A VIRAL-BASED TOOLBOX FOR EFFICIENT GENE EDITING IN *NICOTIANA* SPECIES

A Dissertation

by

WILL BRYAN CODY

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Herman B. Scholthof
Committee Members,	Karen-Beth G. Scholthof
	John V. da Graça
	Kranthi K. Mandadi
Head of Department,	Leland S. Pierson III

August 2018

Major Subject: Plant Pathology

ABSTRACT

The CRISPR/Cas9 site-specific nuclease from a prokaryotic viral defense mechanism has been adapted into a gene editing tool for molecular biologists. Consisting of two deliverable parts, a Cas9 nuclease and sequence specific single guide RNA (sgRNA), CRISPR/Cas9 simplicity of design is its hallmark feature. While the system has brought gene editing to virtually every laboratory bench, there remain many aspects of the system not yet explored.

Here, I have combined the overexpression capabilities of a plant viral vector, along with CRISPR technology, for efficient targeted gene mutagenesis of a plant host. Through delivery of sgRNAs with a viral vector I have addressed two issues currently hindering gene editing in plant biology: low efficiency associated with current delivery infrastructure and the reliance on transgenic integration of gene editing parts. The results in *Nicotiana benthamiana* plants demonstrate the novel delivery of sgRNAs with substantial 5′ upstream and 3′ downstream RNA sequences yields efficient editing *in vivo*, which previously was considered to not produce catalytically active complexes. In addition to viral vector-based gene editing, I also rationally designed a subgenomic RNA where viral-based protein overexpression remains intact while simultaneously delivering sgRNAs.

Following these results, I looked to explore DNA double stranded breaks (DSB) capabilities of Cas9 with transcripts carrying 5′ and 3′ sgRNA "overhang" sequences. *In vitro* experiments demonstrated that 5′ sgRNA overhangs inactivate the catalytic activity of Cas9/sgRNA complexes, which hints that a native 5′ RNA processing mechanism in *N. benthamiana* must be present for viral delivery of biologically competent sgRNAs. Furthermore, 5′ sgRNA processed products were verified from viral delivered sgRNAs, *in planta*, and further

exploration revealed that processing events localized to the cytosol. A contemporary model is proposed for the processing events with specific host pathways implicated.

Lastly, I focused on the application of protein and sgRNA co-delivery using a viral vector for imaging of virus replication and movement while inactivating host genes. A survey revealed that the application of this system in diverse *Nicotiana* species is feasible. This is a potentially pivotal tool to further our understanding of plant-virus interactions.

DEDICATION

I dedicate this dissertation to my parents, Jay and Katie Cody. Without their support I could have not made it to where I am today. My parents set a foundation of the importance of not just having degrees signifying an education, but more importantly to be an educated person.

ACKNOWLEDGEMENTS

I would like to thank Herman, Karen, Kranthi, and John for serving on my committee and their incredible support over the past five and a half years.

I would especially like to thank Karen for her unwavering support of me since my days as an undergraduate student in the department. Karen, I still am not sure how I managed to gain your support as a distracted undergraduate student in 2011, but I am immensely thankful for you consistently looking out for my best interest and your belief in me.

I would like to thank Herman for taking a chance on me as a rotation student especially because I joined without a single day of molecular lab experience. This adventure into molecular virology could have been a massive failure, but if I was driven by nothing else it was to impress both Herman and Karen. As I move on in my career it would be disingenuous of me to say that it will not continue to be a motivating factor. Herman has been the best possible mentor I could have ever asked for. I hope one day to have a lab of my own and show the dedication to those students as he has shown to me.

I would like to thank Kranthi during his time as a postdoc in Karen's lab for being the best possible model for how scientists should conduct themselves. My efficiency in the lab, writing manuscripts, and the inquisitive nature it takes to be a scientist is a direct reflection of his time in the lab, and later as a member of my committee. Additionally, I would like to Jesse Pyle a previous lab member (and my good friend) for being my scientific muse during his time in the lab and continued support from a distance. I will never forget the late nights at my old house creating bizarre experiments on my makeshift chalk-board.

I am not sure what makes lab environments such a magical place for foraging friendships, but I am tremendously thankful for it. In addition to Kranthi and Jesse, I would also like to thank the other lab members over the years for creating such a fantastic work environment. In particular graduate students Maria Mendoza, April DeMell, and Kelvin Chiong. Thank you all for all of your help, support, and love for science.

Finally I would like to thank my family. Not only my parents (for whom this dissertation is dedicated), but also my loving fiancé Lauren. Her willingness and instance to drive 10 hours a weekend so we could see each other every week for a year and a half is perhaps the reason I had the strength to complete this degree. I would also like to thank Millie Cody for her sensitive nature, nightly cuddles, and somewhat unconditional love. Last but not least, but certainly the toughest, I would like to thank Wesley Cody for finding me in Weslaco when I needed a friend and for being “the most fantastic dog ever”.

CONTRIBUTORS AND FUNDING SOURCES

The work was supervised by a dissertation committee comprised of Dr. Herman B. Scholthof (Advisor), Dr. Karen-Beth G. Scholthof (Member), Dr. Kranthi K. Mandadi from the Department of Plant Pathology and Microbiology [Home department], and Dr. John V. da Graça from the Department of Horticultural Sciences [Outside member]. All work from this dissertation was performed, analyzed, and made to presentation formatting by Will B. Cody.

Graduate study was funded and supported by a first year graduate research assistantship provided by the Department of Plant Pathology and Microbiology, and a graduate research assistantship provided by Southern Garden Citrus-US Sugar obtained by Dr. T. Erik Mirkov. Graduate study funding and supplies for the last two years was supported by a USDA-NIFA Predoctoral Fellowship awarded to Will B. Cody. Additional research supplies were obtained through grants from USDA-NIFA-AFRI awarded to Dr. Herman B. Scholthof, and Genomics and Bioinformatics Texas A&M AgriLife Plant Water Use Seed Grant awarded to Will B. Cody, Dr. Herman B. Scholthof, and Dr. T. Erik Mirkov.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

PLANT VIRUS VECTORS 3.0:

VIROLOGISTS TRANSITION INTO THE SYNTHETIC BIOLOGY ERA

INTRODUCTION

Plant virus vectors 3.0: Virologists transition into the synthetic biology era

The uses of plant viruses as heterologous protein overexpression tools has been an area of interest in the community for the past three decades (Siegel 1985). Since the initial development of viral vectors, a large body of work has focused on the most effective viruses for heterologous protein production as well as the modification of specific viral vectors to enhance the ability to produce proteins. In either case, this area of viral vector technology development has been discussed in other review articles (Gleba et al. 2007; Scholthof et al. 1996) and will not be the focus of the present review. Rather, I will focus on three areas of viral vector technology development which could not have been predicted over two decades ago (Scholthof et al. 1996): 1) The vast diversity of viruses and their genomic components developed as heterologous expression tools, 2) viral vectors applications in functional genetics specifically on non-model (not *Arabidopsis thaliana*) plants, and 3) the recent uses of viral vectors for delivery of gene editing tools. It is the development of the tools we discuss here that has transitioned the utilization of viral vectors from a plant virology novelty instrument, to the lab bench of virtually every plant molecular biologist. This review will serve as a brief introduction to how I have implemented viral vector technology in the three subsequent research chapters.

Viral vectors 1.0: Use of plant viruses as protein expression tool

Viruses (Fiers et al. 1976), and more specifically plant viruses, were some of the first "organisms" genomes to be sequenced (Goelet et al. 1982; Franck et al. 1980). This was mostly due to the small size –most plant viruses are under 10 kb –of the viruses allowing for a reasonable, but certainly not trivial number of nucleotides to be deciphered using sequencing gels. Sequencing of DNA and RNA plant viruses ultimately led to the development of full length DNA (Gardner et al. 1981; Stanley et al. 1986) and cDNA (Dawson et al. 1986; Ahlquist et al. 1984), respectively, infectious clones. These painstaking methods of sequencing, and the resulting sequenced viral genomes, ultimately gave rise to the genomics era in biological sciences. The use of sequenced infectious clones made viral functional genetics not only attainable, but fairly straight forward.

Plant virology took on a new role following the discovery of *Agrobacterium* based transformation of plant nuclear genomes in the 1980s (Bevan 1984). While the utility of this tool seemed to have no bounds in the plant biology community, there was an overall lack of understanding on how to integrate the expression of protein coding genes. Considering plant viruses have been studied throughout the beginning of sequencing technologies through development of infectious clones and functional genetic studies, the basis for most genomic tools such as promoters, enhancers and terminators were known for plant viruses. Organisms with much larger genomes, such as plants, were far beyond the scope of the technology at the time to have an understanding of how native genetic parts functioned. Perhaps it was the limitation of plant genomic sequences as well as the realization by researchers at the time that viruses are uniquely qualified for the task of gene expression which led to their use in transformation technology (Restrepo et al. 1990). Ultimately these viral genetic parts (promoters, enhancers, and

terminators) were used in the early *Agrobacterium* plasmid vectors (Kay et al. 1987). Remnants of the viral-heavy influences in plant biology are still present today; for example, the 35S promoter from dsDNA *Cauliflower mosaic virus* (CaMV) as well as its transcriptional terminator. However, the application of viral parts to facilitate protein over-expression extends beyond the use of promoters and has grown to the employment of transcriptional enhancers, such as the 5′ and 3′ UTRs from *Tobacco etch virus* (TEV) and the 5′ leader sequence (omega) from *Tobacco mosaic virus* (TMV) (Gallie et al. 1987). Additionally, viral proteins have been used to supplement expression of proteins such as the RNA silencing suppressors HC-Pro from potyviruses (Anandalakshmi et al. 1998a) and *Tomato bushy stunt virus* P19 (Saxena et al. 2011). Interestingly enough, we are now so far removed from the initial development of tumor-inducing (Ti) plasmids for heterologous protein expression, it is not uncommon for newly trained plant biologists to not know their plant expression cassettes are mostly of viral origin.

With plant virus genetic parts already being used for protein expression, it was at the forefront of researchers' minds, and not an unreasonable leap, when virologists hypothesized that they could bypass plant transformation steps and manipulate infectious viral clones to transiently produce proteins (Siegel 1985). The development of viral vectors as overexpression tools was further aided by the use of *Agrobacterium* to launch viral expression (Grimsley et al. 1986). However, the magnitude of the development of viruses as tools, Agro-infection/infiltration using infectious clones, would take a few more years to come to fruition (Sean et al. 1992). Ultimately the development of Agro-launchable viral vectors to express heterologous proteins—the first iteration of viral vectors or viral vectors 1.0—has split plant virology labs, generally, into two aspects of research, fundamental and technology development.

