenerated plants exhibited targeted mutations of any type. Furthermore, regenerated plants from systemically infected tissue also carried the desired target mutations, with genome editing efficiencies between 90-100% assessed by band intensity in PCR/RE assays (137).

In this work, a protocol for plant regeneration from *N. tabacum* leaves was set up, obtaining fully grown plants from leaf sections after approximately two months. Non-selective media prepared using 0.2 mg/L of NAA and 1 mg/L of BAP or kinetin proved to be effective for shoot induction, as described previously by Ariga *et al.* (2020) (139). It has been reported that high ratios of cytokinin to auxin promotes shoot regeneration, while the opposite has been described for root development (213). Regenerated shoots were moved to non-selective regeneration media, where 1 mg/L of thiamine, 0.5 mg/L of nicotinic acid and 0.5 mg/L of pyridoxine as vitamin

mix was used for root development. These three vitamins are usually used together and are involved in many plant pathways, as cofactors for several key enzymes or precursor for the synthesis of NAPDH, ATP, amino acids, nucleic acids, or other essentials compounds (215).

As a gelling agent, Gellan Gum, a bacterial polysaccharide, was used instead of Agar, a polysaccharide isolated from seaweed. Preliminary observations in our group determined a better performance for callus induction and visualization of root development of gellan Gum over Agar or Phytagel. Mohamed *et al.* (2021) tested different gelling agents, as agar, bacto agar, phytagel or gelrite (both types of gellan gum) on different stages of rice regeneration (callus, shoot and root regeneration). Physical characteristics of the media were described, with media prepared using gellan gum being more transparent, rigid, and brittle. Shoot regeneration was not affected across the four solidifying agents, however better callus induction and a significant higher root regeneration frequency was observed when phytagel or gelrite were used. The authors hypothesized that this might be because of the impurities found in agar, which is a mixture of linear agarose and agropectin, while gellan gum is a natural, highly purified agent (218).

Once conditions for *N. tabacum* regeneration were established, leaves were infected with PVX or TRV virus vector expressing *Sp*Cas9 full-length, *Sa*Cas9 full-length or a split version of both. Leaf sections were placed on non-selective shoot regeneration media and the obtained shoots were screened for genome editing events using PCR coupled with a digestion step assay. It is important to consider that each cell is infected by the virus independently, so even though the tissue culture is a clonal reproduction, different shoots derived from different cells should exhibit different patterns of genome editing.

Samples where genome editing had been previously confirmed by sequencing were used as positive controls in CAPS assay (Figure 50). Edited protoplast samples were partially resistant to *Dde*I digestion, indicating a mixture of WT and edited DNA in the protoplast pool. A clone of *NtFT4* exon 2 target region edited with *Sp*Cas9 full-length TobFT1, which by Sanger sequencing exhibited a deletion of 4 bp in the target site (Data not shown), was fully resistant to *Dde*I digestion. In contrast, a PCR product of *NtFT4* exon 2 target region from a WT sample was fully digested by the

enzyme. Moreover, it was determined that a purification step is not required between PCR and digestion reactions, indicating that *Dde*I restriction enzyme is active under GoTaq® Polymerase reaction buffer, hence this is a quick and cheap assay to screen for mutations.

Several detection methods have been developed to detect indels in mutated samples, reviewed extensively by Bennet *et al.* (2020) (219). The CAPS (or PCR/RE) assay was one of the first methods used to assess genome edits in plants (33-35). This analysis is an easy, low cost and effective technique to test the presence of SNPs or INDELs, with a sensitivity of 2-5% (219). It is based in the elimination of a restriction enzyme recognition site in the position of the DSB induced by Cas9, therefore if a genome edit is introduced it will generate digestion-resistant products. For this assay, the edited region is amplified and then the PCR product is incubated with the appropriate restriction enzyme. Digestion products are analysed by gel electrophoresis and indel frequency can be estimated by a quantification of the decrease band intensity of the digested (WT) versus non-digested (edited) amplicons using ImageJ software (49). Developed in 1993, this technique distinguishes between wild type, heterozygous and homozygous genotypes, hence cell pools or clonal cell lines can be analysed using this method (219, 220). It was successfully used to determine genome editing events in complex genomes, such as *N. tabacum* and *N. benthamiana*, where duplicated copies of certain genes are found (137, 139, 142), or rapeseed, a triplicated diploid genome (221). However, the major disadvantage of the CAPS assay is the Cas9 cut site (3-4 bp upstream the PAM) being within a restriction enzyme site. In this thesis work, samples edited with TobFT4 guide targeting the *NtFT4* exon 4 gene couldn't be analysed, since the target site doesn't meet this criterion.

As an alternative approach, an enzyme mismatch cleavage (EMC) assay could have been performed to analyse regenerated shoots where the *NtFt4* exon 4 site was targeted (222). EMC protocols are simple and low cost, with a sensitivity similar to the CAPS assay (2-5%) (219). In the EMC analysis, the targeted region is amplified, and the resulting PCR product is denaturated and reanneal. If indels are introduced into the target site a heteroduplex DNA will be produced, generating ssDNA bubbles at the mismatch point. Reannealed samples are then incubated with specific endonucleases that cleavage heteroduplex DNA (edited), but no homoduplex DNA

(WT). As in the CAPS assay, editing efficiency can be calculated using ImageJ software, by the quantification of digested amplicons (edited) versus non-digested amplicons (WT) visualized by gel electrophoresis (49). Many endonuclease enzymes are commercially available to pursue this analysis, as CEL-I and CEL-II endonuclease (Surveyor®), T7 endonuclease I (T7EI), T4 endonuclease VII (T4E7) or endonuclease V (EndoV). A proper WT control must be included when performing this analysis, since naturally occurring SNP will also form an heteroduplex, leading to a false-positive result. Therefore, the amplified PCR region preferably shouldn't include an SNP (219). Alternatively, it has been reported that heteroduplex DNA migrate slower than homoduplex DNA in native PAGE electrophoresis, then the digestion step with the endonuclease could be skipped (223). Nonetheless, for CAPS and EMC assays, Sanger sequencing must be followed to determine the type of indels introduced into the target gene.

It was confirmed that the CAPS analysis is a sensitive method to determine genome edits by testing the assay using samples where genome edits had previously been confirmed by sequencing. However, when regenerated shoots from infected leaves using TRV or PVX and the different constructs to targeted genome edit *NtFT4* exon 2 were tested, all evaluated samples were WT. A total of 134 shoots were analysed (3 to 39 shoots per construct), where successful amplification and digestion was obtained for 125 shoots. Previous studies using shoot regeneration from virus infected leaves reported a similar number of analysed shoots. For example Ma *et al.* (2020) evaluated mutations in 30 shoots, where 17 shoots were homozygous and 28 shoots exhibited mutations of any type, while Ariga *et al.* (2020) assessed genome editing in 50 shoots, where 31 carried mutations in the target site (137, 139). Therefore, probably in these particular experiments, viral infection of these plants was not enough to induce genome editing events in the infected leaves. Particularly, our PVX vector is *in vitro* transcribed to produce viral RNA for infection, rather than using intact viral particles or an engineered vector for *Agro*infiltration, which could affect the efficiency of the vector. As discussed in CHAPTER 4, *Agro*infiltration of TRV is able to deliver the CRISPR/Cas9 components and introduce mutations in infected leaves, but even regenerated shoots from TRV-infected leaves didn't show genome edits. These results contrast with previous reports, where high efficiency of mutations was found in regenerated plantlets from *Agro*-infiltrated PVX infected

leaves (134, 139). Nevertheless, the analysis of a higher number of shoots could potentially detect genome edits in the *NtFT4* exon2 target, while an EMC assay could be used to assess genome edits in the *NtFT4* exon 4 target site.

Previous studies also confirmed the presence of virus in regenerated shoots, either by RT-PCR to detect viral mRNA or by the presence of viral symptoms. Interestingly, when the progeny of these regenerated plants was analysed, the introduced mutations were inherited, however no virus was detected, confirming that PVX and SYVN aren't transmissible through seeds (134, 137, 139). Hence, inherited mutations can be obtained by regeneration from somatic cells, indicating the development of a DNA-free genome editing approach. These types of analysis weren't conducted in this research. Therefore, further experiments are necessary to support these findings.

Plant regeneration from either protoplasts or leaves is a standard procedure in plant biotechnology. However, it has some disadvantages since it's time-consuming, laborious, and only works for some species/cultivars/accessions of species. To overcome this problem, the use of highly conserved morphogenetic regulators to induce meristem morphogenesis in somatic cells has been reported. Maher *et al.* (2020) described the generation of genome edited plants by *de novo* meristem induction, where developmental regulators and genome editing reagents are delivered into somatic cells, which develop into meristems that generate edited shoots in approximately two weeks. The authors tested this technology in *N. benthamiana*, potato, tomato and grapes indicating a broad range of species for its application, decreasing the time and work for obtaining new edited crop varieties (224).

CHAPTER 6: GENERAL DISCUSION

This thesis proposes the development of a non-integrative CRISPR/Cas9 system using plant virus vectors to introduce target genome editing into the floral inducer *NtFT4* of *N. tabacum* var. Maryland Mammoth. Particularly, the use of non-integrative (+) ssRNA virus vectors to express the CRISPR/Cas9 was investigated. Since virus vectors have a limited cargo capacity, the utilisation of a split Cas9 approach using two different Cas9 orthologs was tested.

6.1 GENOME EDITING OF COMPONENTS OF THE FLOWERING TIME PATHWAY

Horticultural crops, such as fruits, vegetables, and ornamental plants (floriculture) are a key component of the agriculture production system. Most food crops are flowering plants, constituting a major component of the human diet. Changing the seasonal timing of reproduction is a long-term goal of plant breeding to produce new varieties that can adapt better to shifting climatic conditions, local environments, and higher demand to feed the rising global population (152). For example, early flowering is desirable in cereals to prolong their grain filling phase and avoid certain climate conditions that can affect production. On the other hand, in plants used for biofuel production, animal feeding (forage crops) or crops grown for their leaves, delayed flowering is preferable to achieve higher biomass yields. Also, a delayed flowering phenotype is wanted for certain types of trees that grow in cold environments which can damage their flowers and fruits, but in others where the vegetative phase lasts many years an acceleration of flowering would be advantageous (152). Using conventional breeding programmes, such as hybridization and mutational breeding, many desirable traits have been introduced into crops. However, a limited gene pool available for new traits in some species, high heterozygosity, or the inability to produce seeds of some cultivars highly restrict the applicability of such methods (225). Therefore, a major aim of plant biotechnology is to develop new methods to maximise the production of horticultural crops in a sustainable manner. In recent years, the development of several genome editing tools, but particularly the CRISPR/Cas9 system, has enabled the introduction of desirable agronomic traits in an efficient, simple, and low-cost way.

There are numerous examples of the use of CRISPR/Cas9 technology to manipulate the flowering time in crops. Considering that apple trees flower after 4 – 8 years (152), Charrier *et al.* (2019) edited the apple *TFL1* gene, a flowering repressor, to induce an early flowering phenotype (226). The expected phenotype was observed in 93% of the transgenic lines analysed. Furthermore, using the same constructs the authors edited the pear *TFL1* gene, however a lower efficiency was found (9% of the transgenic lines were edited), most probably due to the presence of a mismatch between the sgRNAs tested and the target site or because both *TFL1* homolog genes (*PcTFL1.1* and *PcTFL1.2*) must be mutated for a complete release of the floral repression (226). Chinese cabbage (*Brassica rapa* spp. *pekinensis*) is an important vegetable crop, source of dietary fibre, vitamins, and minerals. Late flowering Chinese cabbage has been produced by gene editing, to avoid its premature bolting, which affects crop yield and quality (227). The successful targeted mutagenesis using four sgRNAs of *VERNALIZATION 1* (*BrVRN1*) gene, a repressor of FLC to promote flowering, produced a late-bolting phenotype compared to WT plants under vernalization conditions (227). Additionally, early flowering cultivars of Chinese cabbage have been generated, that in contrast with the previous study, don't require vernalization (228). Genome editing of double knockout lines of *BraFLC2* and *BraFLC3* genes showed an early flowering phenotype that did not depend on vernalization (228). In tomato, genome editing of the *SELF-PRUNING 5G* (*SlSP5G*) gene, a flowering repressor, conferred rapid flowering and a quick burst of flower production that translates into an improved fruit yield (229). The targeted mutagenesis of two kiwifruit *CENTRORADIALIS*-like (*CEN*) genes, *AcCEN4* and *AcCEN*, transformed a climbing woody perennial which grows axillary inflorescences after many years of juvenility into a compact plant with rapid terminal flowering and fruit development (230). Interestingly, a green bristlegrass (*Setaria viridis*) variety with delayed flowering time obtained by deactivating the *S. viridis* homolog of the maize *ID1* gene is being exempted of GMO regulations in the US, being in the pipeline for commercial release (231). Epigenetic regulation of flowering has also been described. In *A. thaliana* histone H3 lysine 4 trimethylation (H3K4me3) and H3K36me3 are involved in the activation of *FLC* expression, while H3K4 demethylation, H3K9me2, H3K27me2/3 and histone arginine methylation are involved in the repression of *FLC* expression. The methyltransferase SET DOMAIN GROUP 8 (SDG8) controls the methylation level of H3K36 at the *FLC* locus (232).

Jiang et al. (2018) characterized two homologs of *SDG8* in *Brassica napus*, *BnaSDG8.A* and *BnaSDG8.C* genes and showed that the knockdown of these genes by CRISPR/Cas9 led to an early flowering phenotype compared with the WT control, related to a reduced levels of H3K36me2/3 and an increased levels of H3K36me1 at the *FLC* loci (232).

These examples show the complex network that regulates flowering in plants and illustrate how the editing key components can result in beneficial agronomical traits. An integrator of these pathways is FT, a protein synthesized in the companion cells of the leaves that moves to the shoot apex where floral transition occurs (156, 157). In sorghum, *SbFT* gene mutants generated by CRISPR/Cas9 exhibited a delay in the flowering time of eight to ten days compared with the WT plants (233). Soya bean is an important legume crop, sensitive to seasonal changes in day length, which limits its geographical range of cultivation (234). Homozygous T1-line mutants with a frameshift of 1 bp in the *GmFT2a* gene generated by CRISPR/Cas9 showed a late flowering phenotype under natural conditions (summer) in Beijing, China, while the T2 generation from these plants exhibited late flowering under both long-day and short-day conditions, broadening the possibilities of regional introduction (234). In *N. tabacum*, several FT proteins have been described, where some of them act as floral repressors (NtFT1, NtFT2 and NtFT3) and others as floral inducers (NtFT4 and NtFT5) (163). The knock-out of the *NtFT5* gene by CRISPR/Cas9 prevents LD flowering in *N. tabacum* cv. SR1 plants, while heterozygous $Ntft5^{-}$ /$NtFT5^{+}$ plants showed a delay of flowering of around 2 days, conferring traits such as an increase in the vegetative leaf biomass, the production of more seeds and better performance under abiotic stress (165).

In this thesis work, the targeted CRISPR/Cas9 editing of the *NtFT4* gene of *N. tabacum* var. Maryland Mammoth was tested. Most varieties of tobacco plants are day neutral, however the Maryland Mammoth variety flowers exclusively under SD conditions, producing 1 – 2 kg of leaf biomass per plant prior to flowering (235). In contrast to NtFT5, which is expressed regardless of the day length, the floral inducer NtFT4 is expressed mostly under SD.

Targeted genome edits were successfully introduced into the *NtFT4* gene in *N. tabacum* protoplasts and plants. *Nt*FT4 is a floral inducer belonging to the PEPB

protein family. Apart from the previously mentioned NtFTs proteins, eight more NtFTs has been identified, name as NtFT6 to NtFT13, but their function as floral inducer or repressor could not be predicted (159, 162, 163). *N. tabacum* FT proteins are closely related, with their amino acid sequences having an identity >70%. For example, NtFT1 and NtFT3 are 89% similar, NtFT3 is 70% similar to NtFT4, and NtFT5 is 76.8% similar to NtFT4 (159, 162). Considering this high degree of conservation, the design of guide RNAs specific to the *NtFT4* gene is important to avoid off-target activity of the CRISPR/Cas9 system. When designing the gRNAs, some online tools such as Cas-OFFinder indicate the number of mismatches in potential off-targets, where gRNAs with zero to two mismatches in off-target sites should be selected (51, 167, 168). Moreover, mismatches closer to the PAM site seem to be more critical than mismatches at distal positions (43, 44). It is suggested that to reduce off-target activity a perfect complementation between the target site and the gRNA in the 7-12 bp closest to the PAM (seed sequence) should be considered (46). In this work, the possible off-target activity of TobFT1 and TobFT4 gRNAs in other *NtFT*s genes was analysed by sequence alignment. A sequence identity of 100% was found in the gRNA target region of exon 2 between *NtFT4*, *NtFT6* and *NtFT11*, however *NtFT6* gene carries a nonsense mutation, hence is inactive (163), while *NtFT11* exhibited a mismatched base in the PAM site for both Cas9. No off targets were detected for TobFT4, since *NtFT4* and other *NtFT*s exhibited multiple mismatched bases in the exon 4 target region. Even though in this thesis the presence of off-target mutations in edited plants wasn't analysed, Kaya *et al.* (2016) reported that using *Sa*Cas9 and the same guide used to target the exon 2 of *NtFT4* (TobFT1), no off-targets were found in the *NtFT1* (Two mismatches) or *NtFT2* (Four mismatches) genes (142).

Special attention has been given to the off-target activity of the CRISPR/Cas9 system, particularly in humans where several studies have been conducted to understand the specificity of the system because of its potential use for gene therapy (43, 44, 46). In plants this has been less studied, mainly because undesired mutations can be removed or segregated away by backcrossing. Some studies suggest a low off-target activity of Cas9 in plants due to the usually low levels of expression of the Cas9 protein (56). Nonetheless some strategies proposed to reduce the off-target activity of the CRISPR/Cas9 system in plants include the use of transient expression

systems (especially RNP complexes), Cas9 orthologs with more specific PAMs, shorter gRNAs, lower concentrations of delivered Cas9/sgRNA, exposing plants to higher temperature conditions, or the use of "high fidelity" engineered Cas9 proteins (55, 56).

6.2 PLANT VIRUS VECTORS AS A NON-TRANSGENIC APPROACH FOR GENOME EDITING

Genome editing using CRISPR/Cas9 is a valuable tool to develop new crop varieties with improved agronomical traits. However, the major bottleneck to use this technology in plants is the delivery of the CRISPR reagents. Larger and complex genome structures due to common occurrence of polyploidy and other genome re-arrangements and rigid cell walls surrounding the plant cells hinder the efficient entry of external molecules (236). In plants, foreign genes are mostly introduced and stably integrated into the genome via *Agrobacterium*-mediated transformation (145, 236). However, in the case of CRISPR/Cas9, this can lead to off-target effects due to the constitutive expression of the components or even plant death. To avoid GMO constraints, transgene-free edited plants can be obtained by segregation, but for vegetatively propagated plants this is not feasible (145, 236). As an alternative, the use of transient systems, such as autonomously replicating plant virus-based vectors offers a bypass to the stable integration of the CRISPR reagents.

The delivery of the CRISPR components by viruses engineered as vectors offers various advantages. (+) ssRNA virus do not integrate their genome into their plant host genome, thus allowing the development of non-integrative vectors. ssRNA viruses replicate in the cytoplasm of the host cell using their own RNA-dependent RNA polymerases. All the viral proteins necessary for the replication of (+) ssRNA are directly translated upon infection of the host cell. In contrast, (-) ssRNA virus requires that the RdRp protein copies the (-) RNA into (+) RNA for its translation (237). Another advantage of using viruses as vectors is that they spread systemically throughout mature plants within days, expressing transgene products in a short period of time rather than weeks or months as in traditional stable transformation methods (107). In this thesis work, PVX and TRV were used to deliver the *Cas9* gene and sgRNAs into *N. tabacum* plants. It was established that *N. benthamiana* plants infected with PVX and TRV viral vectors developed systemic symptoms

within 7 to 10 dpi, hence they are a suitable option to deliver heterologous genes of interest. To increase the likelihood of delivering the sgRNA into the same cell as the Cas9 protein, a single transcriptional unit was created by fusing self-cleavable ribozymes to each end of the sgRNA to create the RGR unit, which was positioned downstream of the *Cas9* gene. The self-cleavage activity of the HHRz to release mature sgRNAs was demonstrated *in vitro* and *in vivo* by cRT-PCR. 3'-HHRz scar was detected in both conditions, while incomplete cleavage of the 5'HHRz was found in *N. tabacum* leaves. Complete guide RNAs were detected in all the samples analysed, while full-length RNA scaffold sequences (76 bp) were noticed in most of the samples but truncated 3'-ends sequences were also found probably due to RNA degradation during sample preparation (101).

However, the use of plant viruses to deliver the CRISPR reagents offers many challenges. It has been reported that since the RdRp protein lacks proofreading activity, RNA viruses exhibit high mutations rates (10^{-4} to 10^{-6} mutations per base pair per generation). Any mutations that interfere with viral replication will eventually be eliminated from the viral population (237).

Viral cross protection is a phenomenon that occurs when plants are pre-infected with a mild variant of the virus to prevent secondary infections with more severe viruses of the same, or closely related families but remains susceptible to infection by more distantly related families. Initially, it was suggested that its mechanism of action was through RNA silencing, but recent studies using *Turnip crinkle virus* also revealed a protein-based mechanism (140, 141). It has been reported that the expression of two transgenes by co-inoculation of two TRV2 viruses resulted in less efficient expression of both of them (125). Even though, in this thesis work when the split Cas9 approach was tested, mRNA expression of both Cas9 domains was seen when the same virus, or a combination of TRV and PVX were co-inoculated. Similar results were obtained when two FoMV viruses expressing the sgRNA and the Cas9 gene separately were co-inoculated, where no cross-protection was detected (136).

The major disadvantage of using viruses as vectors is their limited cargo capacity. For example, it has been reported that for TRV the cargo capacity is between 2 – 3 kb, while for PVX the longest successful gene expressed is ~2 kb (107, 129, 135). This is related to a negative correlation between the size of the insert and the stability

of the vector (110). *Sp*Cas9 full-length gene is ~4.2 kb, whilst the *Sa*Cas9 full-length gene is ~3.3 kb, and in combination with the RGR unit they greatly exceed the putative size limit of both viral vectors. Western blot analysis showed that it was possible to detect full-size *Sp*Cas9 protein delivered by TRV in infected leaves, suggesting that at least in inoculated leaves the vector can express the protein. Recent papers reported the successful delivery of a functional full-length *Sp*Cas9 by the (+) ssRNA viruses FoMV and PVX, and the (-) ssRNA rhabdoviruses SYNV and BYSMV, resulting in mutation of a target gene (136-139). However, in our experiments no full-length *Sp*Cas9 protein was detected in PVX infected leaves, and no full-length *Sa*Cas9 protein was identified in inoculated leaves with either TRV or PVX. In contrast, it was possible to detect the presence of full-length *SpCas9* and *SaCas9* in TRV or PVX inoculated tobacco leaves at the mRNA level. As mentioned above, large inserts reduce viral replication and expression levels and are more susceptible to the loss of the transgene as early as the first infection passage (110, 111). For instance, plants infected with PVX-GFP lost the transgene progressively and over time, whereas only the WT recombinants prevailed (111)

To overcome the limited cargo capacity problem, a split Cas9 approach was tested. Several papers have described this approach for genome editing, splitting *Sp*Cas9 or *Sa*Cas9 (80, 81, 143). Additionally, to make the system more precise, the fusion of inteins to each half of Cas9 protein has been proposed, which facilitates the functional reconstitution of the protein by the formation of a peptide bond between each end (83). As a proof of concept, split *Sp*Cas9 fused with inteins and split *Sa*Cas9 were delivered by *Agro*infiltration of plant expression vectors into *N. benthamiana* plants, or by PEG transformation into *N. tabacum* protoplasts. Immunoblot results of the transformed tissue showed the presence of the individual split *Sa*Cas9 domains, while for *Sp*Cas9, a reconstituted full-size Cas9 protein was detected. The functionality of the re-assembled proteins was assessed by targeted genome editing of the *PDS* gene in *N. tabacum* protoplasts. Indels in the target region were detected in both cases. Both re-assembled Cas9 proteins exhibit a similar genome editing frequency to their respective full-length versions. Next, *N. tabacum* plants were infected with TRV, PVX or a mixture of both vectors harbouring each N- and C- terminal domains fused to different RGR genes (TobFT1 and TobFT4). Only the presence of the *Sa*Cas9 C-terminal domain was observed by

Immunoblot analysis in inoculated leaves with PVX or TRV where each fragment was delivered by a combination of both viruses, but mRNA for each split domain was detected in all tested samples from inoculated leaves.

The presence of systemic symptoms in TRV viral infections, but no expression of full-length or split *SpCas9* mRNA (Figures 26 B and 28 B) was determined, whilst weak expression of full-length and N-terminal *SaCas9* mRNA (Figures 40 B and 42 B) and strong detection of *SaCas9* C-terminal mRNA (Figures 40 B and 42 B) was found. In the case of PVX, no systemic symptoms were seen in any of the analysed plants, but a faint band was identified when *Sa*Cas9 C-terminal domain was delivered (Figure 42, lane 5). These results suggest that an ideal cargo size for TRV and PVX should not exceed 1.3 kb to avoid a loss of the insert during viral replication events as the virus moves through the plant, as reported by Avesani *et al.* (2007) (110).