Viral vectors 2.0: From protein production to functional genetic tools

While the use of viral vectors was gaining momentum outside of virology settings due their use as protein production tools, viruses were still somewhat of a "black box" for basic plant biologists. I propose two predominantly factors which have led to the wide use of viral vectors outside of virology labs: 1) the discovery of post-transcriptional RNA silencing (PTGS) (Lindbo et al. 1993); and, 2) the development of modern sequencing tools—initially Sanger sequencing and now high throughput sequencing. The use of both technologies led to the development of viral induced gene silencing (VIGS) screens (Ruiz et al. 1998; Kumagai et al. 1995). Due to the synergistic effect that both modern sequencing tools and the discovery of RNA silencing have, it is difficult to separate which led to the other, but ultimately these discoveries allowed plant virologists and plant biologists alike to produce new viral-based tools.

The initial development and use of VIGS for large functional genetic screens and the potential impact the technology would have in functional genetic studies excited the plant biology community, beginning nearly two decades ago (Burch-Smith et al. 2004; Baulcombe 1999). However, what could not be predicted at the time, before high throughput sequencing techniques, was the divergence of the VIGS screening tool from use in "model" plant species, such as *N. benthamiana* and *A. thaliana*, to agriculturally relevant or biologically interesting plants. An exponential increase in virus and plant sequences found in public databases has allowed those working on plant species, previously with only rudimentary cDNA libraries and a high amount of human effort, to now have available robust sequence data to empower their screening techniques. In short, through genomic and transcriptomic information we, as researchers, can actually "screen" for gene function instead of "fish" for phenotypes, as was typical of the initial cDNA VIGS assays. In any case, it is hard to imagine that nearly two

decades ago functional genetics using VIGS screening tools would be developed sufficiently to explore non-model plants such as woody species, which take a considerable amount of time and space to grow.

Citrus is one crop benefitting from the advances in both viral vector and high throughput sequencing. The development of an expressed sequence tag library (Forment et al. 2005) and more recently a reference genome (Xu et al. 2013) has allowed molecular biologists to work on a crop species that has rather simple genetics. This provided a near perfect opportunity to use viruses to understand host gene function, within a reasonable time frame, on a crop where breeding methodologies can be extremely time consuming. Within the past decade several *Citrus* VIGS vectors have been constructed and used to ascertain gene function in *Citrus* hosts (Aguero et al. 2012; Hajeri et al. 2014). Interestingly enough, it turns out the *Citrus tristeza virus* (CTV) based VIGS tool can also be used to down regulate genes in insects feeding on the hosts (Hajeri et al. 2014). The development of these tools in *Citrus* as well as other perennial crops are further reviewed elsewhere (Dawson et al. 2015; Dolja and Koonin 2013).

Viral vectors 3.0: Viruses as the biological "software" of the future

Unlike the previous two sections, here, I will focus on the current and future development on what I will be referring to as viral vectors 3.0. The culmination of using viruses to express foreign genes, viral vectors 1.0, and knocking down host gene expression levels, viral vectors 2.0, into a single deliverable is the primary goal of viral vectors 3.0.

The framework for the development of this technology will be discussed in detail in the ensuing chapters. Ultimately, the drive to create such tools is fundamental in origin. The initial goal was to create a viral vector system that can be used for viral tracking in real-time, as was

proposed and demonstrated over two decades ago (Baulcombe et al. 1995), while targeting specific host factors for knocking down expression. In doing so, virus movement and replication can be visualized in the absence of potentially key host factors in a single screen. After the development of the tool, the potential technological applications became increasingly apparent.

Viral vectors should not be categorized as technology or basic science, rather, it is the application of the tools which defines which category the research belongs to. In each iteration of the viral vector toolkit there have been examples of both applications of technology and uses in fundamental research. Perhaps our energy would be better suited by not focusing on these trivial categorizations of research and instead focus our fundamental understanding of the tool and open our minds to all of its potential applications.

From theoretical to practical applications

The intent of the above short perspective from previous research (historical) and future goals is to provide a framework for a review paper which ultimately in its remaining parts will essentially set the stage, summarize, and place in context, the finding of Chapters 2-4. The overall hypothesis of my dissertation research was that plant virus vectors (3.0) can be used as delivery tools for CRISPR/Cas9 mediated gene editing (**Figure 1.1**). To summarize the findings, I have established a proof-of-concept for a new use of virus vectors for modern functional genomics (3.0) by using TMV to deliver single guide RNAs (sgRNAs) that are used to program Cas9 for gene editing. New findings of particular interest include determining that TMV can be used to co-express a protein (as in 1.0) as well as a sgRNA (creating the framework for 3.0), that, in contrast to expectations, the 5′ and 3′ overhangs on the sgRNA do not interfere with editing.

The final findings demonstrate that thee viral-based delivery platform can be used in multiple *Nicotiana* species.

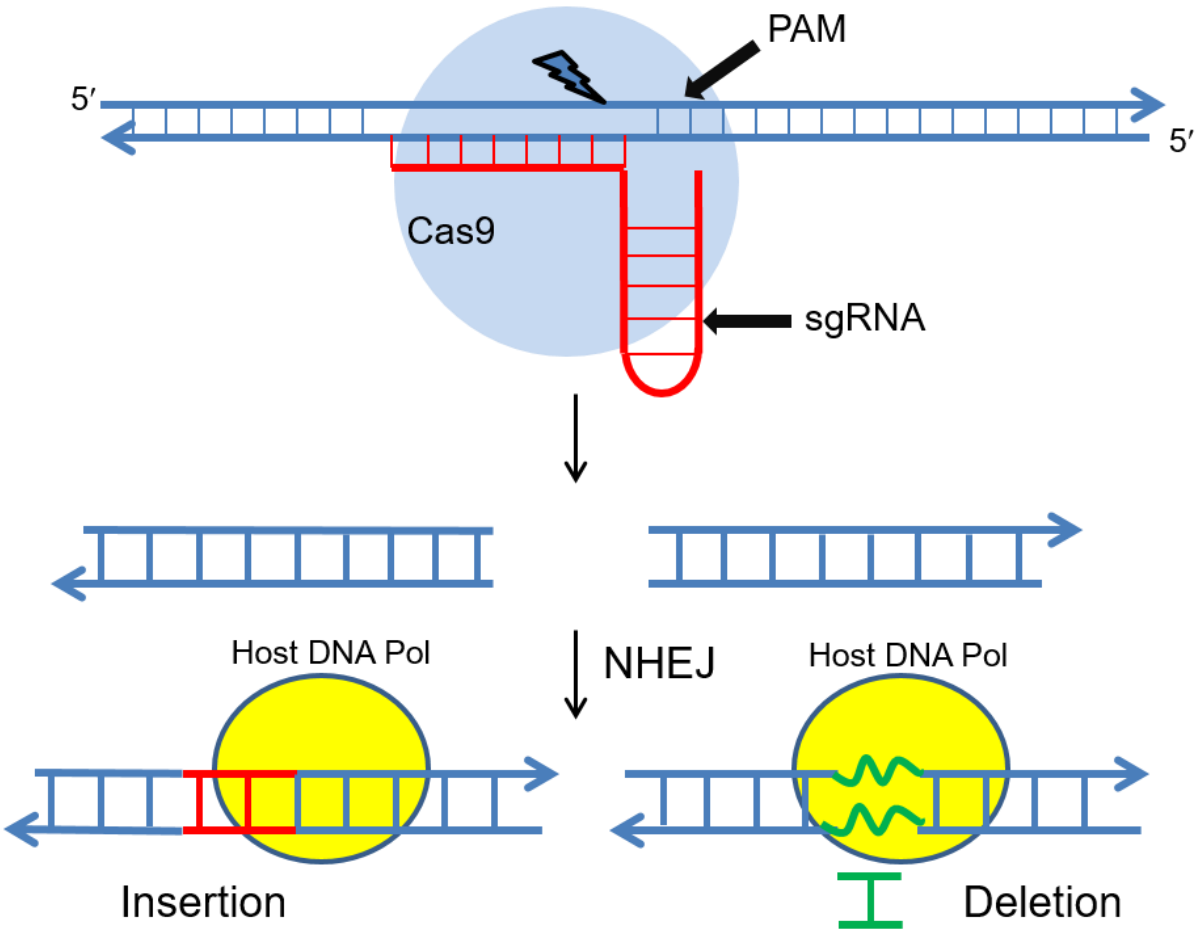


Figure 1.1 Cas9-sgRNA complex binding of complimentary genomic DNA and the hosts mode of repair of DNA double stranded breaks by nonhomologous end joining (NHEJ). The Cas9-sgRNA targets genomic DNA by complementary base pairing of DNA and sgRNA sequences. Cas9 forms a complex with genomic DNA and a sgRNA, which can then induce DNA double stranded breaks. The protospacer adjacent motif (PAM), usually identified as NGG (N representing any nucleotide), is 3′ proximal to the sgRNA-DNA R-loop based structure. The sgRNA molecule is represented by the red lines on the top panel. The bottom panel represents NHEJ-based nucleotide insertions or deletions at double stranded break points. Endogenous host DNA polymerase inserts and deletes nucleotides red and green lines, respectively, on the diagrams on the bottom. DNA is presented with blue lines with arrows pointing in the 3′ direction.

CHAPTER II

MULTIPLEXED GENE EDITING AND PROTEIN OVER-EXPRESSION USING A *TOBACCO MOSAIC VIRUS* VIRAL VECTOR[*]

INTRODUCTION

Gene editing tools allow for the precise targeting of DNA for purposes of changing nucleic acid sequences. This is performed by a variety of systems relying on DNA binding proteins fused to nucleases (Porteus and Carroll 2005; Cermak et al. 2011), ribonucleic acids (Dong et al. 2006) and protein-nucleic acid complexes (Cong et al. 2013; Jinek et al. 2012; Mali et al. 2013). Recently, the *Streptococcus pyogenes* CRISPR/Cas9 system has been adapted as a programmable DNA targeting RNA-guided nuclease complex for gene editing (Doudna and Charpentier 2014). The *S. pyogenes* based system has been widely adopted primarily due to the simplicity of its parts, consisting of a sequence-specific synthetic single guide RNA (sgRNA) and the sgRNA programmable Cas9 nuclease (Mali et al. 2013). CRISPR gene editing technology creates DNA double stranded breaks (DSBs) which activate the native host DNA repair mechanisms of non-homologous end joining (NHEJ) or homologous dependent repair (HDR). NHEJ repairs DSBs by inserting or deleting (indels) nucleic acids to restore the integrity of the host dsDNA. NHEJ causes localized DNA disruption, due to indel formation during the repair process, and has been used for sequence-specific disruption of downstream gene products, such as proteins and long non-coding RNAs (Schiml et al. 2014; Ran et al. 2013).