As an alternative to obtain fully edited plants, shoot regeneration from viral inoculated leaves was attempted. *Cas9* mRNA expression was confirmed in infected leaves with PVX and TRV Cas9-RGR constructs and targeted genome edits were detected in TRV infected leaves. Virus vectors infect each cell independently, thus different shoots derived from different infected cells will show different types of genome editing. Shoots were successfully regenerated from leaves inoculated with PVX or TRV full-length, or split, Cas9-RGR constructs, and these shoots were screened for targeted genome edits by a CAPS assay. This technique distinguishes between wild type, heterozygote and homozygote genotypes and it's broadly used as a first step to screen for mutations in target sites (220). Targeted genome editing was assessed successfully in shoots regenerated from inoculated leaves with PVX or TRV full-length, or split, Cas9-TobFT1, but all tested samples were WT.

Recently, two smaller Cas9 orthologs have been described and characterized. CasΦ or Cas12j (700 – 800 amino acids), a type V CRISPR/Cas protein, was isolated from *Biggiephages* and successful genome editing was observed *in vitro* and *in vivo* (in HEK293 human cells and *A. thaliana* protoplasts) where RNP complexes were delivered (74). Cas14 or Cas12f, another type V CRISPR/Cas protein, is a protein of 400 – 700 amino acids isolated from uncultivated *archaea.* Genome editing using this nuclease has been reported in maize cells (238). Cas12f and gRNAs targeting

the *male sterile 26* and *waxy* genes were delivered by biolistic into immature maize embryos. After 24 hours, embryos were incubated at 28°C, 37°C or 45°C for 4 h once per day for a total of three days. Successful targeted mutagenesis at both genes was detected only in the 45°C-treated embryos, indicating that Cas12f has a preferential activity at higher temperatures (238). Recently, the delivery of Cas12f by a modified PVX vector and its sgRNA by TRV has been described. The co-inoculation of these components into *N. benthamiana* plants challenged with the ssDNA Multan or Krokan viruses, was able to confer systemic resistance against them by the targeted mutagenesis of the viral genomes. Thus, these results indicate that Cas12f may meet the cargo capacity of PVX (239).

Alternatives to the use of plant virus vectors to obtain non-GMO edited plants have been proposed. The transformation of RNP complexes into tobacco, *A. thaliana*, lettuce, rice, petunia, grapevine, apple and potato protoplasts and rice zygotes by PEG-transformation and into maize and wheat somatic embryos by particle bombardment have proved to be successful to generate DNA-free genome edited plants (240). But protoplasts, callus and somatic embryo regeneration protocols have been developed only for some species and are demanding, costly, time-consuming, and laborious processes (145). In contrast, methods to edit fully-grown plants using a non-GMO approach have been published. Haploid induction-mediated method uses one plant as a "delivery vehicle", where the CRISPR reagents are delivered into a haploid inducer line, which is crossed with a WT plant. After fertilization, the haploid inducer genome, harbouring the CRISPR components, will introduce targeted edits into the WT genome. Later, the chromosome from the haploid inducer, is eliminated, and diploid non-GMO plants can be obtained by diploid induction (145). An additional method is the particle bombardment of a transient expression plasmid with the CRISPR components into the L2 cell layer of the shoot apical meristem for germline transmission, generating edited non-GMO lines (241).

6.3 NEW GENOME EDITING TOOLS FOR PRECISE TARGETED MUTAGENESIS

Thus far, most of the CRISPR/Cas9 reports in horticultural crops implicate the NHEJ DNA repairing mechanism to introduce targeted mutagenesis. As a step further to introduce precise changes into the genome, new editing technologies have

been developed. In cells, a DSB undergoes two mechanisms of repairing, NHEJ, which generates indels randomly, or HDR where the cleavage is repaired using a template. NHEJ is more frequent than HDR, since the first takes place during the whole cell cycle, while the second only occurs during S and G2 phases (2). Several methods using CRISPR/Cas9 have been developed to induce the HDR pathway to introduce precise insertions, sequence replacements and nucleotide substitutions using donor DNA templates. However, a low editing efficiency using HDR repairing tools has been reported in plants (54, 242). In contrast, the use of base editors and prime editing has been reported broadly across plant species (54) and it has been thoroughly reviewed by Molla *et al.* (2021) (243)

Base editors allow to precisely change one DNA base into another without using donor templates or involving DSBs. This is particularly useful to avoid frameshifts due to the introduction of random nucleotides to repair the DSB and it has proven to be more efficient as an editing system in plants than the HDR repairing system (54). Deaminase-mediated base editors are a fusion between a Cas9 nickase (nCas9, mutated at the RuvC I domain, D10A) and a deaminase protein, where nCas9 and the sgRNA guides the deaminase to the target site.

Cytosine base editing (CBE) creates C:G>T:A substitutions, where a cytidine deaminase converts cytosine to uracil, which is repaired as thymine by the DNA repair mechanisms (244). It is also possible to introduce non-sense mutations, generating knockout genes using this system (54). Improvements to the system have been introduced by using cytidine deaminases and Cas9 orthologs. Most of the studies reported the use of a rat cytidine deaminase (rAPOBEC1) and a *Petromyzon marinus* cytidine deaminase 1 (*Pm*DCA1), however cytidine deaminases such as human activation induced cytidine deaminase (AID), human APOBEC3A or human APOBEC3B had been reported in order to increase the base editing efficiency (54, 243). To broaden the targeting scope, several Cas9 orthologs have been used to improve the system, such as *Sa*Cas9, *Sc*Cas9 or Cas12a (243). Targeted genome editing using CBE has been reported for many plant species, such as rice, maize, wheat, *A. thaliana*, potato, tomato, watermelon, pear, apple, strawberry, cotton, soybean, moss, poplar and rapeseed (243).

On the other hand, adenine base editing (ABE) converts A:T>C:G substitutions by the use of an adenine deaminase. The enzyme deaminates adenosines to inosines, recognized as guanosines by the DNA polymerase during repair and replication. However, no enzymes have been described that can naturally deaminate adenine in DNA. Instead, Gaudelli *et al.* (2017) engineered an *E. coli* tRNA-specific adenosine deaminase (*ec*TadA) to operate on DNA using directed evolution (245). Improvements to the ABE system have been made through codon optimization of the *ec*TadA and the addition of a bipartite NLS. A miniABE system was developed using a monomer version of *ec*TadA, instead of a dimer version. Similar to the CBE system, Cas9 orthologs have been used to expand the targeting scope (54, 243). In plants, the use of this system has been reported in *A. thaliana*, rice, wheat, rapeseed, *N. benthamiana*, moss and poplar (243).

Based on these systems, some further improvements have been made. Dual base editors are a fusion between CBE and ABE and are known also as Saturated Targeted Endogenous Mutagenesis Editor (STEME). In the CBE system, besides a nCas9 and the cytidine deaminase, an uracil DNA glycosylase inhibitor is fused to prevent the formation of apyrimidinic sites and subsequently the creation of single strand breaks by lyases. In contrast, CBE-based precise DNA deletion (or AFIDs for APOBEC–Cas9 fusion-induced deletion systems) used a WT Cas9 fused to an uracil DNA glycosylase to create a combination of a DSB and a nick, resulting in a precise deletion between the deaminated cytidine and the Cas9 cut site (54). C-to-G and C-to-A transversion base editing have been described in mammalian cells (C-to-G) and bacterial cells (C-to-A). The system comprises a nCas9, a cytidine deaminase and an uracil DNA glycosylase, where C is deaminated to U by the cytidine deaminase, and an apurinic/apyrimidinic site is created by removing that U, which is replaced by a G or A (246, 247). Deaminase-mediated base editors have been also engineered as split proteins, either splitting the cytosine deaminase enzyme (248), or the nCas9 fused to an adenine deaminase (249).

Furthermore, a new technology, called Prime Editing (PE) performs all 12 types of base substitutions (four transitions plus eight transversions), precise insertions of up to 44 bp, deletions of up to 80 bp and a combination of these mutations. This tool was developed by the fusion of a nCas9 (H840A) and a reverse transcriptase, which are guided to the target site by a prime editing guide RNA (pegRNA). This is formed

by the sgRNA that is extended at its 3'end with a reverse transcriptase template containing the information for the mutation, and a primer-binding site which pairs with the nicked ssDNA produced by nCas9 and "primes" the reverse transcriptase to incorporate the mutation (250). Prime editing has been successfully used in plants species such as *A. thaliana*, *N. benthamiana*, potato, tomato, rice, maize and wheat, but its editing efficiency in plants remains limited compared to the editing efficiency achieved in humans (54, 243).

Altogether, the CRISPR/Cas9 system is rapidly evolving and its combination with other enzymes, such as reverse transcriptase, or deaminases, make it the most attractive option to induce targeted genome edits in a variety of crops.

6.4 CROP LEGISLATION FOR CRISPR/CAS9 EDITED PLANTS

Considering the several advantages of CRISPR/Cas9 in creating beneficial agronomical traits in crops, the development of DNA-free genome editing systems is in high demand, especially in cultivars where exogenous DNA cannot be removed by segregation such as vegetatively propagated elite cultivars, or highly heterozygous hybrid cultivars (139). Moreover, since no DNA or integration into the host genome is involved, under some country's regulations they could qualify as non-GMO crops. Even though GMO crop legislation is different in each region, all of them use the definition described in the United Nations Cartagena Protocol on Biosafety of a "living modified organism" as a guideline, where a plant is classed as genetically modified if meets two criteria: "1) The plant contains a novel combination of genetic material, and 2) which was introduced by using modern biotechnology, legally defined as an application of either *in vitro* nucleic acid techniques (which includes recombinant DNA and direct injection of nucleic acid into cells or organelles), or the fusion of cells beyond the taxonomic family" (150). Worldwide GM regulations are divided between process- or product-oriented regulations. The first oversees the processes used to produce the new variety, while the second assesses the new traits of the product in comparison to those that could be obtained by conventional breeding (150). For example, Canada's legislation of GMO crops follows a product-oriented approach where "Whether the novel trait was developed by conventional breeding techniques, traditional mutagenesis, or targeted mutagenesis, the novel plant product is subject to the same Canadian Food

Inspection Agency risk assessment regulations". Hence, the emphasis is on the safety of the new characteristic in the final product rather than the technique used to introduce this new trait (150). On the other hand, the European Union follows a process-oriented regulation. Article 2(2) of the Cultivation Directive defines a GMO crop as "An organism, with the exception of human beings, in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination" (150). Moreover, in 2018, the European Court of Justice included the CRISPR/Cas9 system into this definition, indicating that "Organisms obtained by mutagenesis are GMOs and are, in principle, subject to the obligations laid down by the GMO Directive" (251). This contrasts the announcement by the US Department of Agriculture (USDA) indicating that they will not regulate edited crops with mutations that could have occurred in nature; similar decisions were made by Brazil, Argentina and Australia (252). In the US, in 2016, the Animal and Plant Health Inspection Service, an agency under the USDA, granted a non-regulated status to a CRISPR/Cas9 white button mushroom (*Agaricus bisporus*) engineered to be resistant to browning. This was the first CRISPR-edited crop that could be cultivated and sold without being regulated by the agency (149). Currently, a high-oleic-acid soybean oil, a herbicide-tolerant canola, and a waxy corn with enriched amylopectin, are genome edited products that are now being commercialized, while commercial trials are starting for genome edited lettuces resistant to browning (150, 252).

Most scientists claim that genome editing techniques shouldn't be considered under GMO regulations, since the changes are induced directly into the genome of the organism (so no foreign DNA), and so these methods are not different to changes that can be introduced during conventional breeding or through natural mutation (150). Therefore, researchers are calling for an update to the current European regulation. In April 2021, the European Commission published a report regarding New Genomic Techniques (NGTs), such as genome editing, showing the potential of these tools to produce food in a more sustainable way. The study also indicates that the current European GMO legislation is no longer fit for purpose for some NGTs and their products, thus it needs to be adapted to recent scientific and technological progress. Considering these findings, the Commission will start an open consultation process to discuss and design a new regulation for these methodologies (253).

In the UK, new regulations are being introduced by the Government, such as a statutory instrument to "exempt genome edited crops with changes that could have been achieved via conventional breeding, or which could occur naturally, from the GMO regulation for field trials in England" and a new primary legislation to "amend the regulatory definition of a GMO to exclude these crops" (254). Moreover, a new Bill is being discussed in the Parliament. The Genetic Technology (Precision Breeding) Bill will "make provision about the release and marketing of, and risk assessments relating to, precision bred plants and animals, and the marketing of food and feed produced from such plants and animals; and for connected purposes" (255).

6.5 CHALLENGES AND THE FUTURE OF THE CRISPR/CAS9 SYSTEM FOR TARGETED GENOME EDITING OF HORTICULTURAL CROPS

Genome editing technologies have a great potential to boost the production of horticultural crops. Particularly, the CRISPR/Cas9 system offers an easy, cost-effective way of obtaining new varieties with desirable agronomical traits. CRISPR was developed as a genome editing technology in 2013 and 10 years later its use has been proved to be successful in a variety of species, from model plants like *A. thaliana*, to both monocots (maize, rice, wheat) and dicots (tomato, potato, soybean) (256). Most of these genome edits relies on "loss-of-function" of the gene, by the introduction of random mutations by the NHEJ pathway. As discussed in section 6.3, the development of new genome editing technologies, such as base editors or prime editing, will allow to fine-tune a gene of interest. In recent years, new approaches propose to modulate the expression of a gene at different levels, not only by altering its sequence. The transcription of a gene of interest can be regulated using a dCas9 to block the binding of the transcriptional machinery or the transit of RNA Polymerases to repress the transcription (257, 258). Precise regulation can be achieved by the fusion of dCas9 to transcriptional regulators or epigenetic modulators (54). For example, the fusion of the DNA methyltransferase 3α (DNMT3A) catalytic domain, dCas9 and a SunTag array can alter the DNA methylation pattern of a desired loci (259). A different approach is to directly target factors involved in epigenetic regulation, such as the *SDG8* gene in *Brassica napus* discussed previously (232). At RNA level, editing is possible using new Cas9 orthologs, such as Cas13a (77) or by modulating pre-mRNA splicing, either by editing splicing motifs to prevent it, or

alternative splicing sites, enhancing or decreasing one variant over another (54). The editing of upstream reading frames (uORFs) at the 5' untranslated regions of mRNAs allow efficient translation increase. Usually, these regulatory regions reduce the translation of the downstream ORF and promotes mRNA degradation (54). In strawberry, editing using CBE of b*ZIP1.1* uORF, enhanced its transcription and increase the sweetness of the berry (260).

To date, many plants of agronomical importance haven't been sequenced or their genome sequencing results need improvements, such as avocado, olive, or tropical crops, thus it is not possible to run genome editing approaches. New, cheaper, and easier sequencing technologies may help to overcome this obstacle (261). Polyploidy is another challenge when performing genome editing in crops of interest. To obtain the desired phenotype, mutation in two or more copies of the target gene is necessary. The careful design of a single sgRNA or the delivery of multiple sgRNAs will allow to target every allele variant (261). Moreover, a multiplex sgRNA array could target different sites at the same time, enabling the modification of several quantitative trait loci or complex pathways, such as flowering time (54).

As discussed extensively in this thesis work, one of the major bottlenecks of using the CRISPR/Cas9 system in plants is the delivery of its components. (+) ssRNA viruses offer an alternative to solve this setback, obtaining also non-GMO crops. New Cas9 orthologs, such as CasΦ or Cas14, are great options to meet the required cargo capacity of the viruses, as shown in a recent study by Haider *et al.* (2022) (239).

Altogether, the CRISPR/Cas9 system is rapidly evolving, and its drawbacks are being overcome with new studies and developments. Future uses of CRISPR/Cas9 in plants could help to domesticate orphan crops or wild species, accelerate breeding programs creating novel genetic variations, or alter the metabolism of crops of interest in a faster and easier way than before (54, 256), with the promise to meet the requirement of food by the raising global population.

6.6 CONCLUSIONS

- As a DNA-free genome editing methodology, a system to deliver the CRISPR/Cas9 components using the (+) ssRNA plant virus vectors TRV and PVX was developed. Considering the limited cargo capacity described for virus vectors, full-length and split Cas9 orthologs were tested using this viral delivery system.

- Split *Sp*Cas9 and *Sa*Cas9 protein domains were shown to successfully re-assemble and be functional in introducing edits into the *PDS* gene in *N. tabacum* protoplasts

- A *Cas9* gene and ribozyme-flanked sgRNA unit was created to be delivered from a single virus vector. Self-cleavage activity of the ribozymes was confirmed both *in vivo* and *in vitro*, generating mature sgRNAs to guide the Cas9 to its target site.

- Both PVX and TRV viruses can deliver the Cas9-RGR unit into infected plant cells. Full-length *Sp*Cas9 was detected at the protein level only in TRV inoculated leaves, but not in systemic leaves. In contrast, expression of *Cas9* mRNA was confirmed in all PVX and TRV inoculated leaves analyzed.

- Gene edits were detected in leaves infected with TRV vectors expressing both full-length and split Cas9 proteins. Rate of events are between 3.03% to 0.29%. Gene edits remain to be analysed in plants infected with PVX and combined PVX/TRV.

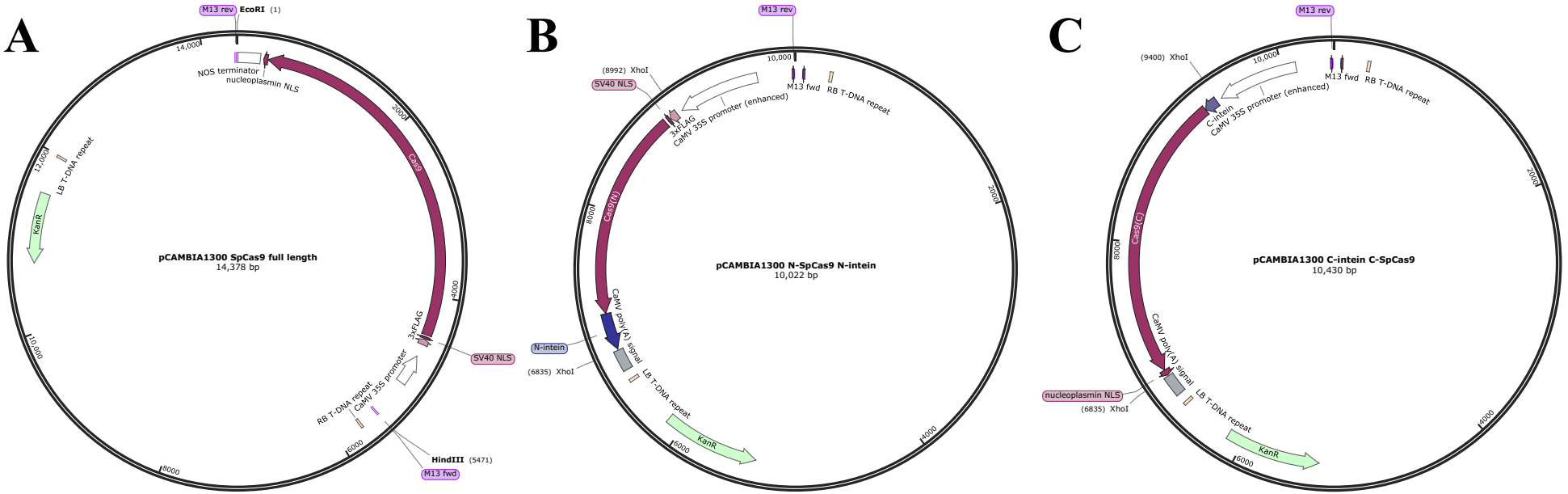
- Systemic symptoms of TRV infection were seen. No full-length or split *SpCas9* RNA was found, while weak expression of full-length and N-terminal *SaCas9* RNA was observed. In contrast, *SaCas9* C-terminal was strongly detected in TRV systemic leaves, whilst weak expression was found in one sample from PVX systemic leaves. Hence, a cargo below 1.3 kb is suggested to avoid a loss of the transgene for viral self-stabilization during its movement through the plant.

- *N. tabacum* shoots were regenerated from virus inoculated leaves with full-length or split Cas9 constructs, but no genome editing was detected in the analyzed plants.

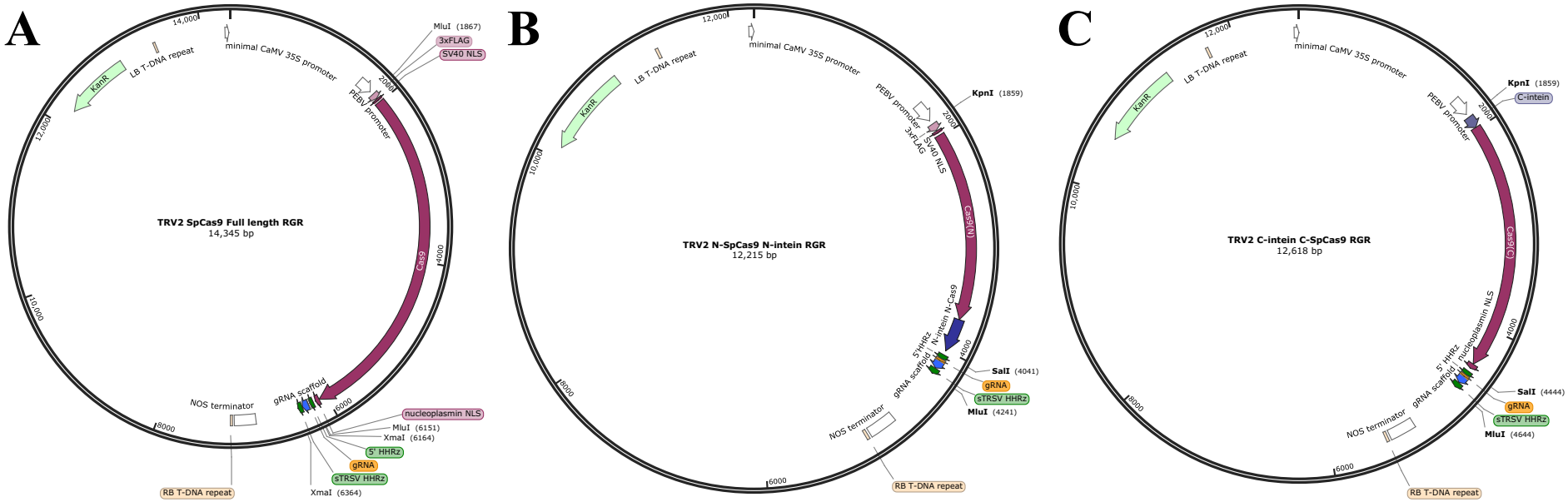
- This work demonstrated the feasibility of using virus vectors to deliver the CRISPR/Cas9 components into *N. tabacum* plants, introducing targeted mutations into the *NtFT4* target gene. In the future, the use of this technology will allow the develop of an alternative, non-integrative gene editing approach to use in crops.

APPENDIX A. SUPPLEMENTARY FIGURES

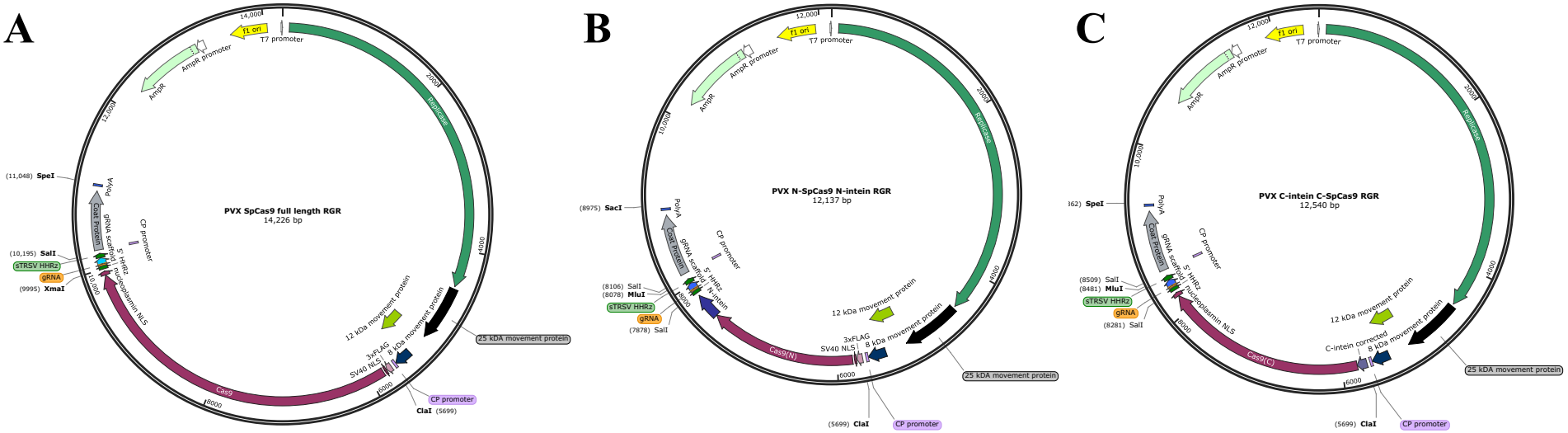
Appendix A1. Vector maps and sequences



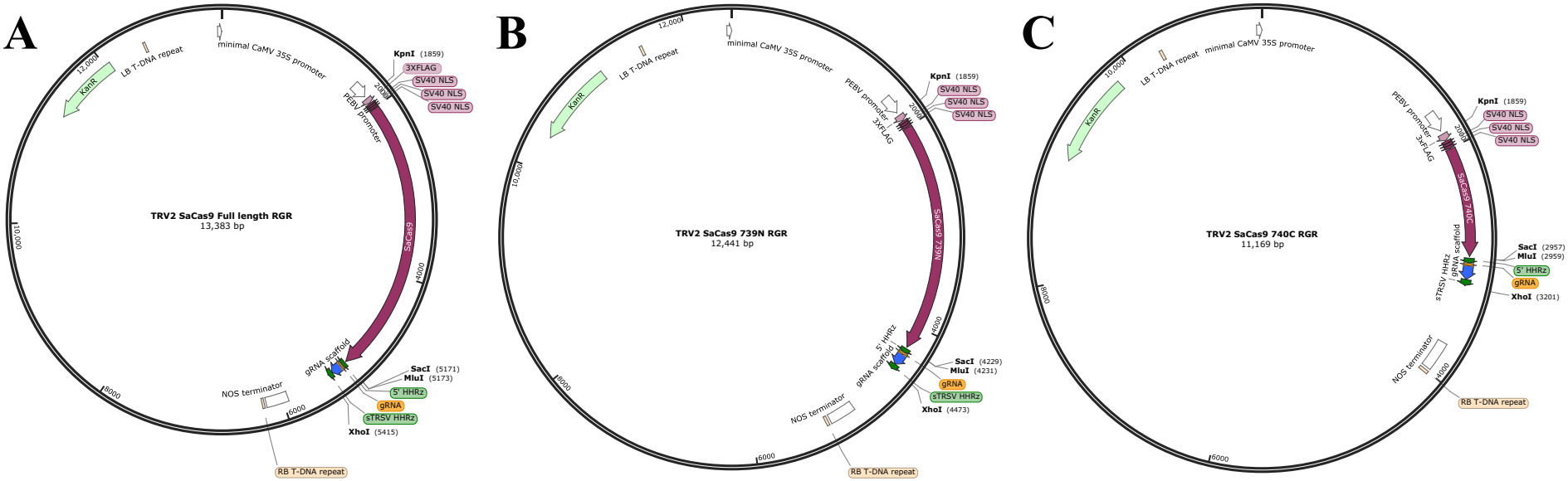
Supplementary figure 1. pCAMBIA1300 vectors expressing A. *Sp*Cas9 full-length (4271 bp). B. N-*Sp*Cas9 N-intein (2145 bp) and C. C-intein C-*Sp*Cas9 (2553 bp). Key restriction sites, CaMV 35S promoter, NosT or CaMV polyA signal and antibiotic resistance (Kan^R) are shown. Vector maps were created using SnapGene Viewer v6.0 software.



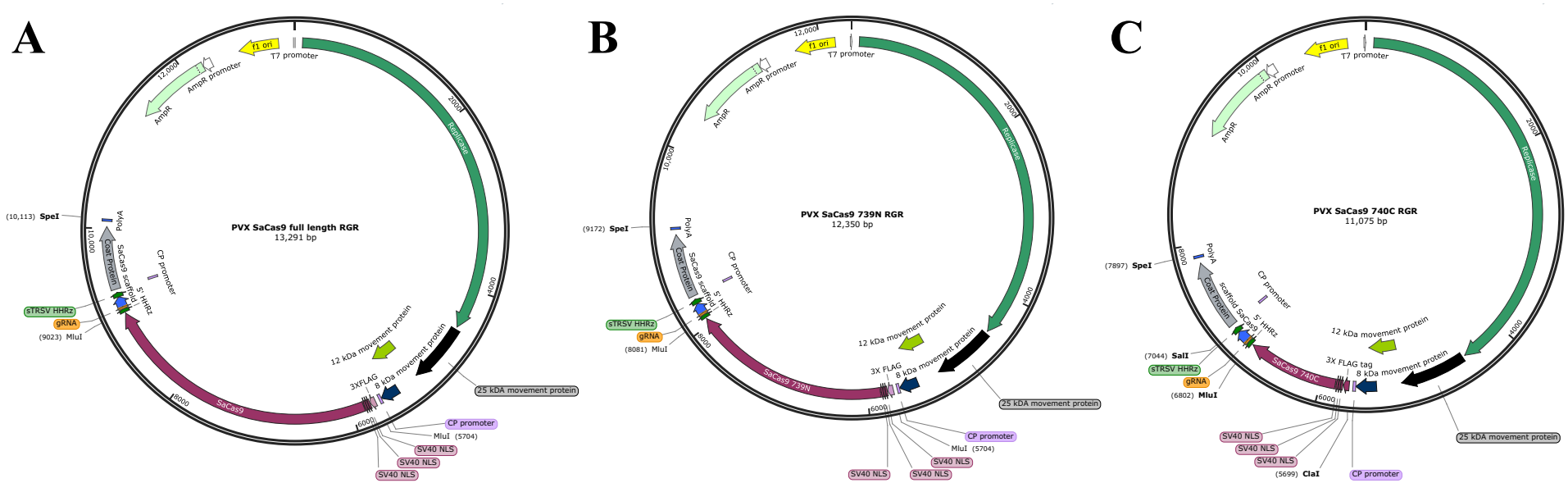
Supplementary figure 2. TRV *Sp*Cas9-RGR vectors. **A. TRV *Sp*Cas9 full-length RGR (4491 bp). B. TRV N-*Sp*Cas9 N-intein RGR (2374 bp). C. TRV C-intein C-*Sp*Cas9 RGR (2777 bp)**. Key restriction sites, PEBV CP promoter for expression of the transgene and antibiotic resistance (Kan^R) are shown. Vector maps were created using SnapGene Viewer v6.0 software.



Supplementary figure 3. PVX *Sp*Cas9-RGR vectors. **A**. PVX *Sp*Cas9 full-length RGR (4491 bp). **B.** PVX N-*Sp*Cas9 N-intein RGR (2374 bp). **C.** PVX C-intein C-*Sp*Cas9 RGR (2777 bp). Key restriction sites, T7 promoter for *in vitro* transcription, CP promoter for expression of the transgene and antibiotic resistance (Amp^R) are shown. Vector maps were created using SnapGene Viewer v6.0 software.



Supplementary figure 4. TRV *Sa*Cas9-RGR vectors. **A. TRV *Sa*Cas9 full-length RGR (3554 bp). B. TRV *Sa*Cas9 739N RGR (2612 bp). C. TRV *Sa*Cas9 740C RGR (1340 bp).** Key restriction sites, PEBV CP promoter for expression of the transgene and antibiotic resistance (Kan^R) are shown. Vector maps were created using SnapGene Viewer v6.0 software.

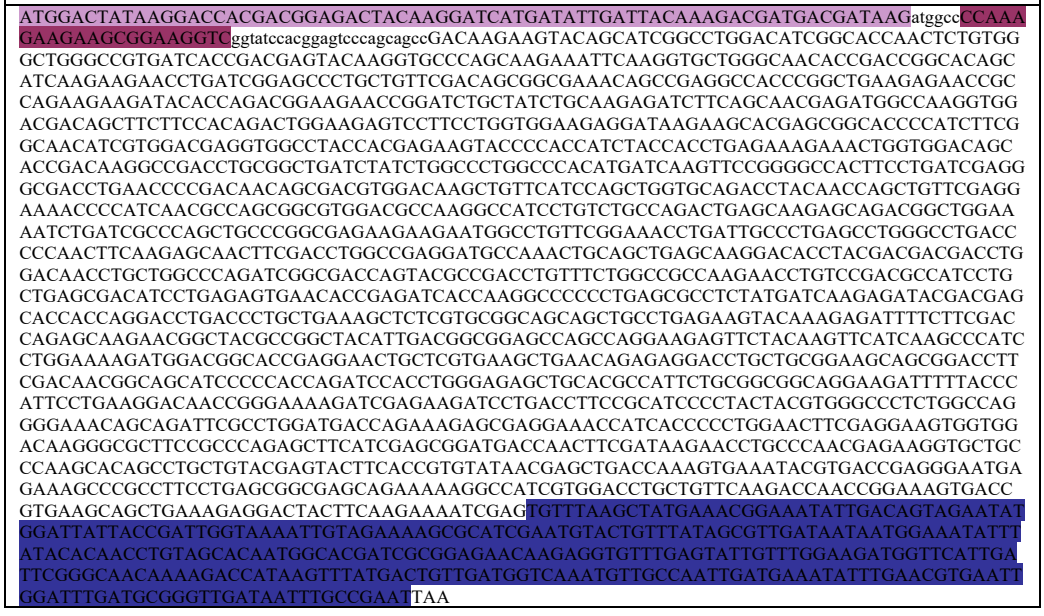


Supplementary figure 5. PVX *Sa*Cas9-RGR vectors. A. PVX *Sa*Cas9 full-length RGR (3554 bp). B. PVX *Sa*Cas9 739N RGR (2612 bp). C. PVX *Sa*Cas9 740C RGR (1340 bp). Key restriction sites, T7 promoter for in vitro transcription, CP promoter for the expression of the transgene and antibiotic resistance (Amp^R) are shown. Vector maps were created using SnapGene Viewer v6.0 software.

A


ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGatggccCCAAA
GAAGAAGCGGAAGGTCggtatccacggagtcccagcagccGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGG
GCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGC
ATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGC
CAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGG
ACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCG
GCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGC
ACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGG
GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGG
AAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAA
AATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACC
CCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTG
GACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTG
CTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAG
CACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGAC
CAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATC
CTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTT
CGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCC
ATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAG
GGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGG
ACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGC
CCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGA
GAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGG
TTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAAC
GAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACC
TATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAA
CAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCA
GGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA
GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC
AGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAG
CCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAG
CTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGC
CCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAG
TTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGT
GGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACA
AGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAA
AGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAA
GTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCT
GGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCC
GGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGC
GGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA
GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAA
GAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACT
TTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGG
AAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT
GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTT
GTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGAC
GCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATC
CACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACA
CCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGT
CTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGTAA

B


ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGatggccCCAAA
GAAGAAGCGGAAGGTCggtatccacggagtcccagcagccGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGG
GCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGC
ATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGC
CAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGG
ACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCG
GCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGC
ACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGG
GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGG
AAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAA
AATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACC
CCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTG
GACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTG
CTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAG
CACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGAC
CAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATC
CTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTT
CGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCC
ATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAG
GGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGG
ACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGC
CCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGA
GAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGTTTAAGCTATGAAACGGAAATATTGACAGTAGAATAT
GGATTATTACCGATTGGTAAAATTGTAGAAAAGCGCATCGAATGTACTGTTTATAGCGTTGATAATAATGGAAATATTT
ATACACAACCTGTAGCACAATGGCACGATCGCGGAGAACAAGAGGTGTTTGAGTATTGTTTGGAAGATGGTTCATTGA
TTCGGGCAACAAAAGACCATAAGTTTATGACTGTTGATGGTCAAATGTTGCCAATTGATGAAATATTTGAACGTGAATT
GGATTTGATGCGGGTTGATAATTTGCCGAATTAA

C

ATGATCAAAATAGCCACACGTAAATATTTAGGCAAACAAAATGTCTATGACATTGGAGTTGAGCGCGACCATAATTTT
GCACTCAAAAATGGCTTCATAGCTTCTAATTGTTTCAATTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGT
TCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCT
ATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGG
AAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAAC
AGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAG
GGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAG
GTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGG
CGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGC
TTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCC
CTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGT
TCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG
GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAA
GCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAA
GTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAG
TACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGA
GCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTG
GCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCG
GGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCG
GCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAG
AAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAG
AAACTGAAGAGTGTGAAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTT
CTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAA
AACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTG
AACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTG
GAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCT
AATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCAC
CTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCA
GCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC
AGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGTAA

D

TTGGGCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGCCCAAGCAACCCTAACCTGGTTTTAGAGCTAGAA
ATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTCCTGTCACCGGATGT
GCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGG

E

ATTCTCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGAGAATCCACAACCATCATTGTTTTAGAGCTAGAA
ATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTCCTGTCACCGGATGT
GCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGG

F

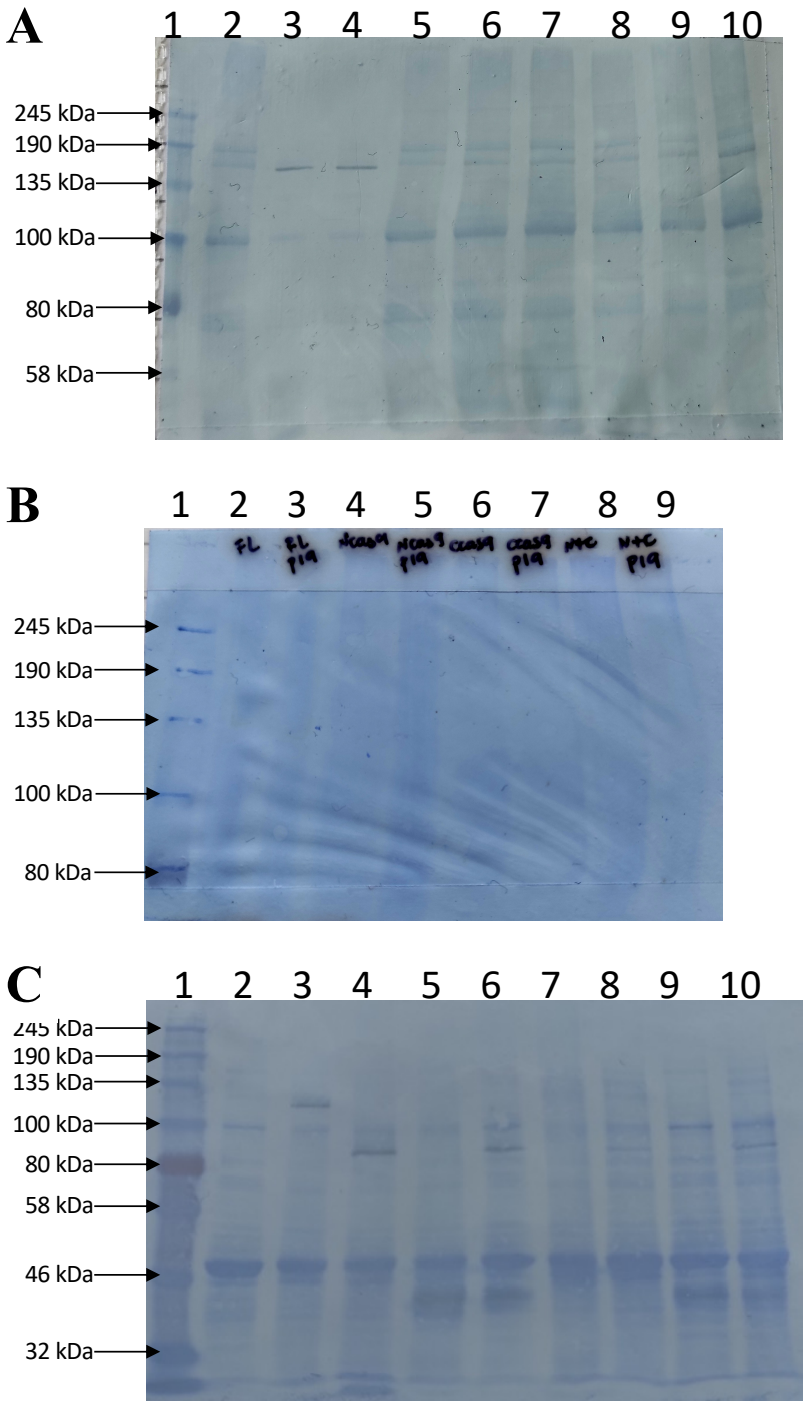
TTGGGCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGCCCAAGCAACCCTAACCTGAGTTTTAGTACTCTGT
AATTTTAGGTATGAGGTAGACGAAAATTGTACTTATACCTAAAATTACAGAATCTACTAAAACAAGGCAAAATGCCGT
GTTTATCTCGTCAACTTGTTGGCGAGACCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGG

G

TTCTCGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCGAGAATCCACAACCATCATTGTTTTAGTACTCTGT
AATTTTAGGTATGAGGTAGACGAAAATTGTACTTATACCTAAAATTACAGAATCTACTAAAACAAGGCAAAATGCCGT
GTTTATCTCGTCAACTTGTTGGCGAGACCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGG

Supplementary figure 6. Sequences of **(A)** *Sp*Cas9 full-length, **(B)** N-*Sp*Cas9 N-intein, **(C)** C-intein C-*Sp*Cas9, **(D)** *Sp*Cas9 TobFT1 HH, **(E)** *Sp*Cas9 TobFT4 HH, **(F)** *Sa*Cas9 TobFT1 HH, **(G)** *Sa*Cas9 TobFT4 HH. 3X FLAG TAG, Nuclear Localization signal (NLS), N-intein, C-intein, HHRz, gRNA and RNA scaffold are highlighted in different colors. Sequences for *Sa*Cas9 full-lenght, *Sa*Cas9 739N and *Sa*Cas9 740C were retrieved from Kaya *et al.* (2017) (143)

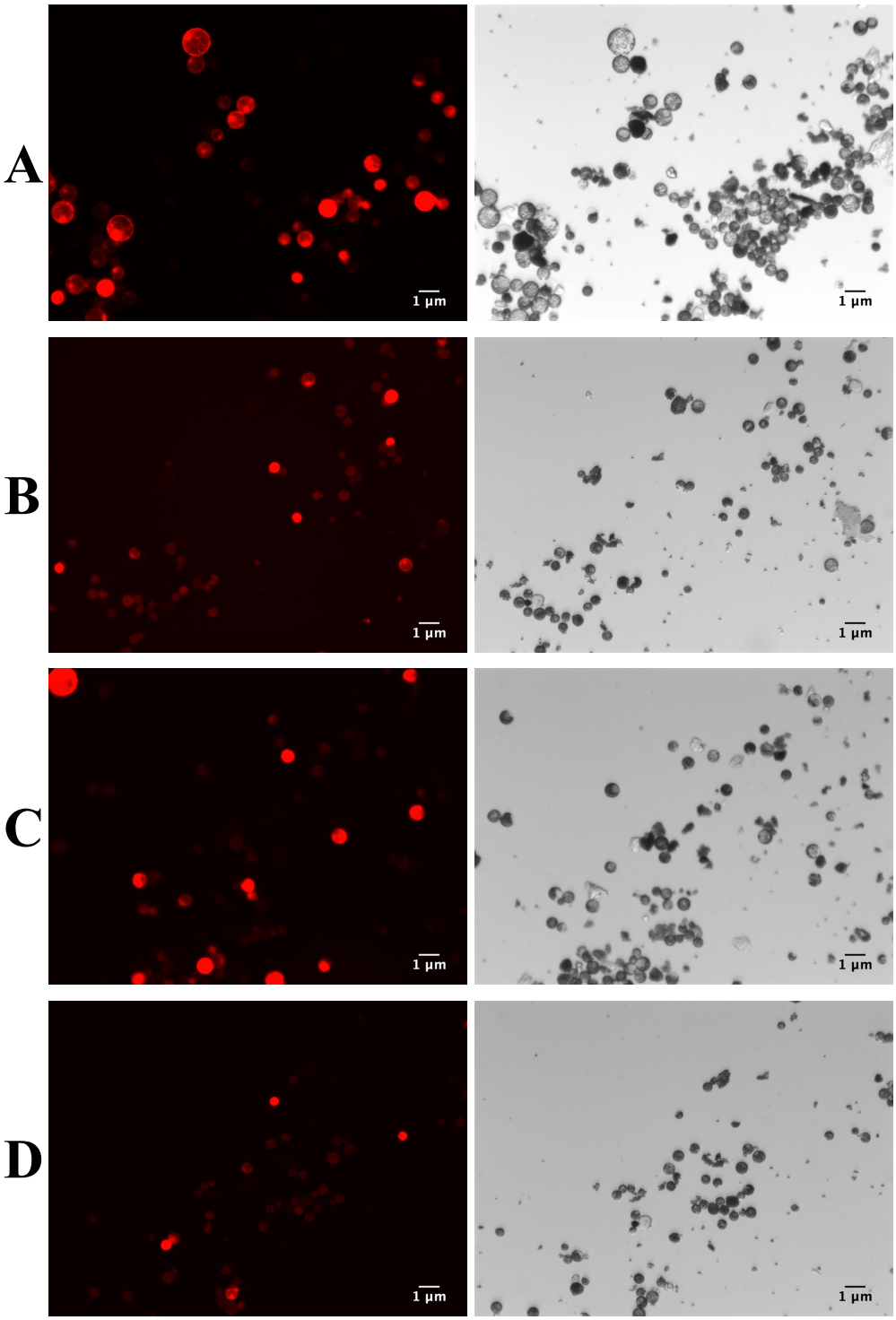
Appendix A2. Immunoblot stained membranes (Cas9 delivered by plant expression vectors)



Supplementary figure 7. *Agrobacterium* plant vectors mediated expression of *Sp*Cas9 and *Sa*Cas9 constructs in *N. tabacum* leaves. Western blot membranes stained with Coomasie blue to confirm total protein transfer. A. Immunoblot using anti-FLAG® M2 antibody 1:1000 to detect *Sp*Cas9 constructs. Lane 1. Protein Standard Ladder (NEB #P7712). Lane 2. Mock. Lane 3. Full-length *Sp*Cas9. Lane 4. Full-length *Sp*Cas9 + p19. Lane 5. C-intein C-*Sp*Cas9. Lane 6. C-intein C-*Sp*Cas9 + p19. Lane 7. N-*Sp*Cas9 N-intein. Lane 8. N-*Sp*Cas9 N-intein + p19. Lane 9. N-*Sp*Cas9 N-intein + C-intein C-*Sp*Cas9. Lane 10. N-*Sp*Cas9 N-intein + C-intein C-*Sp*Cas9 + p19 (*Legend continues in next page*).

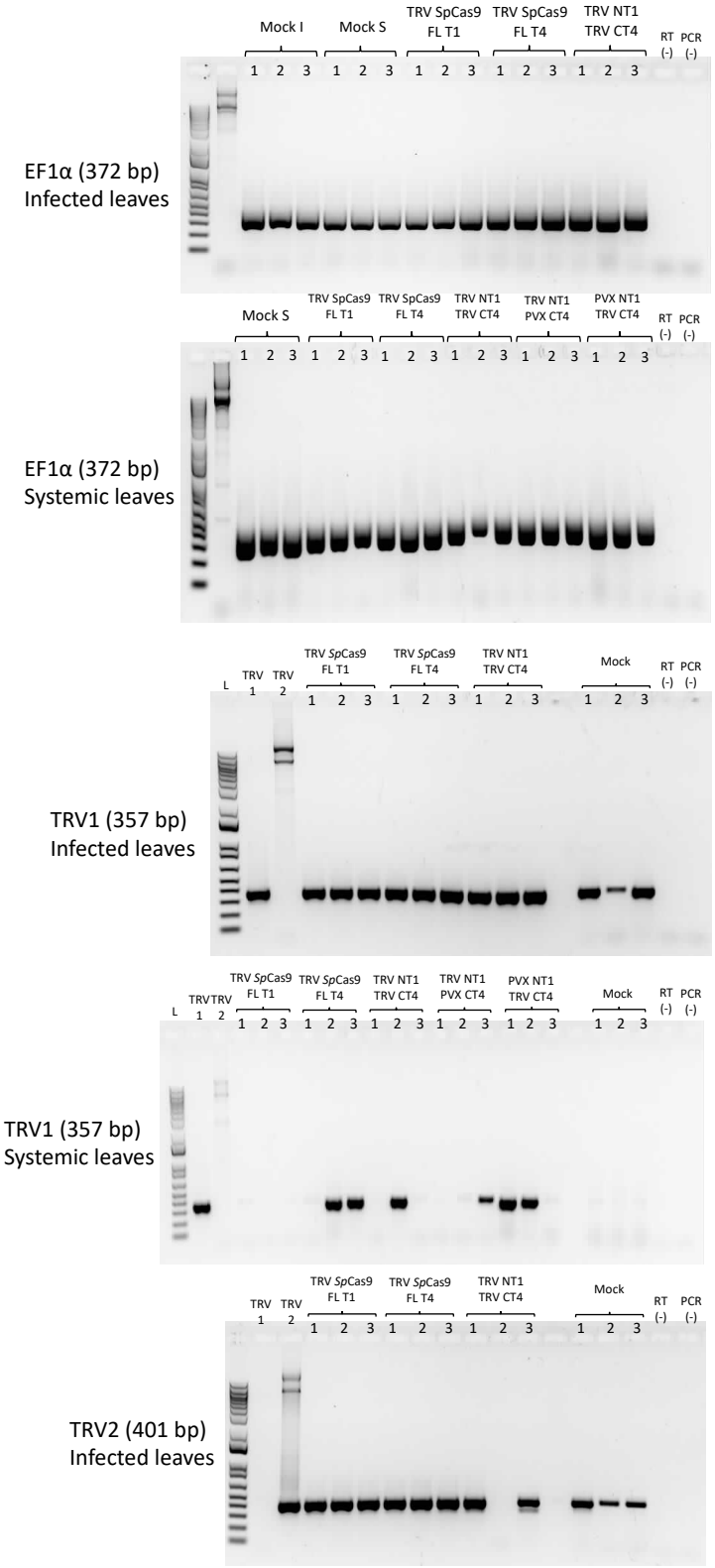
B. Immunoblot using anti CRISPR/Cas9 antibody 1:500 to detect *Sp*Cas9 constructs. Lane 1. Protein Standard Ladder (NEB #P7712). Lane 2. Full-length *Sp*Cas9. Lane 3. Full-length *Sp*Cas9 + p19. Lane 4. N-*Sp*Cas9 N-intein. Lane 5. N-*Sp*Cas9 N-intein + p19. Lane 6. C-intein C-*Sp*Cas9. Lane 7. C-intein C-*Sp*Cas9 + p19. Lane 8. N-*Sp*Cas9 N-intein + C-intein C-*Sp*Cas9. Lane 9. N-*Sp*Cas9 N-intein + C-intein C-*Sp*Cas9 + p19. Black arrows indicate fully re-assembled *Sp*Cas9. **C. Immunoblot using anti-FLAG® M2 antibody 1:1000 to detect *Sa*Cas9 constructs.** Lane 1. Protein Standard Ladder (NEB #P7712). Lane 2. Mock. Lane 3 – 6. Expression with p19; Lane 3. Full-length *Sa*Cas9. Lane 4. *Sa*Cas9 739N. Lane 5. *Sa*Cas9 740C. Lane 6. *Sa*Cas9 739N + *Sa*Cas9 740C. Lane 7 – 10. Expression without p19; Lane 7. Full-length *Sa*Cas9. Lane 8. *Sa*Cas9 739N. Lane 9. *Sa*Cas9 740C. Lane 10. *Sa*Cas9 739N + *Sa*Cas9 740C.