---

[*]**Cody WB, Scholthof HB, Mirkov TE** (2017) Multiplexed gene editing and protein overexpression using a *Tobacco mosaic virus* viral vector. Plant physiol. 175 (**1**):23-35.
www.plantphysiol.org, Copyright American Society of Plant Biologists

Current CRISPR-based editing processes in plants have focused on delivery of the Cas9 nuclease and sgRNAs by transformation technologies or transient delivery to protoplasts (Cong et al. 2013; Nekrasov et al. 2013). Transient protoplast editing, while relatively efficient, requires the regeneration of edited protoplast cells to produce plant lines. Alternatively, crop plants carrying T-DNA with coding regions for Cas9 and sgRNAs, are predisposed to regulation or require traditional breeding techniques to remove unwanted DNA inserts prior to commercialization (National Academies of Sciences 2016). For these reasons, an efficient transient method for generating stable mutations in tissue is desirable for the development of edited plant lines, and as a rapid screening tool.

Viral vectors are used in biotechnology for their ability to replicate and produce recombinant products in a wide range of hosts (Scholthof et al. 2002; Scholthof et al. 1996). The use of viral vectors as delivery devices for gene editing tools offers the potential of producing a high incidence of edited cells without the incorporation of recombinant DNA (Scholthof et al. 1996). DNA (Baltes et al. 2014; Yin et al. 2015) and RNA (Ali et al. 2015) viral based replicons have been tested as vectors for delivery of CRISPR components to create gene knockouts and gene insertions. To date, the efficiency of transient gene editing by viral vectors has been relatively low and delivery methods demonstrating a knockout phenotype have relied on transgenic technology in some capacity (Yin et al. 2015; Ali et al. 2015). CRISPR-based transient screens could be used for functional genetic studies, much like viral induced gene silencing (VIGS) screens (Liu et al. 2002). For example, a *Nicotiana benthamiana* GFP-expressing transgenic line (16c) (Ruiz et al. 1998), has been used extensively as proof-of-concept for RNA silencing studies, due to the phenotype associated with a knock down of GFP expression (Voinnet et al. 2000; Anandalakshmi et al. 1998b; Anandalakshmi et al. 2000), and

was selected in our studies for its potential to establish a proof-of-principle for using viral delivery of CRISPR components for editing efficiency tests.

TRBO, is a coat protein (CP) deletion mutant of the TMV U1 strain, that was initially developed as an agroinfiltration-based delivery tool for expressing recombinant protein coding sequences in host cells (Lindbo 2007). The CP deletion prevents the virus from systemically moving throughout infected plants, while still allowing localized cell-to-cell movement through the movement protein (MP). TRBO has been exploited in biotechnology for its ability to produce large amounts of a protein of interest in hosts such as *N. benthamiana,* while consolidating the infection to infiltrated leaves (Lindbo 2007). Recombinant protein production within infiltrated leaves is due to high quantities of transcript produced through the TRBO CP subgenomic RNA promoter. The transient *Agrobacterium* based delivery of CRISPR components, in current delivery systems that utilize constitutive promoters, is limited by the relatively low production of sgRNAs which lack 5′ and 3′ nucleotide overhangs (Cong et al. 2013; Nekrasov et al. 2013). We speculated that the potentially high output of sgRNAs by TRBO could provide a very effective editing platform to boost efficiency when used in combination with Cas9.

In this study we show that TRBO is a suitable vector for transient delivery of high concentrations of sgRNAs in *N. benthamiana* for Cas9 programming through the quantification of gene-specific indel percentages. We hypothesized that through targeting the GFP coding region in transgenic *N. benthamiana* 16c plants that there would be a phenotypic response along with an increase in indel percentage. We designed a sgRNA delivery platform involving the incorporation of RNA catalytic ribozymes to be transcribed from TRBO (Gao and Zhao 2014). After optimizing deployment of sgRNAs, which surprisingly did not necessitate the use of ribozymes, NHEJ-derived indel values of nearly 70% in the GFP coding region were achieved in

Cas9 co-infiltrated tissue. These results, from *in planta* experiments, contradict what we found using *in vitro* Cas9 cleavage assays, in which Cas9-based DNA DSBs did not occur when predicted TRBO-encoded subgenomic sgRNAs did not have 5′ and 3′ sgRNA overhang removal capabilities, suggesting that an unknown processing event is occurring in plants. Further *in planta* experimentation showed the majority of mutations occurred during 2-3 days post infiltration (dpi) and the occurrence of *mgfp5* indels was reliant on TRBO replication. Furthermore, these plants showed a reduction in GFP expression as well as a noticeable chlorophyll-associated red auto-fluorescent phenotype, confirming GFP protein expression reduction. We also successfully demonstrated the ability to target two paralogues with a single sgRNA, as well as the co-delivery of two adjoining sgRNAs using a single TRBO construct. Additionally, we report the previously unexplored ability of TRBO to deliver a single subgenomic RNA encoding both a biologically functional protein along with a sgRNA for editing a host gene *in planta*.

RESULTS

Development of a genomic target and sgRNA deployment optimization

Constitutive GFP-expressing transgenic *N. benthamiana* 16c plants were selected for our proof-of-concept study for the TRBO-sgRNA delivery system due to the easily detectable red auto-fluorescent phenotype exhibited upon disruption of GFP production (Ruiz et al. 1998). In addition to the phenotype observed with GFP inactivation, GFP protein expression levels can be verified with assays using commercially available antibodies. The *mgfp5* sequence was used to design a sgRNA around a *Bsg*I restriction site for downstream detection of indels using a restriction enzyme resistance assay (Cong et al. 2013; Nekrasov et al. 2013). To evaluate

whether RNA replicons can deliver high quantities of sgRNAs *in planta*, we used a TMV-based coat protein (CP) deletion mutant, TRBO, which replicates to high concentrations in infiltrated *N. benthamiana* (Lindbo 2007; Fu et al. 2014). Due to the specificity of the sgRNA in CRISPR based gene editing systems, we hypothesized that nucleotide overhangs 5′ proximal to the spacer sequence (Dahlman et al. 2015; Fu et al. 2014) or 3′ proximal to the scaffold sgRNA (Konermann et al. 2015; Zalatan et al. 2015) would lower the efficiency of the system, and the subsequent occurrence of DSBs *in planta*. The 5′ and 3′ sgRNA overhangs were expected based on what is known about the TRBO CP subgenomic RNA, which contains both a 5′ 63-nucleotide sequence as well as the TMV 3′ UTR (~200 nt) downstream of the inserted sequence (Dawson et al. 1989; Buck 1999). To circumvent possible compatibility issues of Cas9-sgRNA-DNA binding, and subsequent DSBs due to 5′ and 3′ sgRNA nucleotide overhangs, the construction of three independent sgRNA delivery devices were designed using catalytic RNAs (ribozymes) (**Figure 2.1A**).

The ribozymes were designed and placed to yield sgRNAs that either lacked both 5′ and 3′ overhangs, only removed 3′ overhangs, or that resulted in sgRNA without any flanking ribozymes and thus predicted to yield a substrate with substantial 5′ and 3′ overhangs (Figure 2.1A). With each of these sgRNA delivery constructs it could be readily tested if overhangs affect the generation of Cas9/sgRNA DSBs *in planta.* The first construct, aimed at 5′ and 3′ overhang removal, has the sgRNA sequence flanked by a 5′ hammerhead (HH) ribozyme and a 3′ hepatitis delta virus (HDV) ribozyme (RGR), as described previously (Gao and Zhao 2014). The second construct contained only a 3′ HDV ribozyme located 3′ proximal to the sgRNA (gHDV) sequence. The third construct consisted of a sgRNA without the presence of catalytic HH and HDV units (gGFP) predicted to result in sgRNA that carries both 5′ and 3′ nucleotide

overhangs. Each of the three independent sgRNA delivery combinations (RGR, gHDV, and gGFP) were subcloned downstream of the TRBO CP subgenomic promoter for expression (Figure 2.1A). The activity of the ribozymes was verified *in vitro* (**A-1**).

TRBO-mediated delivery of biologically active sgRNAs

To test the biological activity of TRBO delivered sgRNAs *in planta*, the Cas9 nuclease must be expressed along with the TRBO-sgRNA delivery constructs. The presence of indels can then be assayed and quantified to estimate the efficiency of each sgRNA delivery construct. For Cas9/sgRNA co-delivery, expression of the Cas9 protein within *N. benthamiana* using a minimal binary plasmid was considered to be optimal for the highest possible transient expression. A Cas9 cassette was constructed using a previously synthesized human codon-optimized Cas9 consisting of an N terminus 3X FLAG tag and N and C terminus nuclear localization signal (NLS) (Addgene plasmid: 42230) (Cong et al. 2013). Optimal Cas9 expression in plants was achieved using a CaMV 35S double promoter and translation enhancers including the 5′ and 3′ UTR regions of *Tobacco etch virus* (TEV) (**Figure 2.1B**). Protein expression was confirmed by western blot assays, following agroinfiltration of *N. benthamiana*. Our newly constructed Cas9 delivery cassette (pHcoCas9) had much higher rates of protein expression than a previously developed plant Cas9 expression vector (pFGC-pcoCas9, Addgene plasmid: 52256) (Li et al. 2013), which did not produce detectable Cas9 protein under our experimental conditions (**Figure 2.1C**).