Appendix A3. Protoplasts transformation

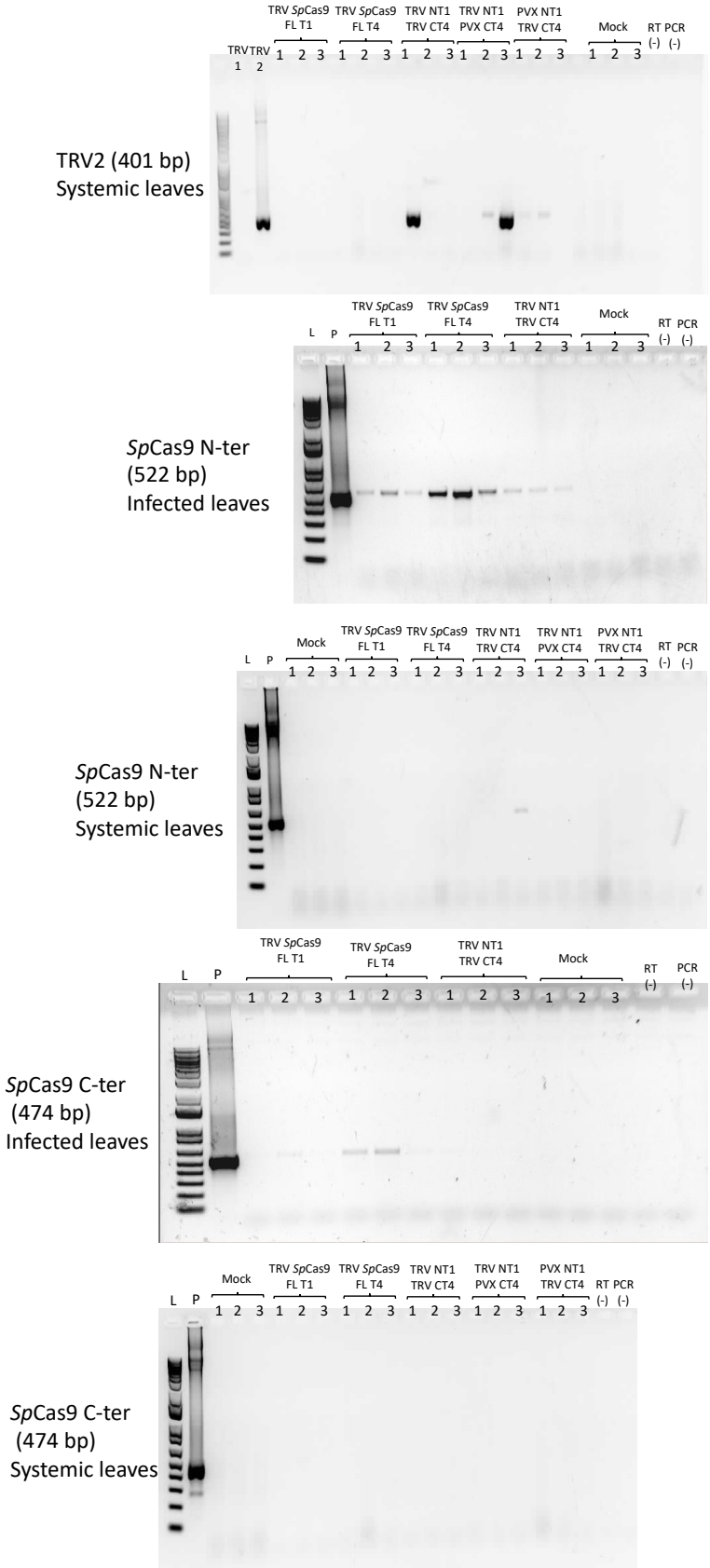


Supplementary figure 8. ***N. tabacum*** **protoplast transformed with Cas9 constructs**. Isolated cells were transformed with plant vectors expressing **A.** mCherry. **B.** mCherry + *Sp*Cas9 Full-length + PDS sgRNA. **C.** mCherry + N-Cas9 N-intein + C-intein C-Cas9 + PDS sgRNA **D.** *Sa*Cas9 Full-length+ PDS sgRNA. Red fluorescence was checked after 24 hours post transformation. For mCherry + *Sp*Cas9 full-length + TobFT1 HH sgRNA, mCherry + *Sp*Cas9 full-length + TobFT4 HH sgRNA, mCherry + *Sa*Cas9 full-length + TobFT1 HH sgRNA and mCherry + *Sa*Cas9 full-length + TobFT4 HH sgRNA fluorescence was confirmed, but pictures of these protoplasts were not taken.

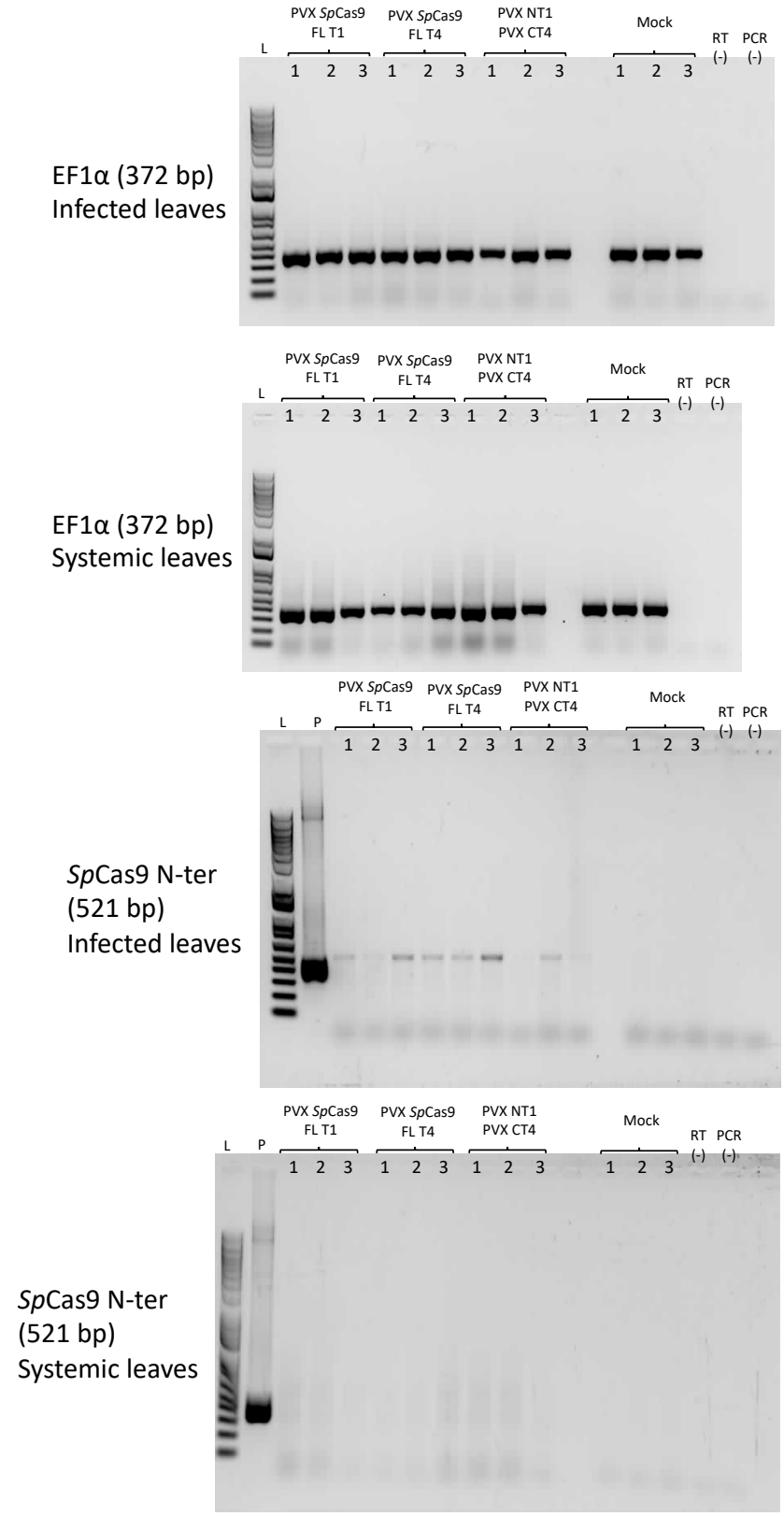
Appendix A4. RT-PCRs agarose gels



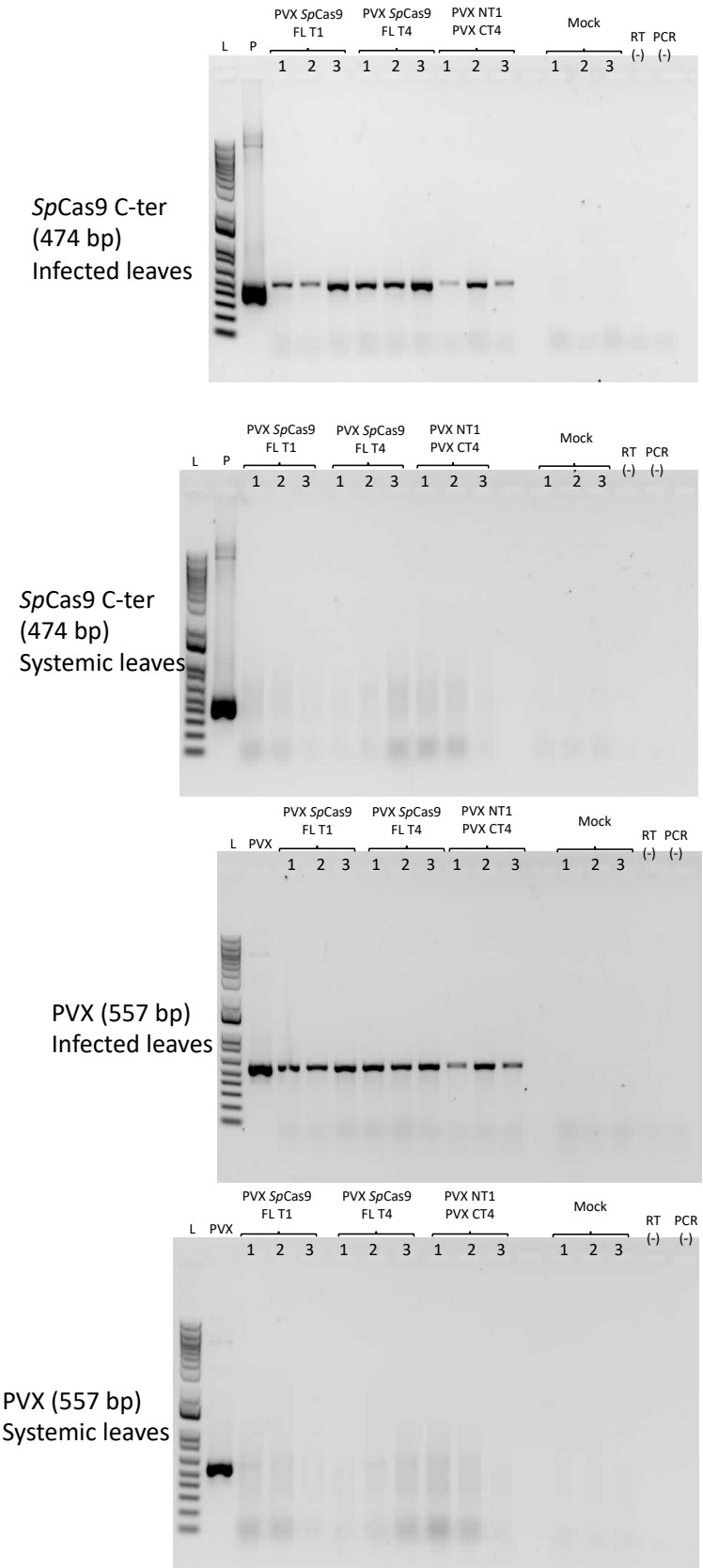
Supplementary figure 9. RT-PCR to detect expression of *SpCas9* mRNA in TRV infected leaves. (L) 1 kb plus DNA ladder. (P) pBlueScript II SK(+) *Sp*Cas9 full-length used as positive control of PCR. TRV1 and TRV2 are PCR positive controls of both vectors. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. Amplicon size and loading order are depicted on each figure (*Continues next page*).



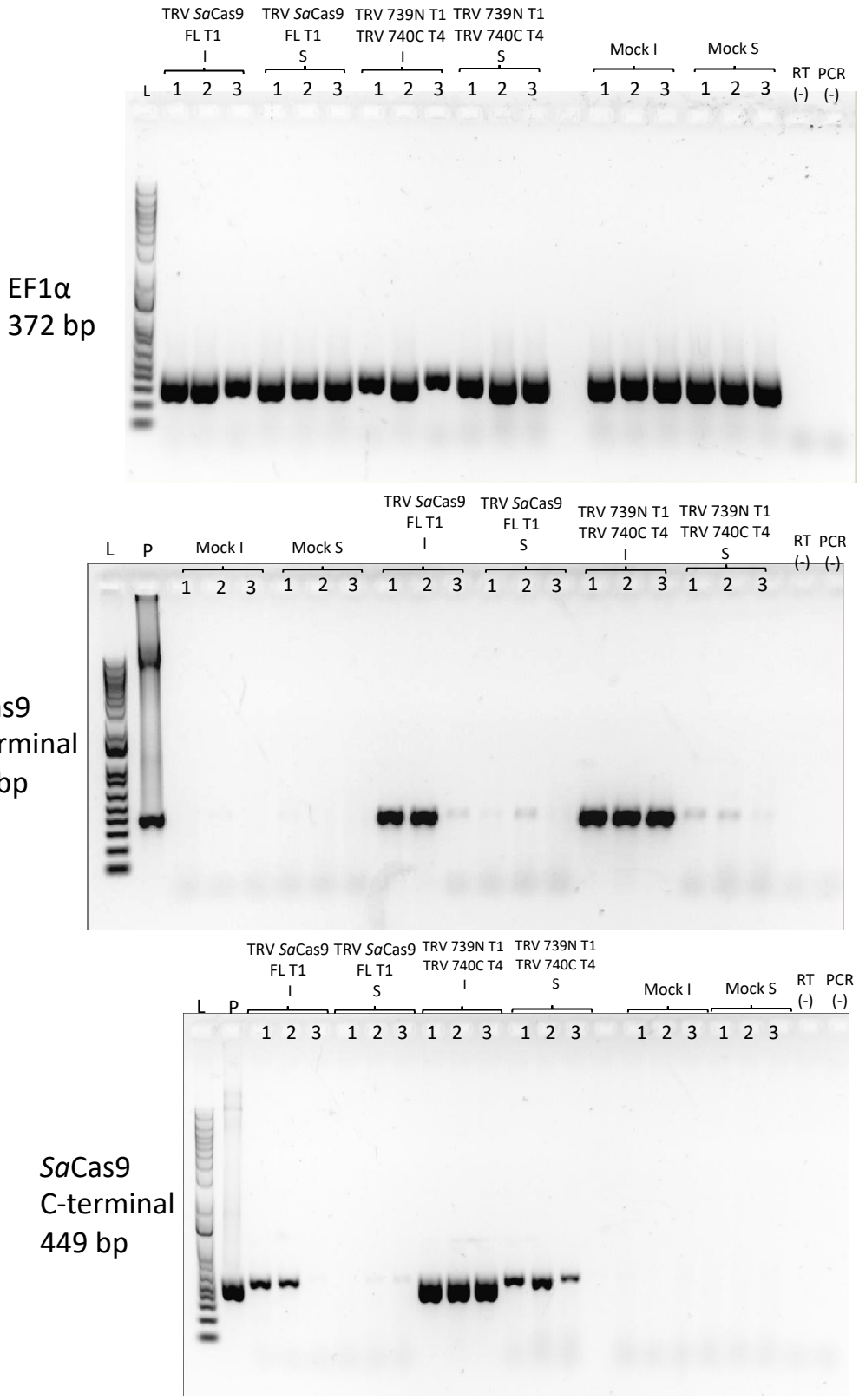
Supplementary figure 9 (*Continuation*). RT-PCR to detect expression of *SpCas9* mRNA in TRV infected leaves. (L) 1 kb plus DNA ladder. (P) pBlueScript II SK(+) *Sp*Cas9 full-length used as positive control of PCR. TRV1 and TRV2 are positive controls of amplification of both vectors. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. Amplicon size and loading order are depicted on each figure.



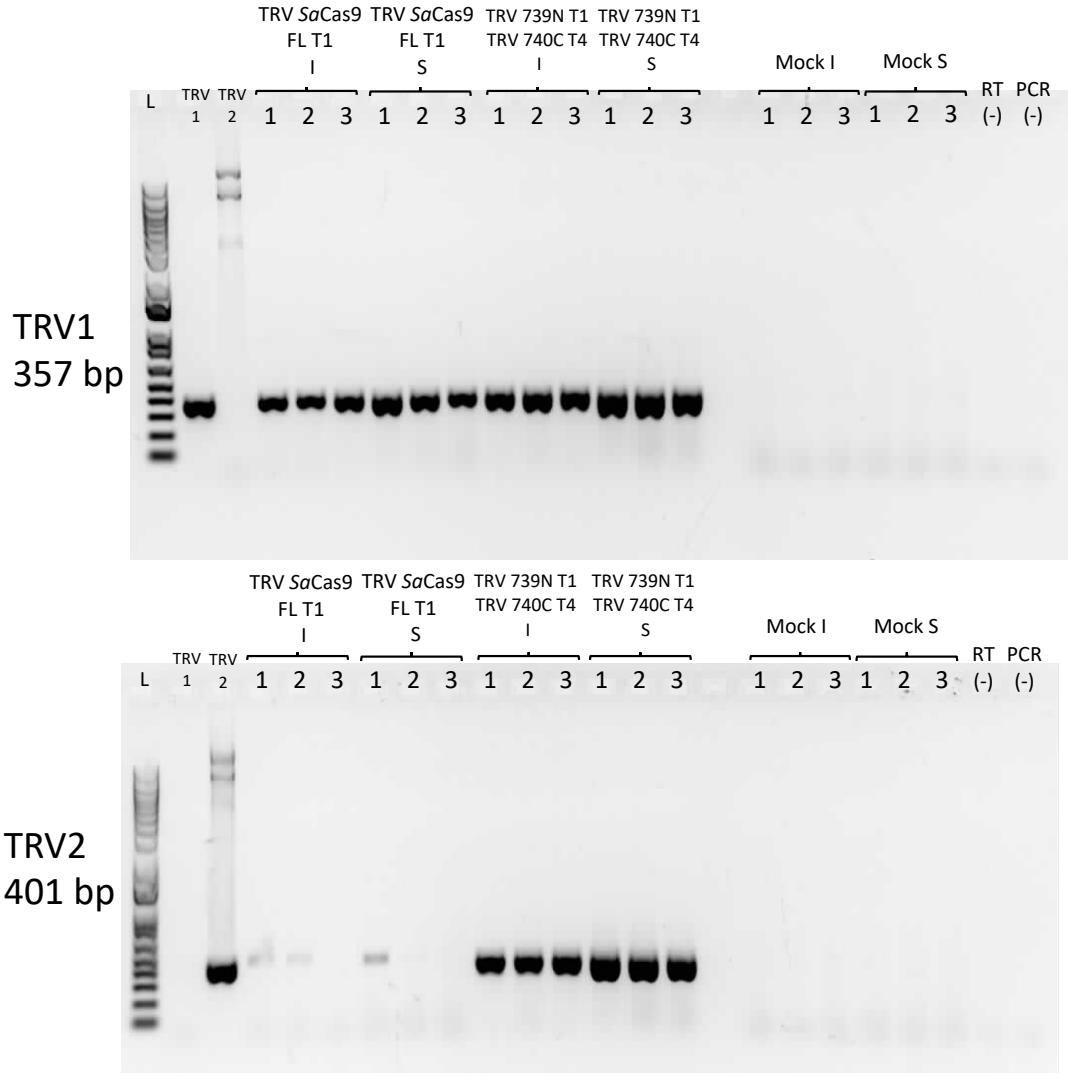
Supplementary figure 10. RT-PCR to detect expression of *SpCas9* mRNA in PVX infected leaves. (L) 1 kb plus DNA ladder. (P) pBlueScript II SK(+) *Sp*Cas9 full-length used as positive control of PCR. PVX is a positive control of amplification of PVX vector. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. Amplicon size and loading order are depicted on each figure (*continues in next page*).



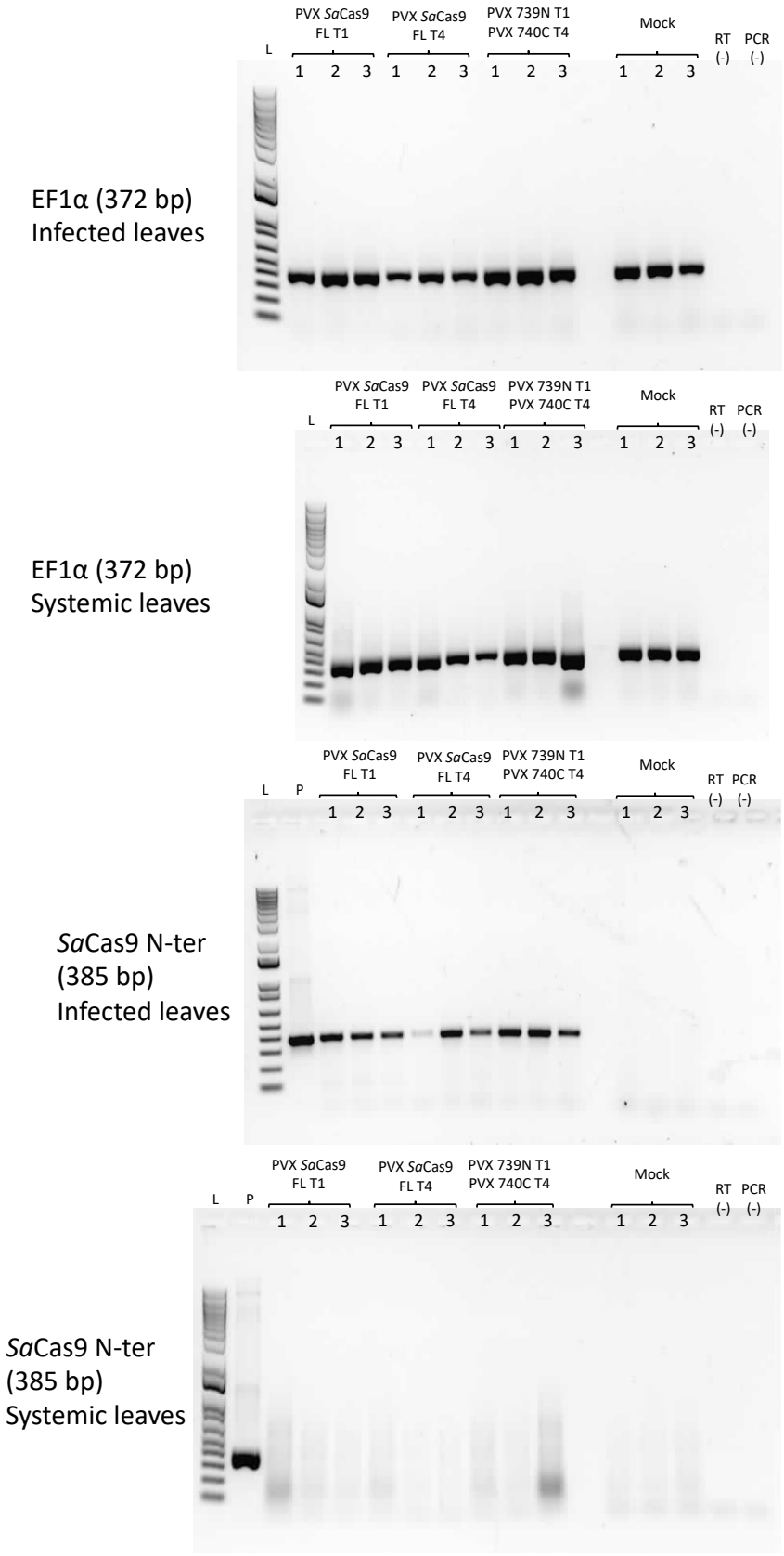
Supplementary figure 10 (*Continuation*). RT-PCR to detect expression of *SpCas9* mRNA in PVX infected leaves. (L) 1 kb plus DNA ladder. (P) pBlueScript II SK(+) *Sp*Cas9 full-length used as positive control of PCR. PVX is a positive control of amplification of PVX vector. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. Amplicon size and loading order are depicted on each figure.



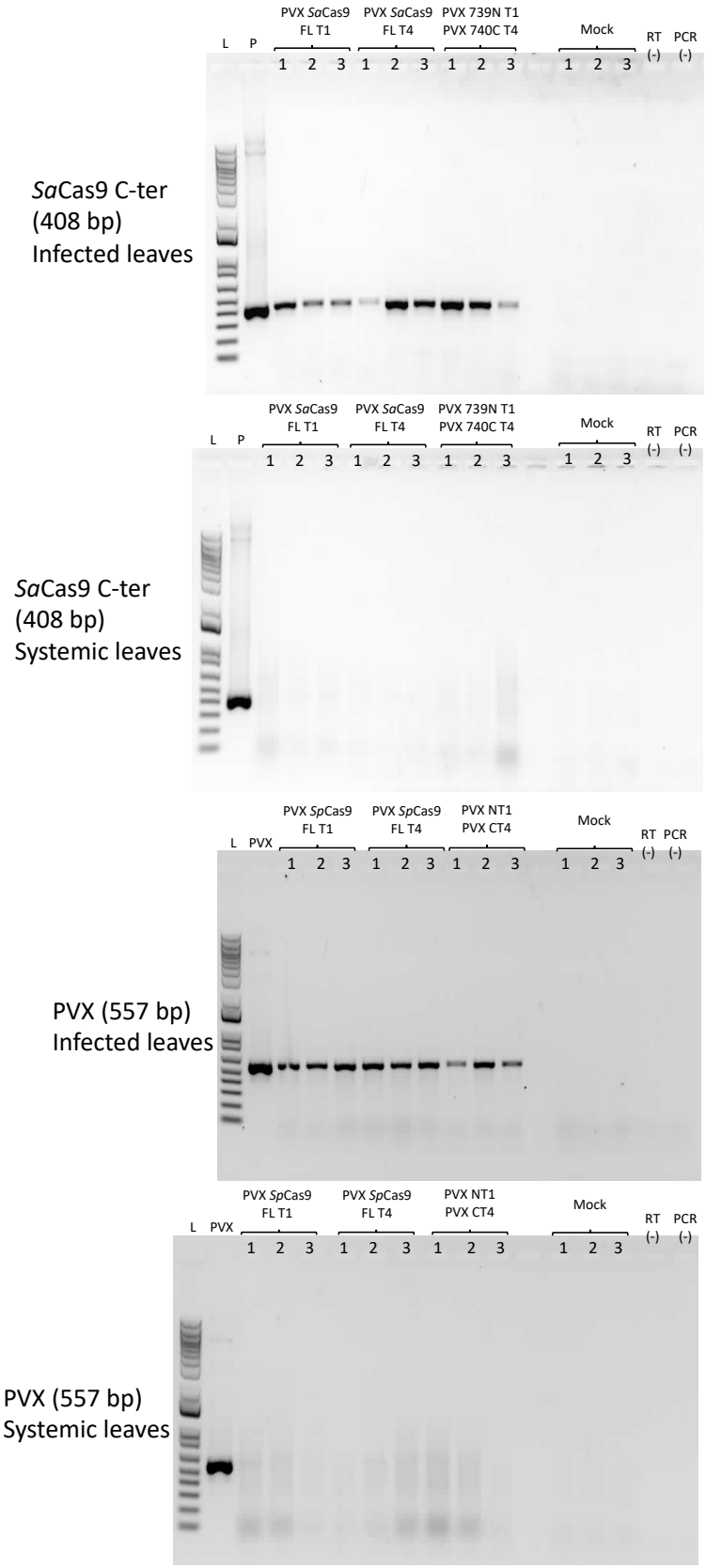
Supplementary figure 11. RT-PCR to detect expression of *SaCas9* mRNA in TRV infected leaves. (L) 1 kb plus DNA ladder. (P) pRI *Sa*Cas9 full-length used as positive control of PCR. TRV1 and TRV2 are positive controls of amplification of both vectors. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. (I) Infected leaves. (S) Systemic leaves. Amplicon size and loading order are depicted on each figure (*Continues in next page*).



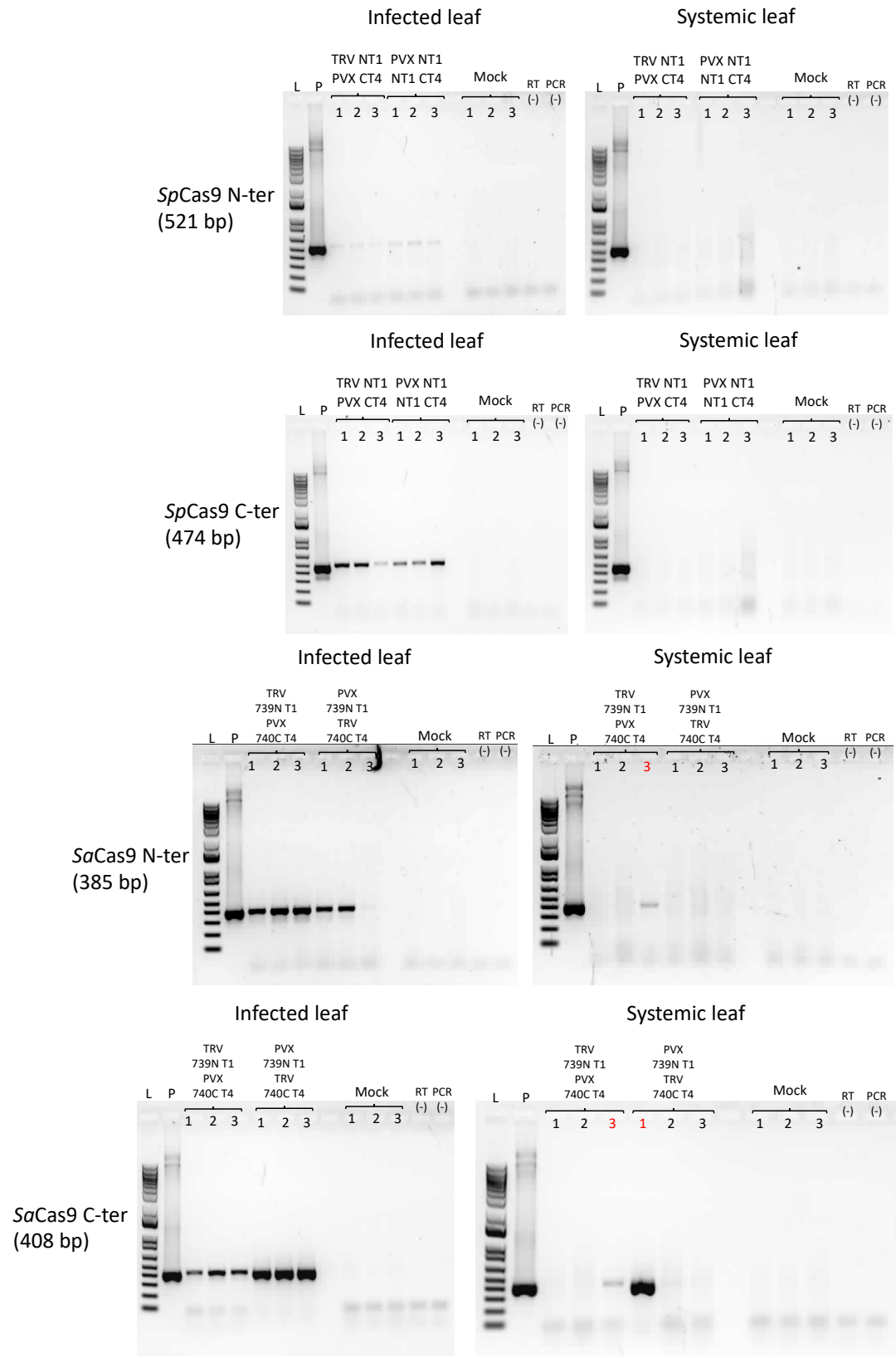
Supplementary figure 11. RT-PCR to detect expression of *SaCas9* mRNA in TRV infected leaves (*Continuation*). (L) 1 kb plus DNA ladder. (P) pRI *Sa*Cas9 full-length used as positive control of PCR. TRV1 and TRV2 are positive controls of amplification of both vectors. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. (I) Infected leaves. (S) Systemic leaves. Amplicon size and loading order are depicted on each figure.



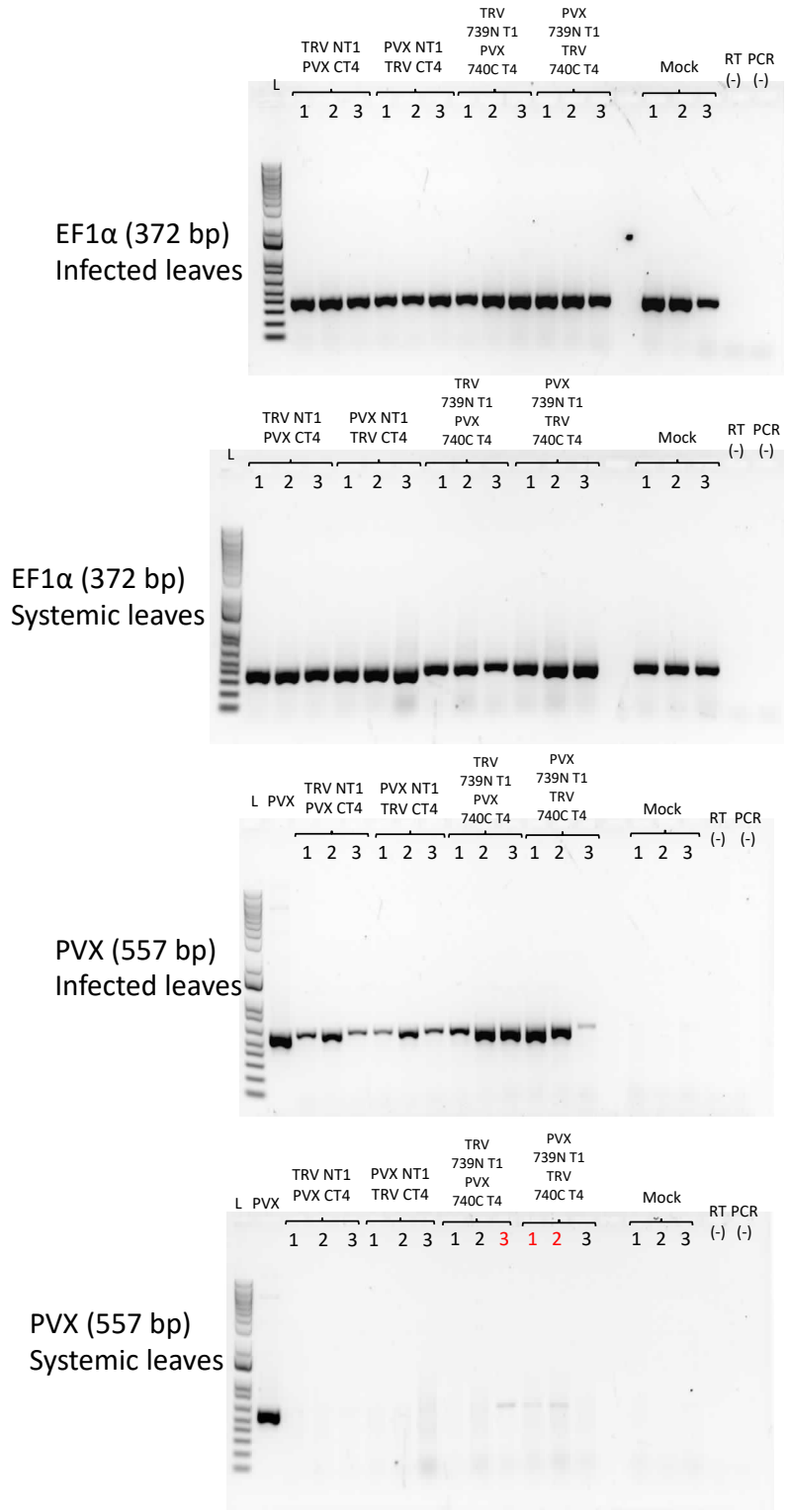
Supplementary figure 12. RT-PCR to detect expression of *SaCas9* mRNA in PVX infected leaves. (L) 1 kb plus DNA ladder. (P) pRI *Sa*Cas9 full-length used as positive control of PCR. PVX is a positive control of amplification of PVX vector. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. Amplicon size and loading order are depicted on each figure (*Continues in next page*).



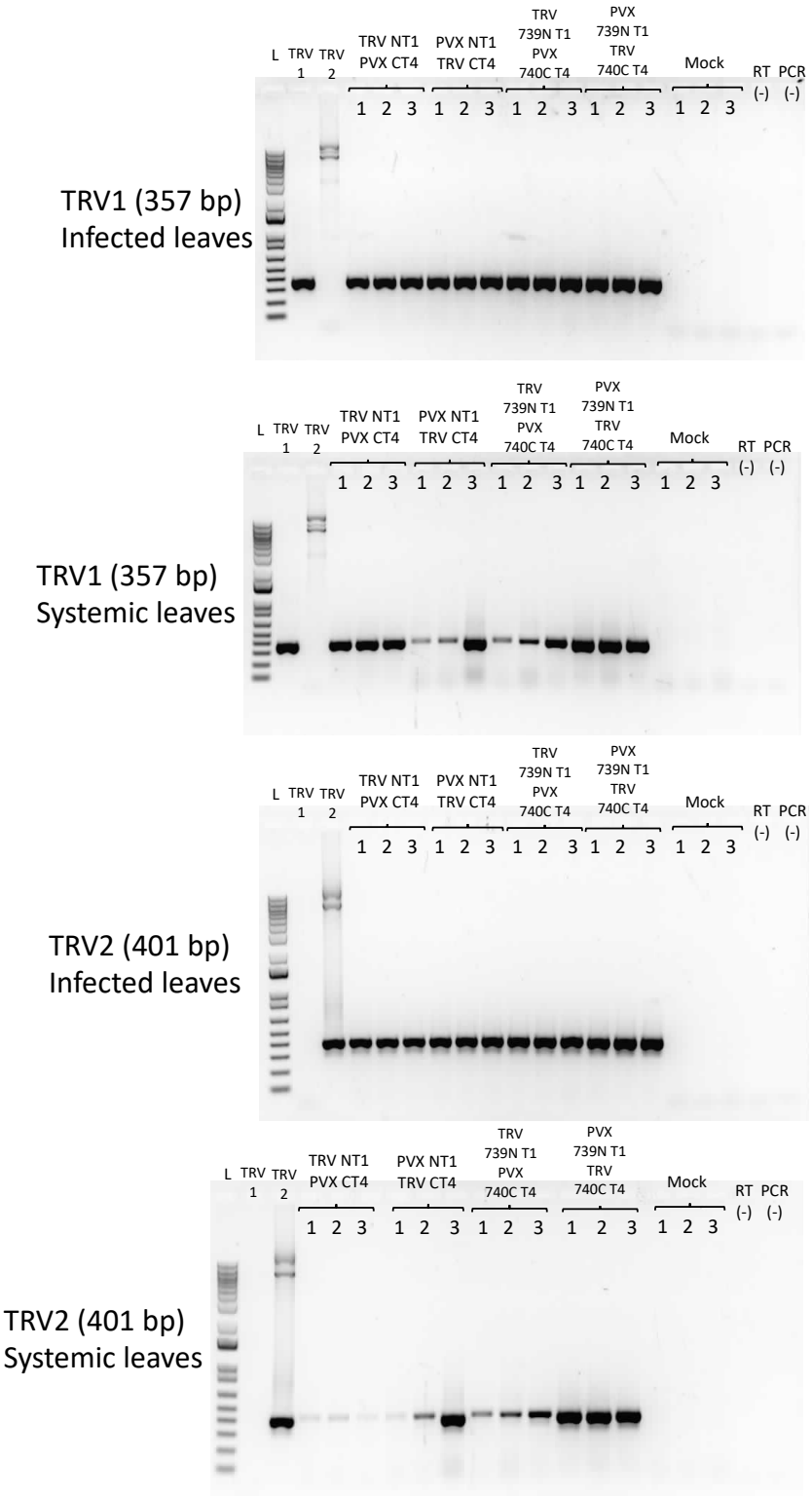
Supplementary figure 12 (*Continuation*). RT-PCR to detect expression of *SaCas9* mRNA in PVX infected leaves. (L) 1 kb plus DNA ladder. (P) pRI *Sa*Cas9 full-length used as positive control of PCR. PVX is a positive control of amplification of PVX vector. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. Amplicon size and loading order are depicted on each figure.



Supplementary figure 13. RT-PCR to detect expression of split *Sp*Cas9 and *SaCas9* mRNA in co-infected leaves with TRV and PVX. (L) 1 kb plus DNA ladder. (P) pBlueScript II SK(+) *Sp*Cas9 full-length or pRI *Sa*Cas9 full-length used as positive control of PCR. TRV1, TRV2 and PVX are positive controls of amplification of each viral vector. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. Amplicon size and loading order are depicted on each figure (*Continues in next page*).

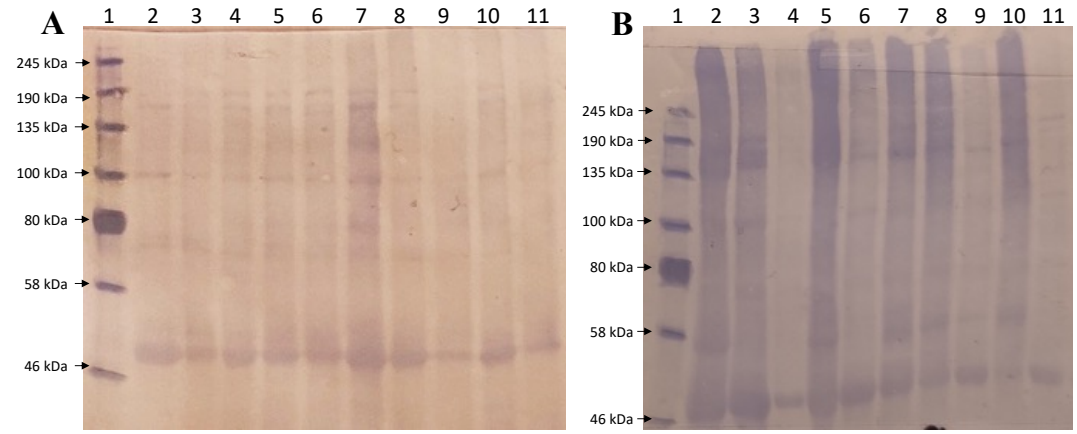


Supplementary figure 13 (*Continuation*). RT-PCR to detect expression of split *Sp*Cas9 and *SaCas9* mRNA in co-infected leaves with TRV and PVX. (L) 1 kb plus DNA ladder. (P) pBlueScript II SK(+) *Sp*Cas9 full-length or pRI *Sa*Cas9 full-length used as positive control of PCR. TRV1, TRV2 and PVX are positive controls of amplification of each viral vector. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. Amplicon size and loading order are depicted on each figure (*Continues in next page*).

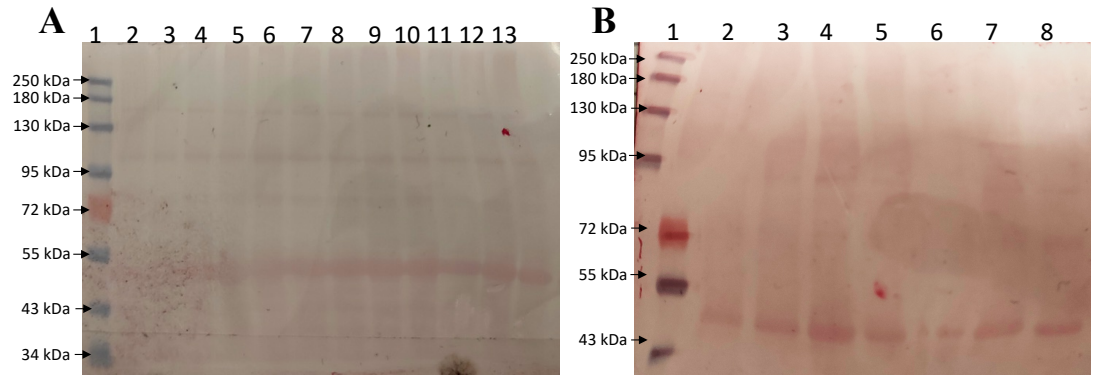


Supplementary figure 13 (*Continuation*). RT-PCR to detect expression of split *Sp*Cas9 and *SaCas9* mRNA in co-infected leaves with TRV and PVX. (L) 1 kb plus DNA ladder. (P) pBlueScript II SK(+) *Sp*Cas9 full-length or pRI *Sa*Cas9 full-length used as positive control of PCR. TRV1, TRV2 and PVX are positive controls of amplification of each viral vector. RT (-) negative control of RT reaction. PCR (-) negative control of PCR amplification. Amplicon size and loading order are depicted on each figure.

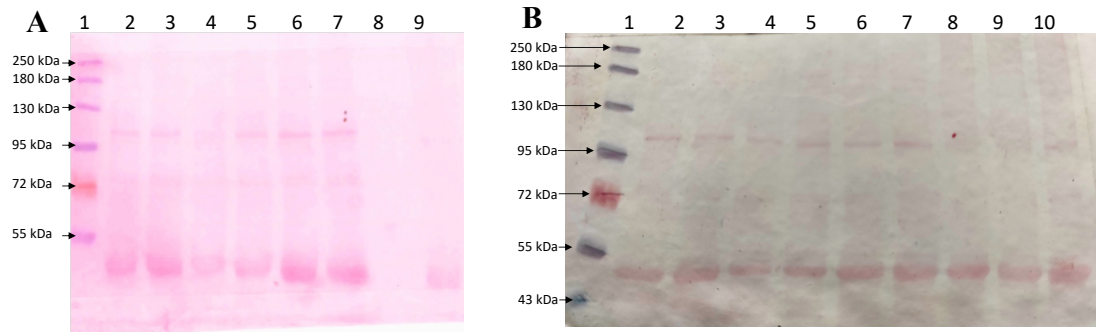
Appendix A5. Immunoblot stained membranes (Cas9 delivered by virus vector)



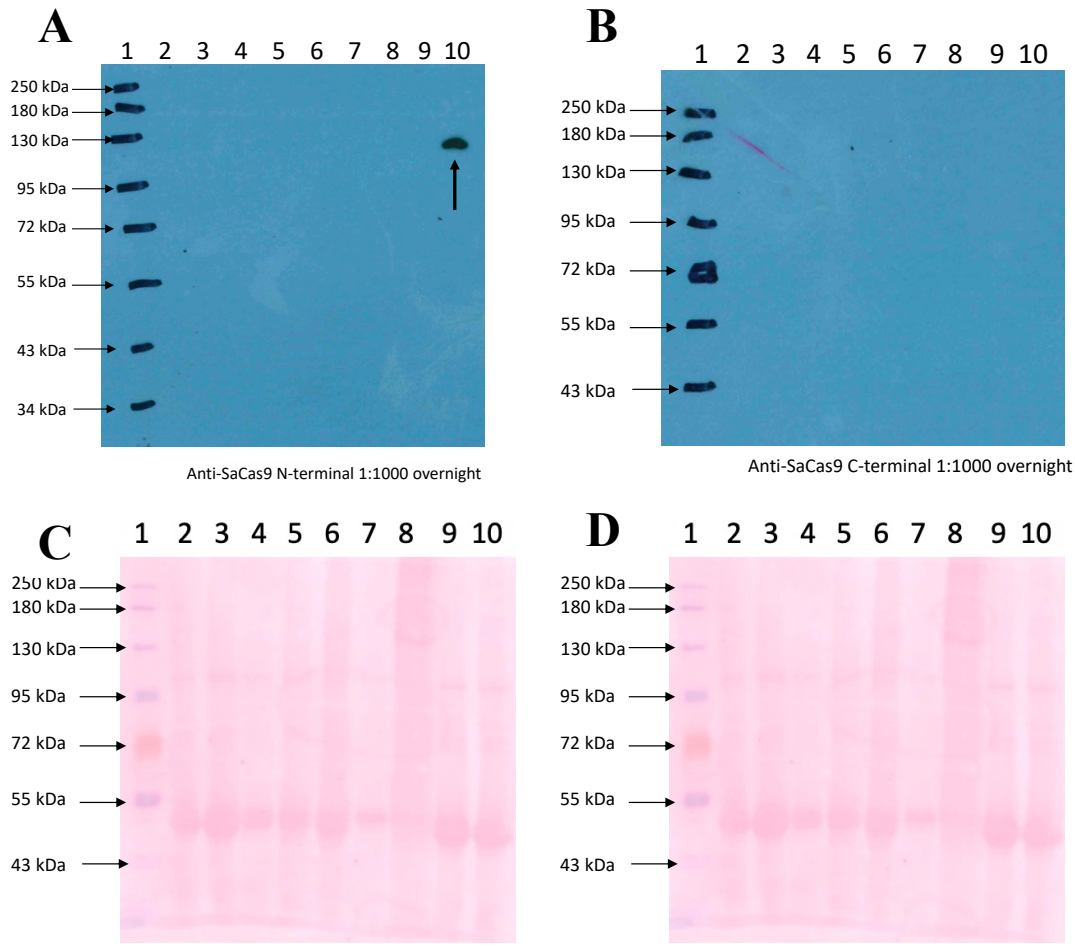
Supplementary figure 14. TRV-mediated expression of *Sp*Cas9 constructs in *N. tabacum* leaves. Western blot membranes stained with Coomasie blue to confirm total protein transfer. Total leaf protein extracts from three biological replicates per construct were analysed using a monoclonal antibody against *Sp*Cas9 (1:1000). The order of both immunoblots is as follow: Lane 1. Protein Standard Ladder (NEB #P7712). Lane 2. Mock. Lanes 3 – 5. TRV2 *Sp*Cas9 full-length TobFT1 HH. Lanes 6 – 8. TRV2 *Sp*Cas9 full-length TobFT4 HH. Lanes 9 – 11.TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4. **A. Infected leaves**. A band of ~160 kDa was identified in samples where full-length *Sp*Cas9 was delivered by TRV. **B. Systemic leaves**. No *Sp*Cas9 protein was detected in these samples.