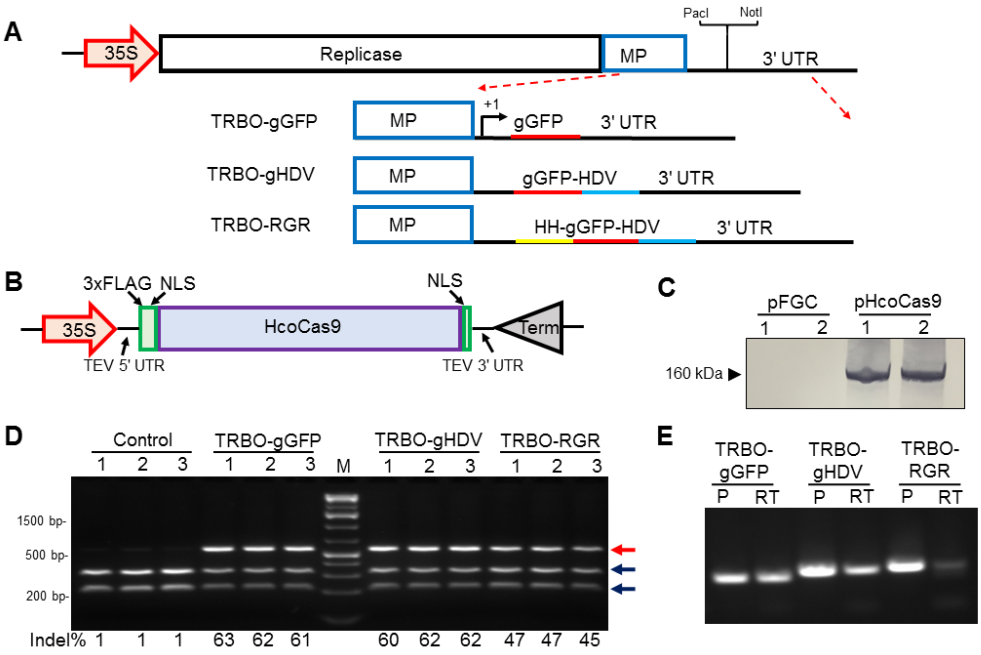


Figure 2.1. TRBO-sgRNA delivery *in planta*. A) Schematic of TRBO-sgRNA constructs. The predicted subgenomic RNA transcriptional start site is indicated by "+1". Three TRBO-sgRNA constructs were made to optimize sgRNA delivery: TRBO-gGFP consisting of the guide GFP (gGFP) only (red), TRBO-gHDV consisting of gGFP with a 3′ proximal hepatitis delta virus (HDV) ribozyme (blue), and TRBO-RGR consisting of gGFP with a 5′ hammerhead ribozyme (yellow) and a 3′ HDV ribozyme. **B)** The pHcoCas9 construct contains an N-terminal triple FLAG-tag (3xFLAG), nuclear localization signal (NLS), and a human codon-optimized Cas9 nuclease followed by a C-terminal NLS. Transcription is initiated by a 35S promoter and terminated by a 35S terminator. The pHcoCas9 transcripts contain both a 5′ TEV UTR and 3′ TEV UTR for increased translation efficiency. **C)** Immunoblot assays from *N. benthamiana* leaf sampled at 3 dpi were taken from two replicates of two different Cas9 expression plasmids. Anti-Flag probes were used for Cas9 protein expression detection of a previously constructed vector (pFGC) and pHcoCas9. **D)** Three technical replicates of *Bsg*I restriction enzyme resistance assay performed on mgfp5 amplicons from genomic DNA samples taken 12 dpi. Infiltration treatments are as follows: pHcoCas9 only (Control), pHcoCas9/TRBO-gGFP (TRBO-gGFP), pHcoCas9/TRBO-gHDV (TRBO-gHDV), and pHcoCas9/TRBO-RGR (TRBO-RGR) co-delivery. Numbers under each lane indicate predicted indel percentages quantified using ImageJ. **E)** PCR and RT-PCR products, respectively amplified from TRBO plasmid for each sgRNA delivery construct (P) or from 14 dpi co-infiltrated pHcoCas9/TRBO-sgRNA 16c *N. benthamiana* plants cDNA (RT) for detection of construct expression and retention *in planta*. Each plasmid amplicon (P) serves as a construct specific molecular weight control for the coinciding RT-PCR products.

N. benthamiana 16c plants were co-infiltrated with *Agrobacterium* cultures harboring the pHcoCas9 expression vector and either gGFP, gHDV, or RGR containing TRBO constructs. Host genomic DNA was sampled at 12 days post infiltration (dpi) and analyzed for indels. (Nekrasov et al. 2013; Cong et al. 2013). PCR amplicons of *mgfp5* from each TRBO-sgRNA delivery system were incubated with *Bsg*I restriction enzyme. Restiction enzyme resistant PCR products were present for each of the TRBO-sgRNA delivery treatments, indicating indel accumulation *in planta*, while control pHcoCas9-only treatments were readily digested by *Bsg*I (**Figure 2.1D**). *Bsg*I resistant DNA was cloned and sequenced to confirm the presence of indels at the predicted Cas9 DNA nucleation sites (**A-2**). Image analysis software (ImageJ) was used to quantify the *Bsg*I resistant amplicons, or the predicted percentage of mutated DNA, compared to the digested "wild-type" *mgfp5* DNA. Upon gel quantification analysis, both gGFP and gHDV showed greater than 60% mutated *mgfp5* DNA and RGR had a mutation rate of 46% (Figure 2.1D).

We anticipated that the lower indel efficiency of the RGR construct, as compared to gGFP and gHDV, was due to a decreased fitness level of the RNA replicon. To test the stability of the TRBO-sgRNA delivery constructs *in planta*, RNA was isolated from Cas9/TRBO-sgRNA co-infiltrated tissue at 14 dpi from plants previously confirmed to contain indels by *Bsg*I restriction enzyme resistance assays (**Figure 2.1D**). RNA extractions from each treatment were subject to TRBO-specific reverse transcription (RT) reactions using equal quantities of total RNA. RT-PCR analysis was used to confirm the integrity of the expected sgRNA constructs followed by cloning and sequencing of each amplicon (**Figure 2.1E**). Both gGFP and gHDV delivery constructs were confirmed to be stable *in planta* through sequencing analysis, but the RGR construct contained a point mutation in the HH ribozyme (**A-3**). Additionally, when

TRBO-RGR was delivered without Cas9 a chimera of RT-PCR products accumulated in *N. benthamiana* 16c plants (**A-3**). The sequencing of one of the most prevalent amplicons from the RGR RT-PCR products indicated a complete deletion of the sgRNA from the TRBO genome (**A-3**). Following these results, we concluded that ribozymes, specifically the 5′ HH ribozyme, had a negative effect on the replication of TRBO. Importantly, the TRBO-sgRNA delivery construct without ribozymes was both sufficient and efficient for delivery of biologically active sgRNAs (**Figure 2.1D**). In summary, the results show a novel feature that *in planta,* Cas9 tolerates the subgenomic RNA-mediated delivery of GFP-sgRNA carrying both 5′ and 3′ overhangs

5′ and 3′ sgRNA overhangs negate Cas9 DNA nuclease activity *in vitro*

Since the *in planta* results using TRBO-sgRNA delivery containing sgRNA nucleotide overhangs was unexpected, we anticipated that these results might have larger implications in Cas9/sgRNA biology in *in vivo* systems. While it has been largely assumed that sgRNA overhangs prevent Cas9 catalytic activity, the proper experiments have not been performed *in vitro* to confirm this assumption (Jinek et al. 2012). However, there has been considerable evidence that insinuates that sgRNA overhangs within the delivery construct do not impede DSBs in both bacteria (Karvelis et al. 2013), human cell lines (Cong et al. 2013), and also has been hinted at in plants (Jia and Wang 2014).

To understand if this phenomenon is fundamental to Cas9/sgRNA or if it is unique interaction that only occurs *in vivo,* we aimed at using our predicted subgenomic RNAs through the TRBO-sgRNA delivery methods previously tested *in planta* (**Figure 2.1D**) and provide them along with Cas9 *in vitro* to see if they create catalytically active complexes outside of cells. Subgenomic transcripts were created by designing a forward primer with a 5′ T7 promoter

sequence at the subgenomic transcriptional start site and a reverse primer located in the 3′ UTR. PCR amplification of TRBO-gGFP, TRBO-gHDV, and TRBO-RGR constructs created the predicted subgenomic RNA templates sub-gGFP, sub-gHDV, and sub-RGR, respectively. Each of the templates was used for T7 RNA synthesis. In addition to the subgenomic sgRNAs, a template encompassing only the 100 nt gGFP region was created (gGFP) as a positive control (**Figure 2.2A**).

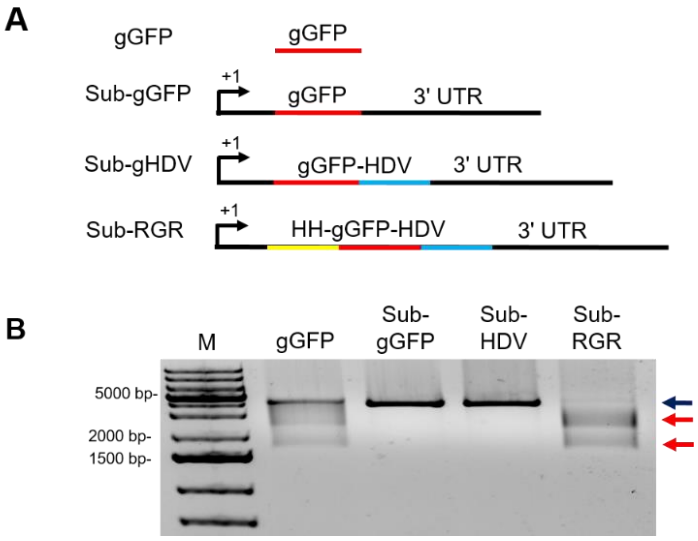


Figure 2.2. *In vitro* Cas9 cleavage assay using a *mgfp5* DNA template. A) Depiction of sgRNAs synthesized for loading into Cas9 *in vitro* Cas9 cleavage assays. The gGFP RNA carries no Cas9 overhangs and serves as the positive control. Predicted subgenomic RNAs (sub) were constructed for each of the construct used for the *in vivo* assay. TRBO-gGFP (sub-gGFP), TRBO-gHDV (sub-gHDV), and TRBO-RGR (sub-RGR) all contain the 63 nt subgenomic promoter sequence and the downstream 3′ UTR from the TRBO vector. **B)** *mgfp5* template was PCR amplified from 16c genomic DNA, cloned into a TA plasmid, and used as a template for Cas9/gGFP cleavage. The plasmid was linearized with *Sca*I restriction enzyme when co-incubated with Cas9 and either gGFP, sub-gGFP, sub-gHDV, or sub-RGR (RNAs depicted in **A**). Red arrows indicate expected DNA cleavage product sizes while the blue arrow represents uncut *mgfp5* template.

A DNA template was created by amplifying the *mgfp5* gene, containing the sgRNA target, from 16c genomic DNA and cloned into a TA vector. Purified Cas9 was incubated with each synthesized sgRNA at room temperature to create Cas9/sgRNA complexes. Purified

plasmid was then co-incubated with each of the Cas9/sgRNA treatments along with *Sca*I restriction enzyme which was used to linearize the plasmid. A positive Cas9/sgRNA reaction can be visualized through gel electrophoresis and the presence of digested linearized plasmid, yielding three DNA bands (**Figure 2.2B**). Following these assays it was confirmed that the subgenomic sgRNAs retaining the 5′ overhangs (sub-gGFP and sub-gHDV) did not yield DNA cleavage while sgRNAs from which 5′ overhangs were removed (gGFP and sub-RGR) did cleave the DNA target (**Figure 2.2B**). These *in vitro* results, along with the above results *in planta*, demonstrate that some unique interaction must be occurring to allow for TRBO-gGFP and TRBO-gHDV to create DSBs *in vivo* and that a comparable interaction and concomitant activity is lacking during the *in vitro* experiments.

Targeted introduction of indels reduces GFP expression

Following confirmation of indel disruption for the TRBO-delivered gGFP, gHDV, and RGR constructs *in planta* (**Figure 2.1D**), we further tested if GFP protein production would be significantly hindered by the development of indels. Half-leaf assays were conducted with half of the leaf being infiltrated with one of the three TRBO-sgRNA delivery methods alone and the other half being subjected to co-infiltrations of a TRBO-sgRNA treatment and pHcoCas9. Protein samples taken 14 dpi were analyzed using western blotting with a GFP specific antibody. GFP expression showed large reductions in GFP protein expression in pHcoCas9 co-infiltrated tissue with both TRBO-gGFP and TRBO-gHDV (Figure 3), confirming our observations for both TRBO-gGFP and TRBO-gHDV restriction enzyme resistance assay. Since TRBO-gGFP was performing optimally and because it allows for simplicity of design compared to constructs carrying ribozymes, we focused on this construct for delivery from here on.

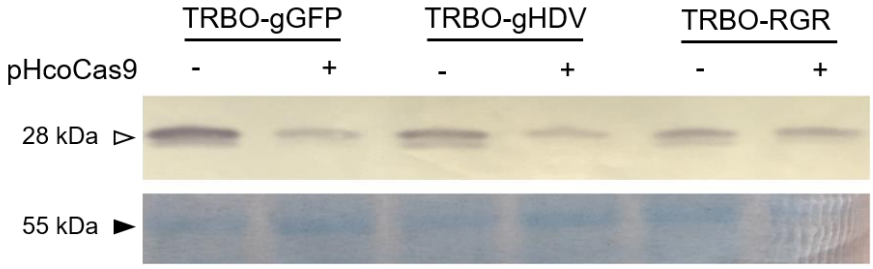


Figure 2.3. 16c *N. benthamiana* GFP protein expression in leaf tissue treated with TRBO-sgRNA constructs with and without Cas9. Anti-GFP HRP western blot for TRBO-sgRNA delivery alone (-) as well as pHcoCas9/TRBO-sgRNA co-delivery (+). The top panel is a western blot used to detect GFP (open arrow). The bottom panel is a Coomassie stain of RuBisCO (filled arrow) used as a protein loading control for each sample.

Rapid TRBO-gGFP derived indels and the corresponding loss-of-function phenotype

To determine if mutations within the *mgfp5* coding region of 16c plants are a direct consequence of viral replication or basal levels of expression from the 35S promoter used to launch the initial TRBO transcripts, a TRBO-gGFP replicase mutant (RM) was constructed by deleting a 1,419 bp region within the replicase coding sequence (**Figure 2.4A**). It was previously reported that TRBO-based recombinant protein expression was detected after just 3 dpi and would continue for several days thereafter (Lindbo 2007). Accordingly, 16c plants were co-infiltrated with pHcoCas9 and either TRBO-gGFP or RM-gGFP. Genomic DNA was isolated, from these treatments, daily over the course of 10 days to have a better understanding of quantity and timing of indel development. *Bsg*I restriction enzyme resistance assays were used to both verify and quantify the percentage of indels in infiltrated tissue (**A-4**). As expected, the negative

control pHcoCas9/RM-gGFP samples showed an indel detection of less than 3%, which is low enough to be accounted for as background from non-indel containing DNA that was undigested, or due to the 35S promoter driving the expression of the replication-negative construct. Cas9/TRBO-gGFP showed no measurable indel buildup until 3 dpi where percentages increased from less than 2% to 48%, and increased to ~70% at 6 to 7 dpi (**Figure 2.4B**). Thus, the time course analysis confirms that the pHcoCas9/TRBO-sgRNA transient knockout screening method is virus replication-dependent and TRBO subgenomic RNAs provide high levels of sgRNA expression for efficient and rapid gene editing.

To visualize how high levels of indel percentages result in a loss-of-function phenotype, at 10 dpi leaves from both RM-gGFP and TRBO-gGFP infiltrations were viewed using a stereomicroscope with a GFP filter (Olympus SZX7). Although GFP expression was present, a clear knockdown of GFP fluorescence was observed in pHcoCas9/TRBO-gGFP co-infiltrated tissue compared to pHcoCas9/RM-gGFP plants (**Figure 2.4 C** and **D**). It should be noted that in absence of Cas9 delivery, GFP expression was still present (**Figure 2.3**), indicating the reliance for both a replication competent TRBO along with co-delivery of Cas9 protein, as expected. Therefore, the clear difference in phenotype based upon a strong constitutively expressed plant chromosomal localized gene demonstrates the potential value of Cas9/TRBO-sgRNA as a transient knockout screening tool.

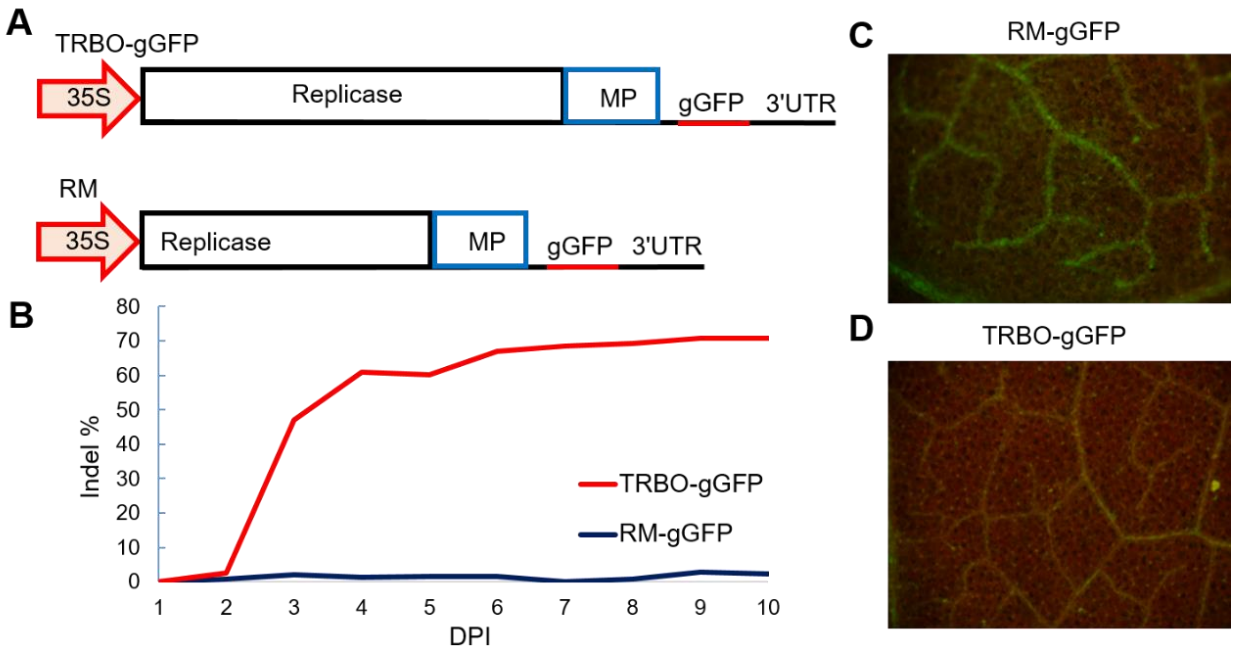


Figure 2.4. TRBO-gGFP delivery indel time course and analysis of GFP protein expression in *N. benthamiana* 16C plants. A) TRBO-based constructs used in time course and imaging study. TRBO-gGFP is replication competent and carries gGFP. RM-gGFP is a TRBO-gGFP mutant that has a 1,419 bp deletion in the replicase protein coding sequence that does not support vector replication or transcription of subgenomic RNA. **B)** Indel percentages over 10 days calculated using restriction resistance assay for TRBO-gGFP and a TRBO replicase mutant carrying gGFP. **C** and **D** are 16c N. benthamiana leaf images using a dissecting microscope with GFP filter. 10 dpi leaf samples were treated with under the following conditions; **C)** RM-gGFP co-infiltrated with pHcoCas9 and **D)** TRBO-gGFP co-infiltrated with pHcoCas9.

TRBO-sgRNA targeting of native genes

Following our success with TRBO-gGFP to induce high percentages of indels in the *mgfp5* transgene, we next tested the ability to target endogenous genes within *N. benthamina*. *ARGONAUTE1* (*NbAGO1*) was selected for gene targeting because of its central role in endogenous RNA silencing in *N. benthamiana*, by presumably serving as the catalytic portion of the RNA-induced silencing complex (RISC). VIGS-based knockdown screens show severe systemic phenotypic responses when *NbAGO1* paralogs are knocked-down using VIGS (Jones et al. 2006). It has previously been well documented that there are two *NbAGO1* paralogs within *N. benthamiana* with 97.61% shared transcript identity (Jones et al. 2006; Gursinsky et al. 2015). The two paralogs have been termed *NbAGO1-H* (KR942296) and *NbAGO1-L* (KR942297) (Gursinsky et al. 2015). While VIGS assays allow for systemic responses, here we aimed to understand if localized gene editing of a native gene could be observed through *NbAGO1* targeting in mature leaf tissue. To test this with the Cas9/TRBO-sgRNA co-infiltration system, we designed a sgRNA spacer using a complementary region of both *NbAGO1-H* and *NbAGO1-L* within the protein coding region for the PAZ domain (**A-5**). The sgRNA target for *NbAGO1* (gAGO1) was cloned into TRBO expression vector, using the same method as described for the TRBO-gGFP, to yield TRBO-gAGO1.