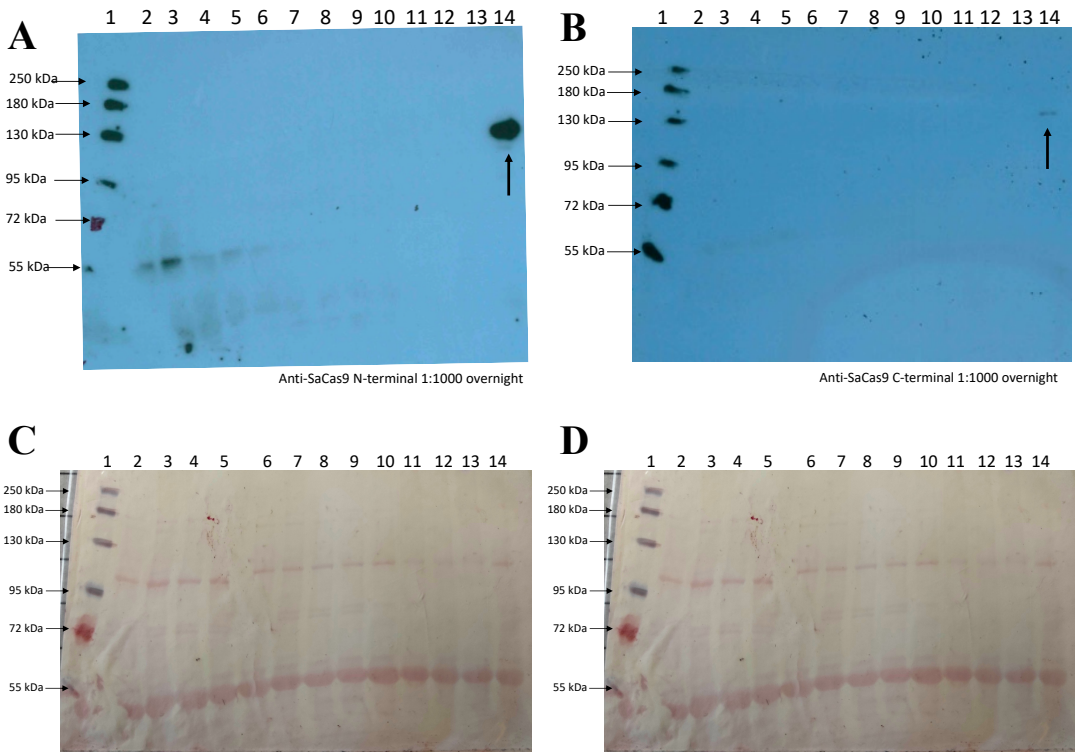
Supplementary figure 15. PVX-mediated expression of *Sp*Cas9 constructs in *N. tabacum* leaves. Western blot membranes stained with Ponceau red to confirm total protein transfer. Total leaf protein extracts from three biological replicates per construct were analysed using a monoclonal antibody against *Sp*Cas9 (1:1000), but no protein was detected. **A. Full-length or split *Sp*Cas9-RGR delivered by PVX**. Lane 1. Protein Standard Ladder (NEB #P7719). Lanes 2 – 4. PVX *Sp*Cas9 full-length TobFT1 HH. Lanes 5 – 7. PVX *Sp*Cas9 full-length TobFT4 HH. Lanes 8 – 10. PVX N-*Sp*Cas9 N-intein TobFT1 HH + PVX C-intein C-*Sp*Cas9 TobFT4. Lane 11 – 13. Mock controls. **B. Co-delivery of split *Sp*Cas9 constructs by PVX and TRV viral vectors.** Lane 1. Protein Standard Ladder (NEB #P7719). Lanes 2 – 4. TRV N-*Sp*Cas9 N-intein TobFT1 HH and PVX C-intein-C-*Sp*Cas9 TobFT4 HH replicates. Lanes 5 – 7. PVX N-*Sp*Cas9 N-intein TobFT1 HH and TRV C-intein-C-*Sp*Cas9 TobFT4 HH replicates. Lanes 8. Mock control.



Supplementary figure 16. Expression of split *Sa*Cas9 protein in *N. tabacum* leaves co-infected with PVX and TRV virus vectors. Western blot membranes stained with Ponceau red to confirm total protein transfer. Three biological replicates per constructs were tested. The order of the lanes is as following: Lane 1. Protein Standard Ladder (NEB #P7719). Lanes 2 - 4. PVX *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH. Lanes 5 - 7. TRV *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH. For panel A. Lane 8. Empty. Lane 9. Mock control, while panel B Lanes 8 – 10 correspond to mock inoculated control plants. **A. Immunoblot using anti-*Sa*Cas9 N-terminal antibody (1:1000**). **B. Immunoblot using anti-*Sa*Cas9 C-terminal antibody (1:1000**). A band at ~43 kDa, corresponding to *Sa*Cas9 C-terminus is identified in samples where this construct was delivered by TRV (Lanes 2 – 4) or PVX (Lanes 5 – 7).

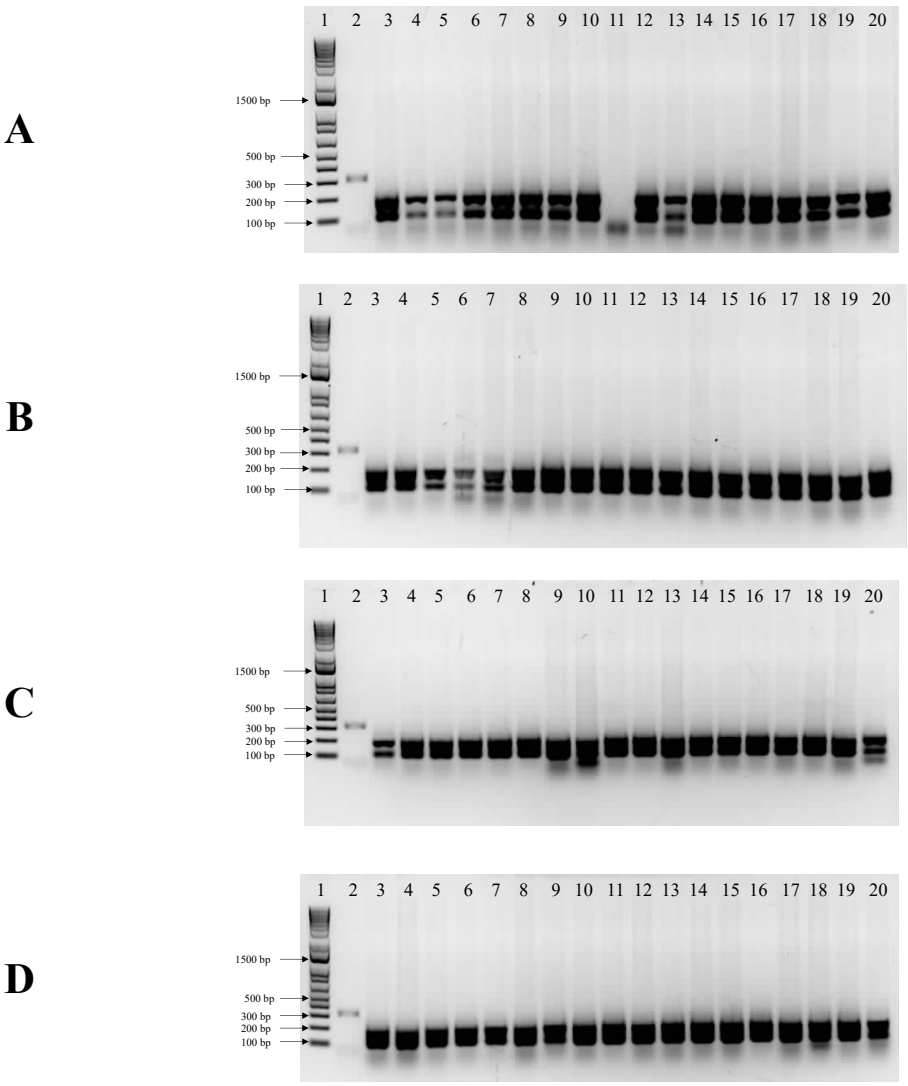


Supplementary figure 17. TRV-mediated expression of *Sa*Cas9 protein in *N. tabacum* infected leaves. Three biological replicates per constructs were tested. The lane order in all figures is as followed: Lane 1. Protein Standard Ladder (NEB #P7719). Lanes 2 - 4. TRV2 *Sa*Cas9 full-length TobFT1 HH. Lanes 5 - 7. TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH. Lane 8. TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH systemic from replicate 1. Lane 9. Mock. Lane 10. pRI201N *Sa*Cas9 full-length expression vector. **A. Immunoblot using anti-*Sa*Cas9 N-terminal dilution 1:1000, exposure time overnight**. A band at ~130 kDa is detected in the positive control (Lane 10), corresponding to *Sa*Cas9 full-length (129.7 kDa). **B. Immunoblot using anti-*Sa*Cas9 C-terminal dilution 1:1000, exposure time overnight**. **C and D. Western blot membranes stained with Ponceau red to confirm total protein transfer**.



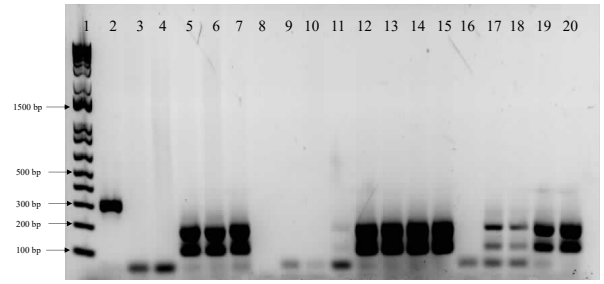
Supplementary figure 18. PVX-mediated expression of *Sa*Cas9 protein in *N. tabacum* infected leaves. Three biological replicates per constructs were tested. In both immunoblots a band at ~130 kDa is detected in the positive control (Lane 14), corresponding to *Sa*Cas9 full-length (129.7 kDa). The order of the samples in all figures is as followed. Lane 1. Protein Standard Ladder (NEB #P7719). Lanes 2 - 4. PVX *Sa*Cas9 full-length TobFT1 HH. Lanes 5 - 7. PVX *Sa*Cas9 full-length TobFT4 HH. Lanes 8 – 10. PVX *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH. Lanes 11 – 13. Mock controls. Lane 14. pRI201N *Sa*Cas9 full-length expression vector. **A. Immunoblot using anti-*Sa*Cas9 N-terminal dilution 1:1000, exposure time overnight**. A faint unspecific band at ~55 kDa is detected in some samples. **B. Immunoblot using anti-*Sa*Cas9 C-terminal dilution 1:1000, exposure time overnight. C and D. Western blot membranes stained with Ponceau red to confirm total protein transfer**.

Appendix A6. CAPS assay for gene editing screening

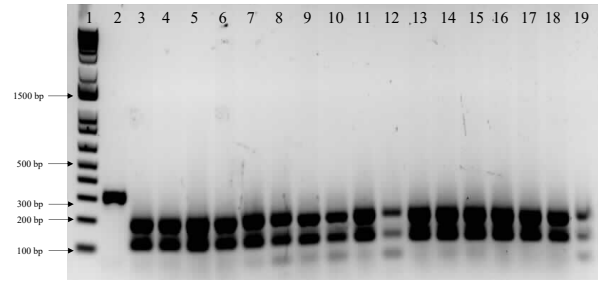


Supplementary figure 19. CAPS analysis to determine gene edits in regenerated shoots from infected leaves with the different Cas9-RGR constructs delivered by PVX and/or TRV. All samples exhibit two bands of 199 bp and 201 bp, indicating no gene editing events. In all agarose gels Lane 1 is 1 kb plus DNA ladder and Lane 2 is an undigested sample used as control of size comparison if a mutation is detected (320 bp). **A.** Lanes 3 – 20. Shoots regenerated from TRV *Sa*Cas9 full-length TobFT1 HH infected leaves. **B.** Lanes 3 – 5. Shoots regenerated from TRV *Sa*Cas9 full-length TobFT1 HH infected leaves. Lanes 6 – 8. Shoots regenerated from PVX *Sp*Cas9 full-length TobFT1 HH infected leaves. Lanes 9 – 20. Shoots regenerated from TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH infected leaves. **C.** Lanes 3 – 10. Shoots regenerated from TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH infected leaves. Lanes 11 – 20. Shoots regenerated from TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4 HH infected leaves. **D.** Lanes 3 – 10. Shoots regenerated from TRV N-*Sp*Cas9 N-intein TobFT1 HH + TRV C-intein C-*Sp*Cas9 TobFT4 HH infected leaves

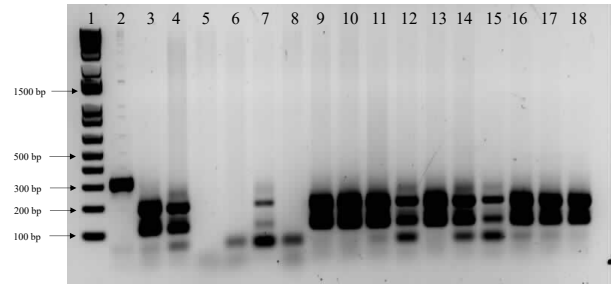
A



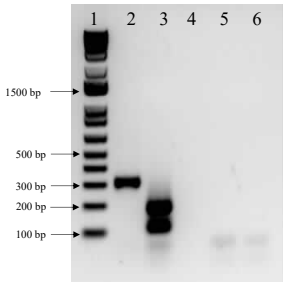
B



C



D



Supplementary figure 19 (continuation). CAPS analysis to determine gene edits in regenerated shoots from infected leaves with the different Cas9-RGR constructs delivered by PVX and/or TRV. All samples exhibit two bands of 199 bp and 201 bp, indicating no gene editing events. In all agarose gels Lane 1 is 1 kb plus DNA ladder and Lane 2 is an undigested sample used as control of size comparison if a mutation is detected (320 bp). **A.** Lanes 3 – 20. Shoots regenerated from PVX *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH infected leaves. **B.** Lanes 3 – 4. Shoots regenerated from PVX *Sa*Cas9 739N TobFT1 HH + PVX *Sa*Cas9 740C TobFT4 HH infected leaves. Lanes 5 – 19. Shoots regenerated from PVX *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH infected leaves. **C.** Lanes 3 – 9. Shoots regenerated from TRV N-*Sp*Cas9 N-intein TobFT1 HH + PVX C-intein C-*Sp*Cas9 TobFT4 HH infected leaves. Lanes 10 – 18. Shoots regenerated from TRV *Sa*Cas9 739N TobFT1 HH + TRV *Sa*Cas9 740C TobFT4 HH infected leaves. **D.** Lane 3. WT control. Lane 4. Empty. Lane 5. Negative control of DNA extraction. Lane 6. Negative control of amplification

APPENDIX B. LIST OF PRIMERS

Appendix B1. General primers

	Primer	Sequence 5' -> 3'	Ta^G
M13	M13 F	TGTAAAACGACGGCCAGT	55°C
	M13 R	GGAAACAGCTATGACCAT	
PVX PP82	PP82 F	CAGTGTTGGCTTGCAAACTAG	55°C
	PP82 R	TTGTGGTAGTTGAGGTAGTTGACCC	
TRV2 PEBV promoter	TRV2 promoter F	AAA**GAATTC**GAGCATCTTGTTCTGGGGTTT	69°C*
	TRV2 promoter R	ATT**TCTAGA**TCGGGTAAGTGATACAGTAACC	
Colony PCR pCAMBIA1300 *Sp*Cas9 FL	Cas9 seq F5b	CAGCACAAGCACTACCTGGA	55°C
	M13 R	GGAAACAGCTATGACCAT	
Colony PCR pCAMBIA1300 N-*Sp*Cas9 N-intein	Cas9 seq F1b	CCCCAACTTCAAGAGCAACT	60°C
	Cas9 seq R4b	AGAATGGCGTGCAGCTCT	
Colony PCR pCAMBIA1300 C-intein C-*Sp*Cas9	Cas9 seq F4b	GAAAGCGAGTTCGTGTACGG	60°C
	Cas9 seq R1b	TCCAGAAAGTCGATGGGATT	
Colony PCR PVX (PP82 R) or TRV2 (TRV2 R seq) *Sp*Cas9 FL and C-intein C-*Sp*Cas9	Cas9 seq F5b	CAGCACAAGCACTACCTGGA	55°C
	PP82 R	TTGTGGTAGTTGAGGTAGTTGACCC	
	TRV2 R seq	CGAGAATGTCAATCTCGTAGG	
Colony PCR PVX (PP82 R) or TRV2 (TRV2 R seq) N-*Sp*Cas9 N-intein	Cas9 seq F2b	ACGAGAAGGTGCTGCCCAAG	55°C
	PP82 R	TTGTGGTAGTTGAGGTAGTTGACCC	
	TRV2 R seq	CGAGAATGTCAATCTCGTAGG	
Colony PCR PVX (PP82 R) or TRV2 (TRV2 R seq) *Sa*Cas9 FL and *Sa*Cas9 740C	SaCas9 4F	TGCTCAACAGGATCGAAGTGAA	55°C
	PP82 R	TTGTGGTAGTTGAGGTAGTTGACCC	
	TRV2 R seq	CGAGAATGTCAATCTCGTAGG	
Colony PCR PVX (PP82 R) or TRV2 (TRV2 R seq) *Sa*Cas9 739N	SaCas9 3F	CACGCTGAGGATGCTCTCAT	55°C
	PP82 R	TTGTGGTAGTTGAGGTAGTTGACCC	
	TRV2 R seq	CGAGAATGTCAATCTCGTAGG	

The overhangs for *EcoR*I (GAATTC) and *Xba*I (TCTAGA) are showed in bold. Ta^G Annealing temperature using GoTaq® G2 Flexi DNA Polymerase. (*) Ta using Phusion® High-Fidelity DNA Polymerase

Appendix B2. List of sgRNAs

Gene	*Sp*Cas9		*Sa*Cas9	
	Primer	**Sequence 5' -> 3'**	**Primer**	**Sequence 5' -> 3'**
NtFT4 TobFT1	TobCRISPR FT4 1 F	**GATTG**GCCCAAGCAACCCTAACCTG	SaCas9 TobFT1 F	**ATTG**GCCCAAGCAACCCTAACCTGA
	TobCRISPR FT4 1 R	**AAAC**CAGGTTAGGGTTGCTTGGGCC	SaCas9 TobFT1 R	**AAAC**TCAGGTTAGGGTTGCTTGGGC
NtFT4 TobFT4	TobCRISPR FT4 4 F	**GATTG**GAGAATCCACAACCATCATT	SaCas9 TobFT4 F	**ATTG**GCGAGAATCCACAACCATCATT
	TobCRISPR FT4 4 R	**AAAC**AATGATGGTTGTGGATTCTCC	SaCas9 TobFT4 R	**AAAC**AATGATGGTTGTGGATTCTCGC
NtPDS	SpCas9 PDS F	**GATTG**TGCGATGCCTAACAAGCCAG	SaCas9 PDS F	**ATTG**GTTGCGATGCCTAACAAGCCAG
	SpCas9 PDS R	**AAAC**CTGGCTTGTTAGGCATCGCAC	SaCas9 PDS R	**AAAC**CTGGCTTGTTAGGCATCGCAAC

The overhangs for *BbsI* are showed in bold. The additional G necessary for higher AtU6-26 promoter efficiency is underlined

Appendix B3. List of primers used for overlapping extension PCR and cloning into pCAMBIA1300 vector

Gene	Primer	Sequence 5' -> 3'	Length (bp)	Ta^P
N-SpCas9	N intein Cas9 F1	TTA**GGTACCATCGAT**ATGGACTATAAGGACCAC	1836 bp	70°C
	N intein Cas9 R2	TTCATAGCTTAAACACTCGATTTTCTTGAA		
N-intein	N intein Cas9 F2	TTCAAGAAAATCGAGTGTTTAAGCTATGAA	309 bp	70°C
	N intein Cas9 R	*TTA*ATTCGGCAAATTATCAACCCGCATC		
C-SpCas9	C intein Cas9 F2 corrected	AGCTTCTAATTGTTTCAATTGCTTCGACTCCGT	2436 bp	70°C
	C intein Cas9 R	*TTA*CTTTTTCTTTTTTGCCTGGCCGGC		
C-intein	C intein Cas9 F1	TTA**GGTACCATCGAT**ATGATCAAAATAGCCACAC	117 bp	70°C
	C intein Cas9 R2 corrected	ACGGAGTCGAAGCAATTGAAACAATTAGAAGCT		
N-SpCas9 N-intein	N intein Cas9 F1	TTA**GGTACCATCGAT**ATGGACTATAAGGACCAC	2145 bp	70°C
	N intein Cas9 R	*TTA*ATTCGGCAAATTATCAACCCGCATC		
	N intein Cas9 primer F XhoI	AAAA**CTCGAG**ATGGACTATAAGGACCACGACGG		65°C
	N intein Cas9 primer R XhoI	AAAA**CTCGAGACGCGT**TTAATTCGGCAAATTATCAACCCGCATC		
C-intein C-SpCas9	C intein Cas9 F1	TTA**GGTACCATCGAT**ATGATCAAAATAGCCACAC	2553 bp	70°C
	C intein Cas9 R	*TTA*CTTTTTCTTTTTTGCCTGGCCGGC		
	C intein Cas9 primer F XhoI	AAAA**CTCGAG**ATGATCAAAATAGCCACACGTAAAT		65°C
	C intein Cas9 primer R XhoI	AAAA**CTCGAGACGCGT**TTACTTTTTCTTTTTTGCCTGGCCGGC		

Restriction sites *Kpn*I (GGTACC), *Cla*I (ATCGAT), *Xho*I (CTCGAG) and *Mlu*I (ACGCGT) are shown in bold. The overlapping overhangs are underlined, while in italics it is indicated the stop codon added.
Ta^P Annealing temperature is using Phusion® High-Fidelity DNA Polymerase

Appendix B4. List of primers to create the *Sp*Cas9 Ribozyme sgRNA Ribozyme (RGR) unit

Forward Primer	Sequence 5' -> 3'	Reverse Primer	Sequence 5' -> 3'	Ta^P
1st Tob1 RZF	GAGGACGAAACGAGTAAGCTCGTCGCCC AAGCAACCCTAACCTG	1st general RZR	TCAGACCGGAAAGCACATCCGGTGACA GGAAAGCACCGACTCGGTGCCACTTT	65°C
1st Tob4 RZF	GAGGACGAAACGAGTAAGCTCGTCGAGA ATCCACAACCATCATT			65°C
2nd Tob1 RZF SalI	AAA**GTCGAC***TTGGGC*CTGATGAGTCCGT GAGGACGAAACGAGTAAGCT	2nd general RZR (MluI)	TCT**ACGCGT**CCTGTTTCGTCCTCACGGA CTCATCAGACCGGAAAGCACATCC	65°C
2nd Tob4 RZF SalI	AAA**GTCGAC***ATTCTC*CTGATGAGTCCGTG AGGACGAAACGAGTAAGCT			65°C
2nd Tob1 RZF XmaI	AAA**CCCGGG***TTGGGC*CTGATGAGTCCGT GAGGACGAAACGAGTAAGCT	2nd general (XmaI)	AAA**CCCGGG**CCTGTTTCGTCCTCACGG ACTCATCAGACCGGAAAGCACATCC	65°C
2nd Tob4 RZF XmaI	AAA**CCCGGG***ATTCTC*CTGATGAGTCCGTG AGGACGAAACGAGTAAGCT			65°C