Using genomic DNA sequences, specific PCR amplification of *NbAGO1-H* and *NbAGO1-L* were designed by aligning primers within allele specific intron regions that vary significantly between the paralogs. Half leaf assays consisting of pHcoCas9/TRBO-gAGO1 and TRBO-gAGO1 only were infiltrated into *N. benthamiana* 16c plants. *NbAGO1-H* and *NbAGO1-L* alleles were amplified from genomic DNA at 7 dpi and amplicons were subject to the surveyor nuclease assay, where digested products contain indels. Both *NbAGO1-H* and *NbAGO1-L*

showed significant indel percentages in both paralogs (**Figure 2.5B**), albeit much lower percentages than observed when previously targeting *mgfp5* (**Figure 2.1D**). An observable phenotype was yet to be associated with the mutated *NbAGO1* alleles at 7 dpi, suggesting that an NbAGO1 knockout phenotype might take longer to establish or might not develop in mature leaves. Nevertheless, our virus-mediated system successfully achieved editing of a native gene in mature leaves with relatively high efficiency.

TRBO can efficiently deliver multiple sgRNAs on a single subgenomic RNA

After successfully editing endogenous genes and a transgene using the TRBO delivery system, we further explored the system's utility through the delivery of multiple sgRNAs from one construct (multiplex). The ability to edit multiple genes at once, upon transient delivery, could greatly aid functional genetic studies and regenerating plants with multiple gene knockouts. Due to the ability of the Cas9/TRBO-sgRNA system in *N. benthamiana* 16c plants to tolerate sgRNAs with 5′ and 3′ nucleotide overhangs while still producing indels, we hypothesized that multiple sgRNAs could be co-delivered using a single TRBO construct. We predicted that both the gGFP and gAGO1 sgRNA constructs could be positioned "side-by-side" under control of the same CP subgenomic promoter, without a linker sequence separating the sgRNAs, and efficiently induce Cas9-mediated DSBs. Additionally, we also tested if the sgRNA orientation proximity to the 5′ or 3′ end of the subgenomic RNA effected the efficiency of indel production.

The TRBO-gGFP vector was used as a template for the addition of gAGO1 directly 3′ to the sgRNA scaffolding of gGFP, creating TRBO-gGFP-gAGO1 (**Figure 2.5A**). Half leaf assays were carried out with co-infiltrations of TRBO-gGFP-gAGO1 either with, or without, Cas9 in

16c *N. benthamiana* plants. Cas9/TRBO-gGFP and Cas9/TRBO-gAGO1 were used to compare delivery of a singular guide compared to the multiplexed gRNA delivery vector TRBO-gGFP-gAGO1. At 10 dpi genomic DNA was sampled for indel analysis of *mgfp5* and both *NbAGO1* alleles. *NbAGO1-H* and *NbAGO1-L* amplicons were analyzed for indels using the surveyor nuclease assay (**Figure 2.5B**), while *mgfp5* was subjected to the *Bsg*I restriction enzyme resistance assay (**Figure 2.5C**). Remarkably, following indel quantification analysis there was not a decrease, but rather an increase in observed indel percentages for all three genes when sgRNAs are multiplexed within TRBO; from 15 to 27% for *NbAGO1-L*, from 8 to 11% for *NbAGO1-H*, and from 53 to 64% for *mgfp5*. These results indicate that at least two sgRNAs can be delivered using the TRBO-sgRNA delivery system with an increase in delivery efficiency. Due to these results, we concluded that sgRNA positioning does not affect indel percentages during sgRNA multiplexing, offering a substantial level of flexibility not previously known for sgRNA design.

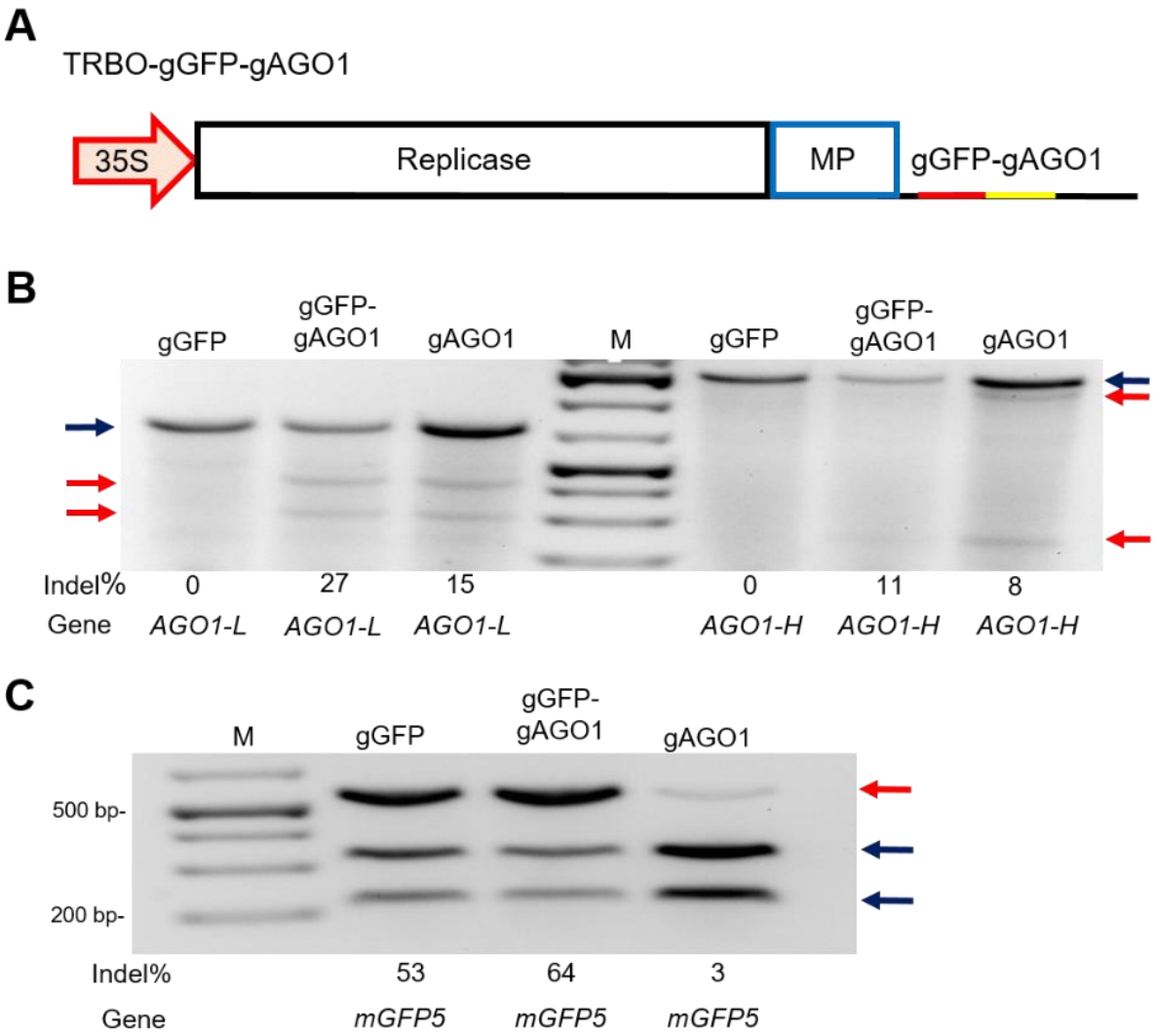


Figure 2.5. Indel analysis for TRBO delivery of single and multiple sgRNAs. A) TRBO-gGFP-gAGO1 co-delivery construct used for simultaneous delivery of both gGFP and gAGO1. **B** and **C** *N. benthamiana* treatments included pHcoCas9 co-infiltrated with one of the following: TRBO-gGFP (gGFP), TRBO-gGFP-gAGO1 (gGFP-gAGO1), and TRBO-gAGO1 (gAGO1). M indicates DNA marker. Blue arrows indicate the presence of non-indel containing DNA and red arrows are indel containing DNA. **B)** NbAGO1-H and NbAGO1-L PCR amplicons treated with surveyor nuclease for indel detection. Marker (M) is same as in **Figure 2.1D**. **C)** The *mgfp5* amplicons treated with *Bsg*I restriction enzyme to detect indels.

TRBO-based protein overexpression and sgRNA delivery using a single TRBO delivery construct

The Cas9/TRBO-sgRNA transient knockout system appears to tolerate both 5′ and 3′ overhangs which was exhibited initially by the efficiency of the TRBO-sgRNA delivery without ribozymes and then through the multiplex delivery of sgRNAs *in planta* through a singular TRBO construct. These data suggest the possibility to deliver sgRNAs along with larger RNA molecules, such as a protein-encoding region. Delivery of biologically active sgRNAs while retaining TRBO-based overexpression of proteins would be of great benefit for biotechnology purposes and as a fundamental tool for functional genomic studies. To examine if an protein open reading frame could be co-delivered with a sgRNA (ORF-sgRNA), gGFP was selected to be delivered along with the *gfpc3* protein coding gene within the TRBO vector (TRBO-G) (Lindbo 2007). gGFP was designed to target only the 16c (host) chromosomal *mgfp5* sequence and therefore should not interfere with TRBO gfpc3 protein expression due to self-targeting from Cas9/gGFP. Viral-based GFP overexpression, which is indicative of TRBO-G replication and subgenomic RNA transcript delivery, should induce several fold higher expression than 16c-based GFP expression, allowing for direct visualization of viral dependent GFP expression.

To test if sgRNA proximity to the protein-coding region would disrupt indel percentages, gGFP was introduced directly 5′ or 3′ proximal to the *gfpc3* coding sequence, forming TRBO-G-5′gGFP and TRBO-G-3′gGFP respectively (**Figure 2.6A**). Specifically, the gGFP spacer sequence carries a start codon (*AUG*) and a stop codon (*UGA*) in the gRNA scaffolding within the same reading frame. We predicted that TRBO dependent GFP expression would not occur using the construct TRBO-G-5′gGFP due to the small ORF being upstream of the GFP start site.