In bold it is shown the restriction sites for *SalI* (GTCGAC), *MluI* (ACGCGT) and *XmaI* (CCCGGG) added for downstream cloning. In italic it is indicated the six complementary nucleotides necessaries for the formation of the hammerhead structure. Underlined it is presented the overhangs nucleotides to create the RGR unit.
Ta^P Annealing temperature is using Phusion® High-Fidelity DNA Polymerase

Appendix B5. List of primers to create the *Sa*Cas9 Ribozyme sgRNA Ribozyme (RGR) unit

Primer	Sequence 5' -> 3'	Primer	Sequence 5' -> 3'	Ta^P
1st TobFT1 SaCas9 RZF	GAGGACGAAACGAGTAAGCTCGTCGCC CAAGCAACCCTAACCTGA	1st general SaCas9 RZR noU	TCAGACCGGAAAGCACATCCGGTGACA GGTCTCGCCAACAAGTTGACGAGA	65°C
1st TobFT4 SaCas9 RZF	GAGGACGAAACGAGTAAGCTCGTCCGA GAATCCACAACCATCATT			65°C
2nd TobFT1 RZF MluI	AAA**ACGCGT***TTGGGC*CTGATGAGTCCGT GAGGACGAAACGAGTAAGCT	2nd general XhoI	AAAA**CTCGAG**CCTGTTTCGTCCTCACGG ACTCATCAGACCGGAAAGCACATCC	65°C
2nd TobFT4 RZF MluI	AAA**ACGCGT***TTCTCG*CTGATGAGTCCGT GAGGACGAAACGAGTAAGCT			65°C

In bold it is shown the restriction sites for *Mlu*I (ACGCGT) and *Xho*I (CTCGAG) added for downstream cloning. In italic it is indicated the six complementary nucleotides necessaries for the formation of the hammerhead structure. Underlined it is presented the overhangs nucleotides to create the RGR unit. Ta^P Annealing temperature is using Phusion® High-Fidelity DNA Polymerase

Appendix B6. List of primers used to amplify *Sp*Cas9 full-length-RGR to clone into PVX

Primer	Sequence 5' -> 3'	Length (bp)	Ta^P
N intein Cas9 F1	TTA**GGTACCATCGAT**ATGGACTATAAGGACCAC	4491 bp	70°C
2nd general SalI	AAA**GTCGAC**CCTGTTTCGTCCTCACGGACTCATCAGACCGGAAAGCACATCC		

In bold it is shown the restriction sites for *Kpn*I (GGTACC), *Cla*I (ATCGAT) and *Sal*I (GTCGAC) added for downstream cloning
Ta^P Annealing temperature is using Phusion® High-Fidelity DNA Polymerase

Appendix B7. List of primers used for cloning *Sa*Cas9 full-length, *Sa*Cas9 739N and *Sa*Cas9 740C into TRV2 vector

Gene	Primer	Sequence 5' -> 3'	Length (bp)	Ta^P
SaCas9 full-length *SaCas9 740C*	SaCas9 FL F EagI KpnI	AAA**CGGCCGGGTACC**ATGGATTACAAGGATCACGATGGTGATT	3306 bp	66°C
	SaCas9 FL R SacI EagI	TTT**CGGCCGGAGCTC**TCAACCCTTCTTAATGATCTGAGGGT	1092 bp	66°C
SaCas9 739N	SaCas9 FL F EagI KpnI	AAA**CGGCCGGGTACC**ATGGATTACAAGGATCACGATGGTGATT	2364 bp	66°C
	SaCas9 739N R SacI EagI	TTT**CGGCCGGAGCTC**TCACTCAGCCTGCTTTTCCTCGAACA		

In bold it is shown the restriction sites for *Eag*I (CGGCCG), *Kpn*I (GGTACC) and *Sac*I (GAGCTC) added for downstream cloning.
Ta^P Annealing temperature is using Phusion® High-Fidelity DNA Polymerase

Appendix B8. List of primers used for cloning *Sa*Cas9 full-length, *Sa*Cas9 739N and *Sa*Cas9 740C with the RGR units into PVX vector

Gene	Primer	Sequence 5' -> 3'	Length (bp)	Ta^P
SaCas9 740C RGR	SaCas9 740C F ClaI	AATT**ATCGAT**ATGGATTACAAGGATCACGATGG	1340 bp	70°C
	2^{nd} general SalI	AAA**GTCGAC**CCTGTTTCGTCCTCACGGACTCATCAGACCG GAAAGCACATCC		
SaCas9 full-length RGR *SaCas9 739N RGR*	PVX SaCas9 sgRNA NEB F	CCACGGAATCGATACGCGTCATGGATTACAAGGATCACGA TG	3554 bp	62.7°C
	PVX SaCas9 sgRNA NEB R	ACTTAACCGTTCATCGGCGGCCTGTTTCGTCCTCACGG	2612 bp	62.7°C

In bold it is shown the restriction sites for *Cla*I (ATCGAT) and *Sal*I (GTCGAC) added for downstream cloning.
Ta^P Annealing temperature is using Phusion® High-Fidelity DNA Polymerase

Appendix B9. List of primers used for RT-PCR

Gene	Primer	Sequence 5' -> 3'	Length (bp)	Ta^G
EF1α	EFNB1 Forward	CTCCAAGGCTAGGTATGATG	372 bp	60°C
	EFNB2 Reverse	CTTCGTGGTTGCATCTCAAC		
SpCas9 N- terminal	Cas9 seq F1b	CCCCAACTTCAAGAGCAACT	522 bp	60°C
	Cas9 seq R4b	AGAATGGCGTGCAGCTCT		
	Cas9 seq F1	ATCTTCGGCAACATCGTGGA	521 bp	60°C
	Cas9 seq R5	AGAAACAGGTCGGCGTACTG		
SpCas9 C-terminal	Cas9 seq F3b	CGAGAACATCGTGATCGAAAT	474 bp	60°C
	Cas9 internal 1R	GCCTTATCCAGTTCGCTCAG		
SaCas9 N-terminal	SaCas9 1F	GGTGACCTCTACCGGAAAGC	385 bp	60°C
	SaCas9 2R	ATCCACGAGGGTGGTAGGAA		
SaCas9 C-terminal	SaCas9 2F	GCCTGAGATCGAGACTGAGC	449 bp	60°C
	SaCas9 3R	GGTGTGCGTTGAGCTTGTTT		
	SaCas9 5F	AGGGTAACACCCTCATCGTG	408 bp	65°C
	SaCas9 5R	CGAGGTTCTTCACGGTCACG		
PVX coat protein	PVX CP 1F	GCTTCAGGCCTGTTCACCAT	557 bp	65°C
	PVX CP 1R	TCTAGGCTGGCAAAGTCGTT		
TRV1 replicase	TRV1 replicase Forw	GAGTCCGGTGAGACCGTTTT	357 bp	64.3°C
	TRV1 replicase Rev	TGTACCGCTTGTTTCCCCTC		
TRV2 coat protein	TRV2 CP forw	CTGGGTTACTAGCGGCACTGAATA	401 bp	64.3°C
	TRV2 CP rev	TCCACCAAACTTAATCCCGAATAC		

Ta^G Annealing temperature using GoTaq® G2 Flexi DNA Polymerase

Appendix B10. List of primers used for circular RT-PCR (cRT-PCR)

Primer	Sequence 5' -> 3'	Primer	Sequence 5' -> 3'	Ta^{Q5}
Scaffold SpCas9 F	GTTTTAGAGCTAGAAATAGCAAG	TobCRISPR FT4 1 R	AAACCAGGTTAGGGTTGCTTGGGCC	59°C
		TobCRISPR FT4 4 R	AAACAATGATGGTTGTGGATTCTCC	59°C

Ta^{Q5} Annealing temperature using Q5® High-Fidelity DNA Polymerase

Appendix B11. List of primers used for gene editing screening

Gene	Primer	Sequence 5' -> 3'	Length (bp)	Ta^{G}	Ta^{Q5}
NtPDS	PDS forw	AGCTGCATGGAAAGATGATGA	886 bp	55°C	66°C
	PDS rev	TGCTTTCTCATCCAGTCCTTAACA			
NtFT4 exon 2	Tob FT4 3F	CCCCATATTCCAAATCCCACCA	320 bp	55°C	66°C
	Tob FT4 4R	TTTTTCCAGCAGTATGTGATAGGTCT			
NtFT4 exon 4	NtabFT4 CRISPR4 F4	GGACCACAAGGGTCTATCGT	309 bp	55°C	66°C
	TobFT4 1R	CTGAAATTCTGACGCCAACC			

Ta^{G} Annealing temperature using GoTaq® G2 Flexi DNA Polymerase
Ta^{Q5} Annealing temperature using Q5® High-Fidelity DNA Polymerase

REFERENCES

1. World Health Organization. Human genome editing 2022 [Available from: https://www.who.int/health-topics/human-genome-editing#tab=tab_1.
2. Yee JK. Off-target effects of engineered nucleases. FEBS J. 2016;283(17):3239-48.
3. Sprink T, Eriksson D, Schiemann J, Hartung F. Regulatory hurdles for genome editing: process-vs. product-based approaches in different regulatory contexts. Plant cell reports. 2016;35(7):1493-506.
4. Shillito RD, Whitt S, Ross M, Ghavami F, De Vleesschauwer D, D'Halluin K, et al. Detection of genome edits in plants—from editing to seed. In Vitro Cellular & Developmental Biology-Plant. 2021;57(4):595-608.
5. Thomas KR, Folger KR, Capecchi MR. High frequency targeting of genes to specific sites in the mammalian genome. Cell. 1986;44(3):419-28.
6. Jasin M. Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet. 1996;12(6):224-8.
7. Bak RO, Gomez-Ospina N, Porteus MH. Gene Editing on Center Stage. Trends Genet. 2018;34(8):600-11.
8. Belfort M, Bonocora RP. Homing endonucleases: from genetic anomalies to programmable genomic clippers. Methods Mol Biol. 2014;1123:1-26.
9. Carroll D. Genome engineering with zinc-finger nucleases. Genetics. 2011;188(4):773-82.
10. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A. 1996;93(3):1156-60.
11. Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, et al. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol. 2001;21(1):289-97.
12. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 2009;326(5959):1509-12.
13. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science. 2009;326(5959):1501.
14. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. 2010;186(2):757-61.
15. Li T, Huang S, Jiang WZ, Wright D, Spalding MH, Weeks DP, et al. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res. 2011;39(1):359-72.
16. Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu JK. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A. 2011;108(6):2623-8.
17. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol. 2011;29(2):143-8.
18. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 2011;39(12):e82.
19. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in

Escherichia coli, and identification of the gene product. J Bacteriol. 1987;169(12):5429-33.
20. Mojica FJ, Juez G, Rodriguez-Valera F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Mol Microbiol. 1993;9(3):613-21.
21. Mojica FJ, Diez-Villasenor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol. 2000;36(1):244-6.
22. Jansen R, Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol. 2002;43(6):1565-75.
23. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol. 2005;60(2):174-82.
24. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology. 2005;151(Pt 3):653-63.
25. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151(Pt 8):2551-61.
26. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct. 2006;1:7.
27. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315(5819):1709-12.
28. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816-21.
29. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A. 2012;109(39):E2579-86.
30. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol. 2013;31(3):230-2.
31. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23.
32. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339(6121):823-6.
33. Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol. 2013;31(8):686-8.
34. Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, et al. Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol. 2013;31(8):688-91.
35. Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nature Biotechnology. 2013;31(8):691-3.

36. Barrangou R. RNA-mediated programmable DNA cleavage. Nat Biotechnol. 2012;30(9):836-8.
37. Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. J Clin Invest. 2014;124(10):4154-61.
38. Pattanayak V, Ramirez CL, Joung JK, Liu DR. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat Methods. 2011;8(9):765-70.
39. Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaeppel C, et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol. 2011;29(9):816-23.
40. Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. 2011;39(21):9283-93.
41. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, et al. Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol. 2011;29(8):731-4.
42. Suzuki K, Yu C, Qu J, Li M, Yao X, Yuan T, et al. Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. Cell Stem Cell. 2014;15(1):31-6.
43. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013;31(9):822-6.
44. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31(9):827-32.
45. Zhang JH, Pandey M, Kahler JF, Loshakov A, Harris B, Dagur PK, et al. Improving the specificity and efficacy of CRISPR/CAS9 and gRNA through target specific DNA reporter. J Biotechnol. 2014;189:1-8.
46. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol. 2013;31(9):839-43.
47. Xu W, Fu W, Zhu P, Li Z, Wang C, Wang C, et al. Comprehensive Analysis of CRISPR/Cas9-Mediated Mutagenesis in Arabidopsis thaliana by Genome-wide Sequencing. Int J Mol Sci. 2019;20(17):4125.
48. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014;32(3):279-84.
49. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154(6):1380-9.
50. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol. 2013;31(9):833-8.
51. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 2014;24(1):132-41.
52. Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat Biotechnol. 2014;32(6):569-76.

53. Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat Biotechnol. 2014;32(6):577-82.
54. Zhu H, Li C, Gao C. Applications of CRISPR–Cas in agriculture and plant biotechnology. Nature Reviews Molecular Cell Biology. 2020;21(11):661-77.
55. Hahn F, Nekrasov V. CRISPR/Cas precision: do we need to worry about off-targeting in plants? Plant Cell Reports. 2019;38(4):437-41.
56. Hajiahmadi Z, Movahedi A, Wei H, Li D, Orooji Y, Ruan H, et al. Strategies to increase on-target and reduce off-target effects of the CRISPR/Cas9 system in plants. Int J Mol Sci. 2019;20(15):3719.
57. Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang D-L, et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. Proceedings of the National Academy of Sciences. 2014;111(12):4632-7.
58. Peterson BA, Haak DC, Nishimura MT, Teixeira PJ, James SR, Dangl JL, et al. Genome-wide assessment of efficiency and specificity in CRISPR/Cas9 mediated multiple site targeting in Arabidopsis. PloS one. 2016;11(9):e0162169.
59. Ishino Y, Krupovic M, Forterre P. History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. J Bacteriol. 2018;200(7).
60. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015;523(7561):481-5.
61. Muller M, Lee CM, Gasiunas G, Davis TH, Cradick TJ, Siksnys V, et al. Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. Mol Ther. 2016;24(3):636-44.
62. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods. 2013;10(11):1116-21.
63. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015;520(7546):186-91.
64. Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, Sontheimer EJ, et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc Natl Acad Sci U S A. 2013;110(39):15644-9.
65. Kim E, Koo T, Park SW, Kim D, Kim K, Cho HY, et al. In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nat Commun. 2017;8:14500.
66. Chatterjee P, Jakimo N, Jacobson JM. Minimal PAM specificity of a highly similar SpCas9 ortholog. Sci Adv. 2018;4(10):eaau0766.
67. Harrington LB, Paez-Espino D, Staahl BT, Chen JS, Ma E, Kyrpides NC, et al. A thermostable Cas9 with increased lifetime in human plasma. Nat Commun. 2017;8(1):1424.
68. Hirano H, Gootenberg JS, Horii T, Abudayyeh OO, Kimura M, Hsu PD, et al. Structure and Engineering of Francisella novicida Cas9. Cell. 2016;164(5):950-61.
69. Fedorova I, Arseniev A, Selkova P, Pobegalov G, Goryanin I, Vasileva A, et al. DNA targeting by Clostridium cellulolyticum CRISPR-Cas9 Type II-C system. Nucleic Acids Res. 2020;48(4):2026-34.

70. Karvelis T, Gasiunas G, Young J, Bigelyte G, Silanskas A, Cigan M, et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol. 2015;16:253.
71. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163(3):759-71.
72. Strecker J, Jones S, Koopal B, Schmid-Burgk J, Zetsche B, Gao L, et al. Engineering of CRISPR-Cas12b for human genome editing. Nat Commun. 2019;10(1):212.
73. Teng F, Cui T, Feng G, Guo L, Xu K, Gao Q, et al. Repurposing CRISPR-Cas12b for mammalian genome engineering. Cell Discov. 2018;4:63.
74. Pausch P, Al-Shayeb B, Bisom-Rapp E, Tsuchida CA, Li Z, Cress BF, et al. CRISPR-CasPhi from huge phages is a hypercompact genome editor. Science. 2020;369(6501):333-7.
75. Liu JJ, Orlova N, Oakes BL, Ma E, Spinner HB, Baney KLM, et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. Nature. 2019;566(7743):218-23.
76. Fonfara I, Richter H, Bratovic M, Le Rhun A, Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature. 2016;532(7600):517-21.
77. East-Seletsky A, O'Connell MR, Burstein D, Knott GJ, Doudna JA. RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. Mol Cell. 2017;66(3):373-83 e3.
78. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014;156(5):935-49.
79. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science. 2014;343(6176):1247997.
80. Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, et al. Crystal Structure of Staphylococcus aureus Cas9. Cell. 2015;162(5):1113-26.
81. Wright AV, Sternberg SH, Taylor DW, Staahl BT, Bardales JA, Kornfeld JE, et al. Rational design of a split-Cas9 enzyme complex. Proc Natl Acad Sci U S A. 2015;112(10):2984-9.
82. Zetsche B, Volz SE, Zhang F. A split-Cas9 architecture for inducible genome editing and transcription modulation. Nat Biotechnol. 2015;33(2):139-42.
83. Truong DJ, Kuhner K, Kuhn R, Werfel S, Engelhardt S, Wurst W, et al. Development of an intein-mediated split-Cas9 system for gene therapy. Nucleic Acids Res. 2015;43(13):6450-8.
84. Fine EJ, Appleton CM, White DE, Brown MT, Deshmukh H, Kemp ML, et al. Trans-spliced Cas9 allows cleavage of HBB and CCR5 genes in human cells using compact expression cassettes. Sci Rep. 2015;5:10777.
85. Chew WL, Tabebordbar M, Cheng JK, Mali P, Wu EY, Ng AH, et al. A multifunctional AAV-CRISPR-Cas9 and its host response. Nat Methods. 2016;13(10):868-74.
86. Ma D, Peng S, Xie Z. Integration and exchange of split dCas9 domains for transcriptional controls in mammalian cells. Nat Commun. 2016;7:13056.
87. Davis KM, Pattanayak V, Thompson DB, Zuris JA, Liu DR. Small molecule-triggered Cas9 protein with improved genome-editing specificity. Nat Chem Biol. 2015;11(5):316-8.

88. Mills KV, Johnson MA, Perler FB. Protein splicing: how inteins escape from precursor proteins. J Biol Chem. 2014;289(21):14498-505.
89. Shah NH, Muir TW. Inteins: Nature's Gift to Protein Chemists. Chem Sci. 2014;5(1):446-61.
90. Zettler J, Schutz V, Mootz HD. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Lett. 2009;583(5):909-14.
91. Shah NH, Dann GP, Vila-Perello M, Liu Z, Muir TW. Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering. J Am Chem Soc. 2012;134(28):11338-41.
92. Cheriyan M, Pedamallu CS, Tori K, Perler F. Faster protein splicing with the Nostoc punctiforme DnaE intein using non-native extein residues. J Biol Chem. 2013;288(9):6202-11.
93. Iwai H, Zuger S, Jin J, Tam PH. Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. FEBS Lett. 2006;580(7):1853-8.
94. Xing H-L, Dong L, Wang Z-P, Zhang H-Y, Han C-Y, Liu B, et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC plant biology. 2014;14(1):1-12.
95. Ma X, Liu YG. CRISPR/Cas9-based multiplex genome editing in monocot and dicot plants. Current protocols in molecular biology. 2016;115(1):31.6. 1-.6. 21.
96. Xie K, Minkenberg B, Yang Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proceedings of the National Academy of Sciences. 2015;112(11):3570-5.
97. Ellison EE, Nagalakshmi U, Gamo ME, Huang P-j, Dinesh-Kumar S, Voytas DF. Multiplexed heritable gene editing using RNA viruses and mobile single guide RNAs. Nature plants. 2020;6(6):620-4.
98. Gao Y, Zhao Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. J Integr Plant Biol. 2014;56(4):343-9.
99. Khvorova A, Lescoute A, Westhof E, Jayasena SD. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol. 2003;10(9):708-12.
100. He Y, Zhang T, Yang N, Xu M, Yan L, Wang L, et al. Self-cleaving ribozymes enable the production of guide RNAs from unlimited choices of promoters for CRISPR/Cas9 mediated genome editing. J Genet Genomics. 2017;44(9):469-72.
101. Tang X, Zheng X, Qi Y, Zhang D, Cheng Y, Tang A, et al. A Single Transcript CRISPR-Cas9 System for Efficient Genome Editing in Plants. Mol Plant. 2016;9(7):1088-91.
102. Mikami M, Toki S, Endo M. In Planta Processing of the SpCas9-gRNA Complex. Plant Cell Physiol. 2017;58(11):1857-67.
103. Zhang Y, Ren Q, Tang X, Liu S, Malzahn AA, Zhou J, et al. Expanding the scope of plant genome engineering with Cas12a orthologs and highly multiplexable editing systems. Nature communications. 2021;12(1):1-11.
104. Chen Q, Lai H. Gene delivery into plant cells for recombinant protein production. Biomed Res Int. 2015;2015:932161.
105. Komarova TV, Baschieri S, Donini M, Marusic C, Benvenuto E, Dorokhov YL. Transient expression systems for plant-derived biopharmaceuticals. Expert Rev Vaccines. 2010;9(8):859-76.

106. Hefferon K. Plant Virus Expression Vectors: A Powerhouse for Global Health. Biomedicines. 2017;5(3).
107. Chapman S, Kavanagh T, Baulcombe D. Potato virus X as a vector for gene expression in plants. Plant J. 1992;2(4):549-57.
108. Martin-Hernandez AM, Baulcombe DC. Tobacco rattle virus 16-kilodalton protein encodes a suppressor of RNA silencing that allows transient viral entry in meristems. J Virol. 2008;82(8):4064-71.
109. MacFarlane SA, Popovich AH. Efficient expression of foreign proteins in roots from tobravirus vectors. Virology. 2000;267(1):29-35.
110. Avesani L, Marconi G, Morandini F, Albertini E, Bruschetta M, Bortesi L, et al. Stability of Potato virus X expression vectors is related to insert size: implications for replication models and risk assessment. Transgenic Res. 2007;16(5):587-97.
111. Dickmeis C, Fischer R, Commandeur U. Potato virus X-based expression vectors are stabilized for long-term production of proteins and larger inserts. Biotechnol J. 2014;9(11):1369-79.
112. Peyret H, Lomonossoff GP. When plant virology met Agrobacterium: the rise of the deconstructed clones. Plant Biotechnol J. 2015;13(8):1121-35.
113. Jeffries CJ, Food, Nations AOotU, Institute IPGR. Potato: Food and Agriculture Organization of the United Nations; 1998.
114. Baulcombe DC, Chapman S, Santa Cruz S. Jellyfish green fluorescent protein as a reporter for virus infections. Plant J. 1995;7(6):1045-53.
115. Lu R, Malcuit I, Moffett P, Ruiz MT, Peart J, Wu AJ, et al. High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. EMBO J. 2003;22(21):5690-9.
116. Lacorte C, Ribeiro SG, Lohuis D, Goldbach R, Prins M. Potatovirus X and Tobacco mosaic virus-based vectors compatible with the Gateway cloning system. J Virol Methods. 2010;164(1-2):7-13.
117. Zaidi SS, Mansoor S. Viral Vectors for Plant Genome Engineering. Front Plant Sci. 2017;8:539.
118. MacFarlane SA. Molecular biology of the tobraviruses. J Gen Virol. 1999;80 (Pt 11):2799-807.
119. Ratcliff FG, MacFarlane SA, Baulcombe DC. Gene silencing without DNA: RNA-mediated cross-protection between viruses. Plant Cell. 1999;11(7):1207-16.
120. Ratcliff F, Martin-Hernandez AM, Baulcombe DC. Technical Advance. Tobacco rattle virus as a vector for analysis of gene function by silencing. Plant J. 2001;25(2):237-45.
121. Liu Y, Schiff M, Marathe R, Dinesh-Kumar SP. Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. Plant J. 2002;30(4):415-29.
122. Liu Y, Schiff M, Dinesh-Kumar SP. Virus-induced gene silencing in tomato. Plant J. 2002;31(6):777-86.
123. Sha A, Zhao J, Yin K, Tang Y, Wang Y, Wei X, et al. Virus-based microRNA silencing in plants. Plant Physiol. 2014;164(1):36-47.
124. Gantner J, Ordon J, Ilse T, Kretschmer C, Gruetzner R, Lofke C, et al. Peripheral infrastructure vectors and an extended set of plant parts for the Modular Cloning system. PLoS One. 2018;13(5):e0197185.
125. Marton I, Zuker A, Shklarman E, Zeevi V, Tovkach A, Roffe S, et al. Nontransgenic genome modification in plant cells. Plant Physiol. 2010;154(3):1079-87.