However, by delivering a gGFP 5′ proximal to *gfpc3* sequence we could determine if gGFP proximity to the protein coding sequence would affect indel rates *in planta*.

Agroinfiltrations of 16c plants with pHcoCas9 alone or co-delivery of pHcoCas9 and TRBO-gGFP, TRBO-G-5′gGFP or TRBO-G-3′gGFP were carried out. At 3 dpi, infiltrated leaves were viewed under UV light to visualize TRBO-dependent GFP expression (**Figure 2.6B**). As we anticipated, neither pHcoCas9 alone nor the co-infiltrations of pHcoCas9 and TRBO-gGFP or TRBO-G-5′gGFP showed GFP expression at levels expected for viral delivery. However, TRBO-G-3′gGFP showed a dramatic increase of GFP expression at 3 dpi (**Figure 2.6B**). Following protein expression visualization, genomic DNA was sampled at 6 dpi then subject to *mgfp5* amplification followed by a *Bsg*I restriction enzyme resistance assay for indel quantification analysis. Remarkably, co-infiltrations of pHcoCas9 and TRBO-G-5′gGFP as well as pHcoCas9 and TRBO-G-3′gGFP showed 5% and 11% higher levels of indel formation, respectively, in comparison with the pHcoCas9/TRBO-gGFP control (**Figure 2.6C**). *Bsg*I resistant bands were subsequently cloned and sequenced to verify the presence of indels at the *mgfp5* target locus. Our results suggest that the additional GFP coding RNA does not hinder gGFP biological activity but might actually increase the occurrence of DSBs through an unknown mechanism. Notably, indel rates of *mgfp5* were observed at an 11% increase for both the ORF-sgRNA, TRBO-G-3′gGFP (**Figure 2.6C**), and multiplex delivery of sgRNAs construct, TRBO-gGFP-gAGO1 (**Figure 2.5C**), compared to the TRBO-gGFP control. These results confirm those above that sgRNA positioning has limited consequence for its activity. Moreover, the developed virus-mediated system offers the novel ability to over-express and inactivate several genes simultaneously.

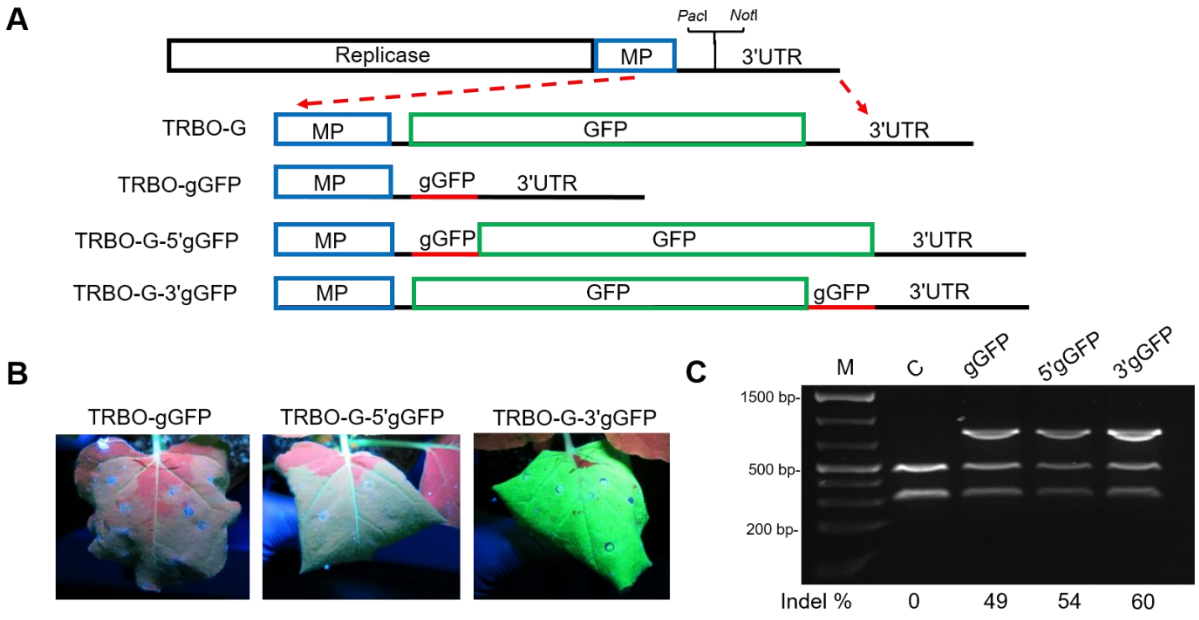


Figure 2.6. Co-delivery of GFP protein and gGFP using TRBO in 16c N. benthamiana. A) Depiction of GFP protein and gGFP co-expression constructs expressed from the TMV CP subgenomic promoter. The 5′ and 3′ preference for gGFP delivery was analyzed by creating constructs that carried both a 5′ and 3′ GFP protein coding sequence proximal gGFP. **B)** 16c *N. benthamiana* leaves co-infiltrated with pHcoCas9 and either TRBO-gGFP, TRBO-G-5′gGFP, or TRBO-G-3′gGFP followed by visualization of GFP protein under UV light at 3dpi. TRBO-G-5′gGFP was not expected to show viral vector-based expression of GFP due to a start and stop codon within the gGFP sequence. **C)** *Bsg*I restriction enzyme resistance assay at 6 dpi. Lane treatments are as follows: pHcoCas9 alone (C), pHcoCas9/TRBO-gGFP co-infiltration (gGFP), pHcoCas9/TRBO-G-5′gGFP co-infiltration (5′gGFP), and pHcoCas9/TRBO-G-3′gGFP co-infiltration (3′gGFP).

DISCUSSION

Several biological systems have benefited from the development of the CRISPR/Cas9 for targeted transient knockdown of gene expression levels, such as bacteria, yeast and human cell lines, but a similar localized transient system has yet to be routinely implemented for plant systems (Larson et al. 2013; Qi et al. 2013; Gilbert et al. 2013). Previously developed viral gene editing delivery methods focused on the creation of gene knockout plant lines (Baltes et al. 2014; Ali et al. 2015). Here we use CRISPR/Cas9 as a method to locally knockout genes in plants and demonstrate its potential use as a transient screening method for virus-plant interactions. First, we explored the potential of a TMV-based delivery vector, TRBO, for its ability to deliver biologically functional sgRNAs through the CP subgenomic promoter *in planta*. Through the design and testing of multiple sgRNA delivery systems in *N. benthamiana*, we were surprised to find that 5′ and 3′ sgRNA overhangs did not impede Cas9/sgRNA mediated DNA cleavage, while sgRNA overhangs negated Cas9 nuclease activity *in vitro*. In fact, the HH ribozyme in the TRBO-RGR construct appears to reduce the fitness of replicons while increasing the accumulation of mutations within the HH sequence *in planta*.

The Cas9/TRBO-sgRNA system transiently delivered and stably sustained indel rates at 60-70% within 7 dpi, creating knockout levels potentially sufficient for transient functional genetic screens. Disruption of the coding regions in the gene paralogs *NbAGO1-H* and *NbAGO1-L* demonstrated the ability to transiently target complementary regions among genes. Remarkably, the delivery of multiple biologically active sgRNAs within the same TRBO construct resulted in an increase in indel induction. Due to the successful delivery of two fused sgRNAs *in planta*, we anticipated that a larger RNA transcript caring both a sgRNA (gGFP) and a protein coding sequence (*gfpc3*) could be delivered using TRBO-G-3′gGFP. Delivery of

Cas9/TRBO-G-3′gGFP successfully expressed GFP protein and increased the percentage of indels compared to the TRBO-gGFP control. One potential explanation for the increase in indels in the ORF-sgRNA co-delivery method, might be due to the longer TRBO 3′ end of TRBO-G-3′gGFP that aided in TMV replicase fidelity and extension from the TRBO CP subgenomic promoter as compared to TRBO-gGFP. For instance, increased distance between the upstream RNA promoter and the 3′ tRNA-like structure has been shown to stimulate transcription (Culver et al. 1993).

As mentioned previously, the current Cas9/sgRNA based expression systems are limited by low expression of sgRNAs *in vivo* (Li et al. 2013; Nekrasov et al. 2013). Using our Cas9/TRBO-sgRNA delivery approach we aimed at saturating the system with sgRNAs to allow for more efficient editing. We believe that the current limitations of the Cas9/TRBO-sgRNA delivery system, explained here, may now be the availability of Cas9 protein. We suggest that increasing Cas9 expression, such as using transgenic Cas9 plants (Ali et al. 2015), could increase the efficiency of the system. However, while transgenic Cas9 expression might improve the accumulation of gene specific indels, it removes the benefit of a transient delivery system. Even though the Cas9/TRBO-sgRNA platform is still reliant on T-DNA that could be integrated into the genome, this may represent a key first-step towards separating gene editing tools from transformation technology. Instead of a transgenic-based system, viral vectors could co-deliver Cas9 and sgRNA to host plants. While such a method might be appealing, there are a limited number of viral vectors that will tolerate an insertion the size of the *S. pyogenes* Cas9 ORF (~ 4.4 kb). Additional complications can arise through the delivery of multiple viral vectors, such as spatial expression differences within tissues and the requirement for shared susceptible hosts among the vectors used.

The TRBO-sgRNA transient gene editing platform varies from other viral-induced gene editing (VIGE) platforms such as geminivirus-based delivery systems (Baltes et al. 2014; Yin et al. 2015), and the *Tobacco rattle virus* (TRV) sgRNA delivery system (Ali et al. 2015). Compared to those initial reports we aimed to extend the concept by primarily focusing on exploiting the high titer of TMV for the purpose of solving the relatively low yield dilemma of sgRNAs from current U6 promoter-based delivery systems. While we did not directly demonstrate that the Cas9/TRBO-sgRNA system increases indel percentages compared to constitutive expression of sgRNAs, the time course analysis of TRBO-sgRNA in comparison with the replication-negative constructs did address this issue in some capacity considering the vector was still driven by the constitutive 35S promoter. Additionally, we aimed to investigate the influence of 5′ and 3′ guide RNA overhangs through the utilization of ribozyme delivery systems (Li et al. 2013; Nekrasov et al. 2013; Xie and Yang 2013). The TRBO vector was of particular interest due to high expression from the 3′ co-terminal CP subgenomic RNA while also restricting sgRNA expression to only infiltrated leaves. This approach differs from the TRV and geminivirus sgRNA delivery tools, where the primary goal was to produce seeds with gene knockouts through virus systemic movement, not localized knockouts with the capacity to be used as a transient screening tool (Ali et al. 2015; Yin et al. 2015). Additionally the geminivirus-based DNA replicon was used to deliver Cas9, sgRNAs, and a DNA donor template for creating knock-in plants (Baltes et al. 2014). The addition of such a large insertion may have possibly reduced the efficiency of genomic DSB events and subsequent genomic recombination events as well as localized movement of the replicon. While these prior pioneering methods created knockout and knock-in plant lines, our strategy aimed to provide localized transient screening while avoiding systemic movement of the viral vector. It is also worth noting that the systemic

movement of the aforementioned viruses relies on CP expression, thus allowing the recombinant virus to be encapsidated and perhaps more easily disseminated (Ali et al. 2015; Yin et al. 2015). The potential unintended biological consequences of creating functional virions should be a consideration when using these technologies. The RNA-based TRBO-sgRNA delivery system may alleviate many of these concerns.