126. Honig A, Marton I, Rosenthal M, Smith JJ, Nicholson MG, Jantz D, et al. Transient Expression of Virally Delivered Meganuclease In Planta Generates Inherited Genomic Deletions. Mol Plant. 2015;8(8):1292-4.
127. Ali Z, Abul-Faraj A, Idris A, Ali S, Tashkandi M, Mahfouz MM. CRISPR/Cas9-mediated viral interference in plants. Genome Biol. 2015;16:238.
128. Ali Z, Abul-faraj A, Li L, Ghosh N, Piatek M, Mahjoub A, et al. Efficient Virus-Mediated Genome Editing in Plants Using the CRISPR/Cas9 System. Mol Plant. 2015;8(8):1288-91.
129. Ali Z, Abul-Faraj A, Piatek M, Mahfouz MM. Activity and specificity of TRV-mediated gene editing in plants. Plant Signal Behav. 2015;10(10):e1044191.
130. Cody WB, Scholthof HB, Mirkov TE. Multiplexed Gene Editing and Protein Overexpression Using a Tobacco mosaic virus Viral Vector. Plant Physiol. 2017;175(1):23-35.
131. Ali Z, Eid A, Ali S, Mahfouz MM. Pea early-browning virus-mediated genome editing via the CRISPR/Cas9 system in Nicotiana benthamiana and Arabidopsis. Virus Res. 2018;244:333-7.
132. Hu J, Li S, Li Z, Li H, Song W, Zhao H, et al. A barley stripe mosaic virus-based guide RNA delivery system for targeted mutagenesis in wheat and maize. Mol Plant Pathol. 2019;20(10):1463-74.
133. Jiang N, Zhang C, Liu JY, Guo ZH, Zhang ZY, Han CG, et al. Development of Beet necrotic yellow vein virus-based vectors for multiple-gene expression and guide RNA delivery in plant genome editing. Plant Biotechnol J. 2019;17(7):1302-15.
134. Uranga M, Aragonés V, Selma S, Vázquez-Vilar M, Orzáez D, Daròs JA. Efficient Cas9 multiplex editing using unspaced sgRNA arrays engineering in a Potato virus X vector. The Plant Journal. 2021;106(2):555-65.
135. Lacomme C, Chapman S. Use of potato virus X (PVX)-based vectors for gene expression and virus-induced gene silencing (VIGS). Curr Protoc Microbiol. 2008;Chapter 16:Unit 16I 1.
136. Zhang X, Kang L, Zhang Q, Meng Q, Pan Y, Yu Z, et al. An RNAi suppressor activates in planta virus-mediated gene editing. Funct Integr Genomics. 2019.
137. Ma X, Zhang X, Liu H, Li Z. Highly efficient DNA-free plant genome editing using virally delivered CRISPR–Cas9. Nature Plants. 2020;6(7):773-9.
138. Gao Q, Xu WY, Yan T, Fang XD, Cao Q, Zhang ZJ, et al. Rescue of a plant cytorhabdovirus as versatile expression platforms for planthopper and cereal genomic studies. New Phytologist. 2019;223(4):2120-33.
139. Ariga H, Toki S, Ishibashi K. Potato virus X Vector-Mediated DNA-Free genome editing in plants. Plant and Cell Physiology. 2020;61(11):1946-53.
140. Zhang X-F, Qu F. Cross Protection of Plant Viruses: Recent Developments and Mechanistic Implications. Current Research Topics in Plant Virology2016. p. 241-50.
141. Zhang X-F, Zhang S, Guo Q, Sun R, Wei T, Qu F. A new mechanistic model for viral cross protection and superinfection exclusion. Frontiers in plant science. 2018;9:40.
142. Kaya H, Mikami M, Endo A, Endo M, Toki S. Highly specific targeted mutagenesis in plants using Staphylococcus aureus Cas9. Sci Rep. 2016;6:26871.
143. Kaya H, Ishibashi K, Toki S. A Split Staphylococcus aureus Cas9 as a Compact Genome-Editing Tool in Plants. Plant Cell Physiol. 2017;58(4):643-9.

144. Datta A. Genetic engineering for improving quality and productivity of crops. Agriculture & Food Security. 2013;2(1):15.
145. Son S, Park SR. Challenges Facing CRISPR/Cas9-Based Genome Editing in Plants. Frontiers in Plant Science. 2022;13:902413.
146. Wang T, Zhang H, Zhu H. CRISPR technology is revolutionizing the improvement of tomato and other fruit crops. Horticulture research. 2019;6.
147. Wolter F, Schindele P, Puchta H. Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. BMC plant biology. 2019;19(1):1-8.
148. Karkute SG, Singh AK, Gupta OP, Singh PM, Singh B. CRISPR/Cas9 mediated genome engineering for improvement of horticultural crops. Frontiers in plant science. 2017;8:1635.
149. Waltz E. Gene-edited CRISPR mushroom escapes US regulation. Nature News. 2016;532(7599):293.
150. Turnbull C, Lillemo M, Hvoslef-Eide TA. Global regulation of genetically modified crops amid the gene edited crop boom–a review. Frontiers in Plant Science. 2021;12:258.
151. Hill CB, Li C. Genetic Architecture of Flowering Phenology in Cereals and Opportunities for Crop Improvement. Front Plant Sci. 2016;7:1906.
152. Jung C, Muller AE. Flowering time control and applications in plant breeding. Trends Plant Sci. 2009;14(10):563-73.
153. Blumel M, Dally N, Jung C. Flowering time regulation in crops-what did we learn from Arabidopsis? Curr Opin Biotechnol. 2015;32:121-9.
154. Andres F, Coupland G. The genetic basis of flowering responses to seasonal cues. Nat Rev Genet. 2012;13(9):627-39.
155. Chen Q, Payyavula RS, Chen L, Zhang J, Zhang C, Turgeon R. FLOWERING LOCUS T mRNA is synthesized in specialized companion cells in Arabidopsis and Maryland Mammoth tobacco leaf veins. Proc Natl Acad Sci U S A. 2018;115(11):2830-5.
156. Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, et al. FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science. 2007;316(5827):1030-3.
157. Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K. Hd3a protein is a mobile flowering signal in rice. Science. 2007;316(5827):1033-6.
158. Sierro N, Battey JN, Ouadi S, Bakaher N, Bovet L, Willig A, et al. The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun. 2014;5:3833.
159. Harig L, Beinecke FA, Oltmanns J, Muth J, Muller O, Ruping B, et al. Proteins from the FLOWERING LOCUS T-like subclade of the PEBP family act antagonistically to regulate floral initiation in tobacco. Plant J. 2012;72(6):908-21.
160. Karlgren A, Gyllenstrand N, Kallman T, Sundstrom JF, Moore D, Lascoux M, et al. Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution. Plant Physiol. 2011;156(4):1967-77.
161. Zheng XM, Wu FQ, Zhang X, Lin QB, Wang J, Guo XP, et al. Evolution of the PEBP gene family and selective signature on FT-like clade. Journal of Systematics and Evolution. 2016;54(5):502-10.
162. Wang G, Wang P, Gao Y, Li Y, Wu L, Gao J, et al. Isolation and functional characterization of a novel FLOWERING LOCUS T homolog (NtFT5) in Nicotiana tabacum. J Plant Physiol. 2018;231:393-401.

163. Beinecke FA, Grundmann L, Wiedmann DR, Schmidt FJ, Caesar AS, Zimmermann M, et al. The FT/FD-dependent initiation of flowering under long-day conditions in the day-neutral species Nicotiana tabacum originates from the facultative short-day ancestor Nicotiana tomentosiformis. Plant J. 2018;96(2):329-42.
164. Zhang M, Li P, Yan X, Wang J, Cheng T, Zhang Q. Genome-wide characterization of PEBP family genes in nine Rosaceae tree species and their expression analysis in P. mume. BMC ecology and evolution. 2021;21(1):1-23.
165. Schmidt FJ, Zimmermann MM, Wiedmann DR, Lichtenauer S, Grundmann L, Muth J, et al. The Major Floral Promoter NtFT5 in Tobacco (Nicotiana tabacum) Is a Promising Target for Crop Improvement. Front Plant Sci. 2019;10:1666.
166. Muona M, Aranko AS, Raulinaitis V, Iwai H. Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro. Nat Protoc. 2010;5(3):574-87.
167. Park J, Bae S, Kim JS. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. Bioinformatics. 2015;31(24):4014-6.
168. Bae S, Park J, Kim JS. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics. 2014;30(10):1473-5.
169. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011;7(1):539.
170. Heckman KL, Pease LR. Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc. 2007;2(4):924-32.
171. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics. 2009;25(15):1974.
172. Hoefle C, Huesmann C, Schultheiss H, Bornke F, Hensel G, Kumlehn J, et al. A barley ROP GTPase ACTIVATING PROTEIN associates with microtubules and regulates entry of the barley powdery mildew fungus into leaf epidermal cells. Plant Cell. 2011;23(6):2422-39.
173. LAEMMLI UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. nature. 1970;227(5259):680-5.
174. Lodhi MA, Ye G-N, Weeden NF, Reisch BI. A simple and efficient method for DNA extraction from grapevine cultivars and Vitis species. Plant Molecular Biology Reporter. 1994;12(1):6-13.
175. Carlson PS. The use of protoplasts for genetic research. Proceedings of the National Academy of Sciences of the United States of America. 1973;70(2):598.
176. Reed KM, Bargmann BOR. Protoplast Regeneration and Its Use in New Plant Breeding Technologies. Frontiers in Genome Editing. 2021;3(20).
177. Yue J-J, Yuan J-L, Wu F-H, Yuan Y-H, Cheng Q-W, Hsu C-T, et al. Protoplasts: From Isolation to CRISPR/Cas Genome Editing Application. Frontiers in Genome Editing. 2021;3(19).
178. Nadakuduti SS, Starker CG, Ko DK, Jayakody TB, Buell CR, Voytas DF, et al. Evaluation of methods to assess in vivo activity of engineered genome-editing nucleases in protoplasts. Frontiers in plant science. 2019;10:110.
179. Lin CS, Hsu CT, Yang LH, Lee LY, Fu JY, Cheng QW, et al. Application of protoplast technology to CRISPR/Cas9 mutagenesis: from single-cell mutation detection to mutant plant regeneration. Plant Biotechnol J. 2018;16(7):1295-310.

180. Hsu C-T, Lee W-C, Cheng Y-J, Yuan Y-H, Wu F-H, Lin C-S. Genome editing and protoplast regeneration to study plant–pathogen interactions in the model plant Nicotiana benthamiana. Frontiers in Genome Editing. 2021;2:39.
181. Woo JW, Kim J, Kwon SI, Corvalán C, Cho SW, Kim H, et al. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nature biotechnology. 2015;33(11):1162-4.
182. Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q, et al. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nature communications. 2017;8(1):1-5.
183. Tanner NK. Ribozymes: the characteristics and properties of catalytic RNAs. FEMS microbiology reviews. 1999;23(3):257-75.
184. Norkunas K, Harding R, Dale J, Dugdale B. Improving agroinfiltration-based transient gene expression in Nicotiana benthamiana. Plant Methods. 2018;14:71.
185. Mei Y, Beernink BM, Ellison EE, Konečná E, Neelakandan AK, Voytas DF, et al. Protein expression and gene editing in monocots using foxtail mosaic virus vectors. Plant direct. 2019;3(11):e00181.
186. Garabagi F, Gilbert E, Loos A, McLean MD, Hall JC. Utility of the P19 suppressor of gene-silencing protein for production of therapeutic antibodies in Nicotiana expression hosts. Plant Biotechnol J. 2012;10(9):1118-28.
187. Angel CA, Hsieh YC, Schoelz JE. Comparative analysis of the capacity of tombusvirus P22 and P19 proteins to function as avirulence determinants in Nicotiana species. Mol Plant Microbe Interact. 2011;24(1):91-9.
188. Siddiqui SA, Sarmiento C, Truve E, Lehto H, Lehto K. Phenotypes and functional effects caused by various viral RNA silencing suppressors in transgenic Nicotiana benthamiana and N. tabacum. Mol Plant Microbe Interact. 2008;21(2):178-87.
189. Harries PA, Palanichelvam K, Bhat S, Nelson RS. Tobacco mosaic virus 126-kDa protein increases the susceptibility of Nicotiana tabacum to other viruses and its dosage affects virus-induced gene silencing. Mol Plant Microbe Interact. 2008;21(12):1539-48.
190. Korth KL, Jaggard DA, Dixon RA. Developmental and light-regulated post-translational control of 3-hydroxy-3-methylglutaryl-CoA reductase levels in potato. Plant J. 2000;23(4):507-16.
191. Liere K, Kaden D, Maliga P, Borner T. Overexpression of phage-type RNA polymerase RpoTp in tobacco demonstrates its role in chloroplast transcription by recognizing a distinct promoter type. Nucleic Acids Res. 2004;32(3):1159-65.
192. Gasic K, Korban SS. Nonspecific binding of monoclonal anti-FLAG M2 antibody in Indian mustard (Brassica juncea). Plant Molecular Biology Reporter. 2012;23(1):9-16.
193. Raitskin O, Schudoma C, West A, Patron NJ. Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: an expanded toolkit for precision genome engineering. PLoS One. 2019;14(2):e0211598.
194. Wu S, Zhu H, Liu J, Yang Q, Shao X, Bi F, et al. Establishment of a PEG-mediated protoplast transformation system based on DNA and CRISPR/Cas9 ribonucleoprotein complexes for banana. BMC plant biology. 2020;20(1):1-10.
195. Xing T, Wang X. Protoplasts in plant signaling analysis: moving forward in the omics era. Botany. 2015;93(6):325-32.
196. Brandt KM, Gunn H, Moretti N, Zemetra RS. A streamlined protocol for wheat (Triticum aestivum) protoplast isolation and transformation with CRISPR-Cas ribonucleoprotein complexes. Frontiers in Plant Science. 2020;11:769.

197. Modrzejewski D, Hartung F, Sprink T, Krause D, Kohl C, Wilhelm R. What is the available evidence for the range of applications of genome-editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map. Environmental Evidence. 2019;8(1):1-33.
198. Modrzejewski D, Hartung F, Lehnert H, Sprink T, Kohl C, Keilwagen J, et al. Which factors affect the occurrence of off-target effects caused by the use of CRISPR/Cas: a systematic review in plants. Frontiers in plant science. 2020;11:574959.
199. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011;471(7340):602-7.
200. Butler NM, Baltes NJ, Voytas DF, Douches DS. Geminivirus-mediated genome editing in potato (Solanum tuberosum L.) using sequence-specific nucleases. Frontiers in plant science. 2016;7:1045.
201. Gil-Humanes J, Wang Y, Liang Z, Shan Q, Ozuna CV, Sánchez-León S, et al. High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. The Plant Journal. 2017;89(6):1251-62.
202. Wang M, Lu Y, Botella JR, Mao Y, Hua K, Zhu J-k. Gene targeting by homology-directed repair in rice using a geminivirus-based CRISPR/Cas9 system. Molecular plant. 2017;10(7):1007-10.
203. Dahan-Meir T, Filler-Hayut S, Melamed-Bessudo C, Bocobza S, Czosnek H, Aharoni A, et al. Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. The Plant Journal. 2018;95(1):5-16.
204. Cody WB, Scholthof HB. Plant virus vectors 3.0: transitioning into synthetic genomics. Annual review of Phytopathology. 2019;57:211-30.
205. Li C, Zhang K, Zeng X, Jackson S, Zhou Y, Hong Y. A cis element within flowering locus T mRNA determines its mobility and facilitates trafficking of heterologous viral RNA. J Virol. 2009;83(8):3540-8.
206. Jackson AO, Li Z. Developments in plant negative-strand RNA virus reverse genetics. Annual review of phytopathology. 2016;54:469-98.
207. Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, Klimyuk V, et al. Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors. Proceedings of the National Academy of Sciences. 2006;103(40):14701-6.
208. Blainey P, Krzywinski M, Altman N. Points of significance: replication. Nature methods. 2014;11(9):879.
209. Su YH, Zhang XS. The hormonal control of regeneration in plants. Current topics in developmental biology. 2014;108:35-69.
210. Ikeuchi M, Ogawa Y, Iwase A, Sugimoto K. Plant regeneration: cellular origins and molecular mechanisms. Development. 2016;143(9):1442-51.
211. Bhatia S, Bera T. Somatic embryogenesis and organogenesis. Modern applications of plant biotechnology in pharmaceutical sciences. 2015:209-30.
212. Sugimoto K, Gordon SP, Meyerowitz EM. Regeneration in plants and animals: dedifferentiation, transdifferentiation, or just differentiation? Trends in cell biology. 2011;21(4):212-8.
213. Skoog F, Miller C, editors. Chemical regulation of growth and organ formation in plant tissues cultured. Vitro Symp Soc Exp Biol; 1957.
214. Saad AI, Elshahed AM. Plant tissue culture media. Recent advances in plant in vitro culture. 2012:30-40.

215. Gerdes S, Lerma-Ortiz C, Frelin O, Seaver SM, Henry CS, de Crécy-Lagard V, et al. Plant B vitamin pathways and their compartmentation: a guide for the perplexed. Journal of experimental botany. 2012;63(15):5379-95.
216. Mao Y, Botella JR, Liu Y, Zhu J-K. Gene editing in plants: progress and challenges. National Science Review. 2019;6(3):421-37.
217. Svitashev S, Schwartz C, Lenderts B, Young JK, Mark Cigan A. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. Nat Commun. 2016;7:13274.
218. Mohamed GM, Amer AM, Osman NH, Sedikc MZ, Hussein MH. Effects of different gelling agents on the different stages of rice regeneration in two rice cultivars. Saudi Journal of Biological Sciences. 2021.
219. Bennett EP, Petersen BL, Johansen IE, Niu Y, Yang Z, Chamberlain CA, et al. INDEL detection, the 'Achilles heel'of precise genome editing: a survey of methods for accurate profiling of gene editing induced indels. Nucleic acids research. 2020;48(21):11958-81.
220. Konieczny A, Ausubel FM. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. The plant journal. 1993;4(2):403-10.
221. Matuszczak M, Spasibionek S, Gacek K, Bartkowiak-Broda I. Cleaved amplified polymorphic sequences (CAPS) marker for identification of two mutant alleles of the rapeseed BnaA. FAD2 gene. Molecular biology reports. 2020;47(10):7607-21.
222. Mashal RD, Koontz J, Sklar J. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. Nature genetics. 1995;9(2):177-83.
223. Zhu X, Xu Y, Yu S, Lu L, Ding M, Cheng J, et al. An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. Scientific reports. 2014;4(1):1-8.
224. Maher MF, Nasti RA, Vollbrecht M, Starker CG, Clark MD, Voytas DF. Plant gene editing through de novo induction of meristems. Nat Biotechnol. 2020;38(1):84-9.
225. Giovannini A, Laura M, Nesi B, Savona M, Cardi T. Genes and genome editing tools for breeding desirable phenotypes in ornamentals. Plant cell reports. 2021;40(3):461-78.
226. Charrier A, Vergne E, Dousset N, Richer A, Petiteau A, Chevreau E. Efficient targeted mutagenesis in apple and first time edition of pear using the CRISPR-Cas9 system. Frontiers in Plant Science. 2019;10:40.
227. Hong JK, Suh EJ, Park SR, Park J, Lee Y-H. Multiplex CRISPR/Cas9 Mutagenesis of BrVRN1 Delays Flowering Time in Chinese Cabbage (Brassica rapa L. ssp. pekinensis). Agriculture. 2021;11(12):1286.
228. Jeong SY, Ahn H, Ryu J, Oh Y, Sivanandhan G, Won K-H, et al. Generation of early-flowering Chinese cabbage (Brassica rapa spp. pekinensis) through CRISPR/Cas9-mediated genome editing. Plant Biotechnology Reports. 2019;13(5):491-9.
229. Soyk S, Müller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R, et al. Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. Nature Genetics. 2017;49(1):162-8.
230. Varkonyi-Gasic E, Wang T, Voogd C, Jeon S, Drummond RS, Gleave AP, et al. Mutagenesis of kiwifruit CENTRORADIALIS-like genes transforms a climbing woody perennial with long juvenility and axillary flowering into a compact plant with rapid terminal flowering. Plant Biotechnology Journal. 2019;17(5):869-80.

231. Waltz E. With a free pass, CRISPR-edited plants reach market in record time. Nature biotechnology. 2018;36(1):6-8.
232. Jiang L, Li D, Jin L, Ruan Y, Shen WH, Liu C. Histone lysine methyltransferases Bna SDG 8. A and Bna SDG 8. C are involved in the floral transition in Brassica napus. The Plant Journal. 2018;95(4):672-85.
233. Char SN, Wei J, Mu Q, Li X, Zhang ZJ, Yu J, et al. An Agrobacterium-delivered CRISPR/Cas9 system for targeted mutagenesis in sorghum. Plant Biotechnology Journal. 2020;18(2):319.
234. Cai Y, Chen L, Liu X, Guo C, Sun S, Wu C, et al. CRISPR/Cas9-mediated targeted mutagenesis of GmFT2a delays flowering time in soya bean. Plant biotechnology journal. 2018;16(1):176-85.
235. McCabe MS, Klaas M, Gonzalez-Rabade N, Poage M, Badillo-Corona JA, Zhou F, et al. Plastid transformation of high-biomass tobacco variety Maryland Mammoth for production of human immunodeficiency virus type 1 (HIV-1) p24 antigen. Plant biotechnology journal. 2008;6(9):914-29.
236. Laforest LC, Nadakuduti SS. Advances in delivery mechanisms of CRISPR gene-editing reagents in plants. Frontiers in Genome Editing. 2022;4.
237. Rampersad S, Tennant P. Replication and expression strategies of viruses. Viruses. 2018:55.
238. Bigelyte G, Young JK, Karvelis T, Budre K, Zedaveinyte R, Djukanovic V, et al. Miniature type VF CRISPR-Cas nucleases enable targeted DNA modification in cells. Nature communications. 2021;12(1):1-8.
239. Haider S, Faiq A, Khan MZ, Mansoor S, Amin I. Fully Transient CRISPR/Cas12f system in plants capable of broad-spectrum resistance against Begomovirus. bioRxiv. 2022.
240. Toda E, Koiso N, Takebayashi A, Ichikawa M, Kiba T, Osakabe K, et al. An efficient DNA- and selectable-marker-free genome-editing system using zygotes in rice. Nat Plants. 2019;5(4):363-8.
241. Hamada H, Liu Y, Nagira Y, Miki R, Taoka N, Imai R. Biolistic-delivery-based transient CRISPR/Cas9 expression enables in planta genome editing in wheat. Scientific reports. 2018;8(1):1-7.
242. Zhang Y, Massel K, Godwin ID, Gao C. Applications and potential of genome editing in crop improvement. Genome biology. 2018;19(1):1-11.
243. Molla KA, Sretenovic S, Bansal KC, Qi Y. Precise plant genome editing using base editors and prime editors. Nature Plants. 2021;7(9):1166-87.
244. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533(7603):420-4.
245. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. Nature. 2017;551(7681):464-71.
246. Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, Langner LM, et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. Nature biotechnology. 2021;39(1):41-6.
247. Zhao D, Li J, Li S, Xin X, Hu M, Price MA, et al. Glycosylase base editors enable C-to-A and C-to-G base changes. Nature biotechnology. 2021;39(1):35-40.
248. Berríos KN, Evitt NH, DeWeerd RA, Ren D, Luo M, Barka A, et al. Controllable genome editing with split-engineered base editors. Nature Chemical Biology. 2021;17(12):1262-70.

249. Yuan G, Lu H, Hassan MM, Liu Y, Li Y, Abraham PE, et al. An intein-mediated split–nCas9 system for base editing in plants. bioRxiv. 2021.
250. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149-57.
251. Union CoJotE. PRESS RELEASE No 111/18: Organisms obtained by mutagenesis are GMOs and are, in principle, subject to the obligations laid down by the GMO Directive. 2018.
252. Ledford H. CRISPR conundrum: strict European court ruling leaves food-testing labs without a plan. Nature. 2019;572(7767):15-6.
253. Commission E. EC Study on New Genomic Techniques. 2021.
254. Parliament U. Genome-Edited Food Crops: Westminster, London: The Parliamentary Office of Science and Technology; 2022 [Available from: https://researchbriefings.files.parliament.uk/documents/POST-PN-0663/POST-PN-0663.pdf.
255. Parliament U. Genetic Technology (Precision Breeding) Bill 2022 [Available from: https://bills.parliament.uk/bills/3167.
256. Nasti RA, Voytas DF. Attaining the promise of plant gene editing at scale. Proceedings of the National Academy of Sciences. 2021;118(22):e2004846117.
257. Shariati SA, Dominguez A, Xie S, Wernig M, Qi LS, Skotheim JM. Reversible disruption of specific transcription factor-DNA interactions using CRISPR/Cas9. Molecular cell. 2019;74(3):622-33. e4.
258. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154(2):442-51.
259. Pflueger C, Tan D, Swain T, Nguyen T, Pflueger J, Nefzger C, et al. A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. Genome research. 2018;28(8):1193-206.
260. Xing S, Chen K, Zhu H, Zhang R, Zhang H, Li B, et al. Fine-tuning sugar content in strawberry. Genome biology. 2020;21(1):1-14.
261. Erpen-Dalla Corte L, M. Mahmoud L, S. Moraes T, Mou Z, W. Grosser J, Dutt M. Development of improved fruit, vegetable, and ornamental crops using the CRISPR/Cas9 genome editing technique. Plants. 2019;8(12):601.