The delivery of multiple sgRNAs as well as the simultaneous delivery of a protein coding sequence and a sgRNA fusion transcript using a single TRBO vector was effective even with the predicted large sgRNA nucleotide overhangs from the subgenomic delivered sgRNA. While the specificity of the complementary region within the 5′ end of the spacer sequence has been studied thoroughly (Dahlman et al. 2015; Fu et al. 2014), it has not been documented, to our knowledge, that significant (>1 nt) 5′ spacer proximal overhangs are tolerated without affecting the induction of DSBs *in vivo*. The effects of sgRNA 3′ overhangs have been passively examined previously with a mutant Cas9 nuclease with deactivated nuclease domains (dCas9) (Konermann et al. 2015; Zalatan et al. 2015). These studies have led to some indication that binding of Cas9 might be hindered by the presence of 3′ overhangs due to the incorporation of RNA binding protein RNA motifs at the 3′ end of the sgRNA scaffold RNA (Zalatan et al. 2015), but this is far from being conclusive.

Complications resulting from sgRNAs harboring 5′ and 3′ overhangs has been an area of concern for researchers since the inception of CRISPR/Cas9 as a gene editing technology (Cong et al. 2013; Jinek et al. 2012; Mali et al. 2013). This is an issue we expected but did not encounter when developing the TRBO-sgRNA delivery system *in planta*. One of three activities could explain the tolerance of 5′ and 3′ sgRNA nucleotide overhangs observed in the Cas9/TRBO-sgRNA delivery system, none of which have been previously explored in detail: 1)

Cas9 tolerates large non-complementary overhangs 5′ proximal to the spacer and 3′ to the scaffold RNA *in planta,* 2) Cas9 has the ability to cleave sgRNA overhangs *in planta*, or 3) endogenous proteins in *N. benthamiana* cleave 5′ and 3′ sgRNA nucleotide overhangs *in planta*. Extensive *in vitro* and *in vivo* studies have been conducted on the activity of Cas9 that have demonstrated the importance of the sgRNA spacer sequence on the nuclease activity of Cas9 (Sternberg et al. 2014; Dahlman et al. 2015; Mali et al. 2013; Fu et al. 2014; Jinek et al. 2012). Largely these studies have assumed that 5′ overhangs reduce the incidence of DSBs created by Cas9, with the exception of the addition of guanine for increased T7 RNA polymerase transcription and RNA polymerase III promoter expression which does not appear to affect Cas9 catalytic activity (Hsu et al. 2014; Nekrasov et al. 2013; Xie and Yang 2013). Based on these assumptions, a reduction in Cas9 DNA cleavage associated with sgRNA 5′ overhangs was expected to occur *in planta* for the gGFP and gHDV constructs. However, here we demonstrated that sgRNA overhangs within the delivery construct do not seem to effect catalytic activity *in planta*, while as expected they do abolish Cas9/sgRNA mediated cleavage *in vitro*. These observations taken together would indicate that the high indel percentages observed using the TRBO-gGFP and TRBO-gHDV delivery systems suggest that the removal of 5′ overhangs must be occurring *in planta*. Although sgRNA delivery systems have been adapted to remove RNA overhangs within the delivery system itself (Gao and Zhao 2014) or to rely on host RNases for programmed nucleotide overhang cleavage (Xie et al. 2015), removal of sgRNA overhangs through existing *in vivo* processes has not been reported, although it has been suggested (Cong et al. 2013; Karvelis et al. 2013).

The most popular method currently used to assess genetic factors and their effect on viral infection has been the VIGS screening method, where the host is typically infected with TRV

before being subjected to the pathogen of interest (Burch-Smith et al. 2004; Scholthof et al. 2011). Subjecting a plant to a bacterium and a virus to study the host effects of another virus elicits concern for the quality of the screen. Results from traditional VIGS screening are often probed for verification through transgenic knockdown techniques such as dsRNA hairpin technology (Odokonyero et al. 2015), T-DNA knockout insertion lines, or more recently CRISPR/Cas9 knockouts (Brooks et al. 2014). The adoption of our viral delivered sgRNAs technique could shorten the time of screening, increase the reliability of the transient screen, and decrease the space and plants needed to conduct screening. Further genetic studies, which are not necessarily accessible when using VIGS screening, such as targeting promoters, enhancers and insulator sequences is now feasible for understanding functionality within the context of a native gene environment (Zhang et al. 2015; Flavahan et al. 2016; Basak and Nithin 2015). Additionally, the adaptation of sgRNA delivery to include different viruses, viral strains, or genetic variants has the potential to help understand host symptom development and viral disease onset (Mandadi and Scholthof 2013).

Although a TRBO-sgRNA screening method is promising, there are certainly some constraints when it comes to CRISPR-based gene targeting particularly when considering large or polyploid plant genomes such as *N. benthamiana*. Additionally, transient CRISPR screens create a chimera of cell types including heterozygous, wide type and true knockout cells which can greatly reduce the quality and effectiveness of the screening method (Morgens et al. 2016). These complications could even be a larger problem when targeting genes that are functionally redundant. In these instances the use of both VIGS and CRISPR based screening methods in parallel could increase the reliability of experimental results and provide findings within one

screen that is overlooked through the others (Morgens et al. 2016; Barrangou et al. 2015; Shalem et al. 2015).

In summary, this study provides new insight into the previously unexplored flexibility of delivering one or more sgRNAs launched from the TRBO vector without the necessity of designing processing elements. Furthermore, it is shown that the same vector can be used for the simultaneous delivery of a sgRNA for gene editing along with an ORF for over-expressing a protein of interest. These findings illustrate new properties associated with virus mediated delivered CRISPR/Cas9 gene editing.

MATERIALS AND METHODS

Design of sgRNA encoding segments and construct development

The sgRNA targets were identified using the CRISPR design toolset (benchling.com). Gene sequences were used from previously reported and deposited sequences for *mgfp5* (Haseloff et al. 1997), *NbAGO-L*, and *NbAGO-H* (Gursinsky et al. 2015). The *NbAGO-L* and *NbAGO-H* sgRNA (gAGO1) was designed through gene sequence alignment and identification of complementary areas for potential co-targeting (benchling.com). Sequences for each of the sgRNA sequences can be found in **A-6**.

RGR was designed as described previously (Gao and Zhao 2014), using a *mgfp5* targeting sgRNA as mentioned above. The RGR construct was synthesized using custom gene synthesis (GenScript) with the inclusion of a 5′ *Pac*I and a 3′ *Not*I site for cloning into TRBO, as described in Lindbo (2007). Towards this, the RGR construct was used as a template to create both the HDV and gGFP construct using PCR fragment cloning. Forward primers were designed carrying an additional 5′ *Pac*I site and complementary regions overlapping the *mgfp5* spacer

sequence of the sgRNA. Reverse primers were designed for both the gHDV and gGFP constructs using the 3′ *Not*I site with complementary regions to the 3′ HDV and scaffold portion of the sgRNA, respectively. The high fidelity Q5 polymerase (New England BioLabs) was used to produce PCR amplicons, corresponding to each construct, and cloned into pGEM-T Easy (Promega). The cloned fragments and the TRBO vector were then subjected to a *Pac*I and *Not*I double digest and a subsequent ligation step to produce TRBO-RGR, TRBO-gHDV, and TRBO-gGFP.

The gAGO1 target was constructed using TRBO-gGFP as a template for site-directed mutagenesis using the Q5 Site-Directed Mutagenesis Kit (New England BioLabs). Primers for site-directed mutagenesis were designed using the online NEBaseChanger (New England BioLabs) tool and the online generated protocol was used to carry out the mutagenesis reaction. TRBO-gGFP-gAGO1 was constructed using DNA assembly where TRBO-gGFP was used as the destination vector. Using TRBO-gGFP overlapping primers, gAGO1 was amplified from TRBO-gAGO1 and inserted 3′ to the gGFP scaffold, of *Not*I linearized TRBO-gGFP, using NEBuilder HiFi DNA Assembly master mix (New England BioLabs).

RM was constructed using the TRBO-gGFP plasmid. TRBO-gGFP was digested with *Sma*I and *Stu*I restriction enzymes to excise 1,419 bp region in the replicase coding region. This DNA fragment cleaned using the Zymoclean Gel DNA Recovery kit (Zymo Research) and used for a blunt-end ligation using T4 ligase (New England BioLabs) to create the final replication incompetent (RM) construct.

TRBO-G was used as destination cloning backbone for both the TRBO-G-5′gGFP and TRBO-G-3′gGFP constructs. TRBO-G-5′gGFP was constructed by linearization of TRBO-G with *Pac*I. The gGFP construct was amplified from TRBO-gGFP using forward and reverse

primers that overlapped with the *Pac*I linearized TRBO-G plasmid. The forward primer consisted of a spacer sequence (TAA) between the TRBO-G overlapping sequence and gGFP overlapping sequence to maintain the *Pac*I site within the final TRBO-G-5′gGFP construct. The TRBO-G digestion and gGFP amplification were visualized using gel electrophoresis followed by excision of bands at the expected molecular weight and cleaned using the Zymoclean Gel DNA Recovery kit (Zymo Research). The final TRBO-G-5′gGFP construct was made using HiFi DNA Assembly master mix (New England BioLabs) with linearized TRBO-G and gGFP fragments following the manufacture's recommended protocol. TRBO-G-3′gGFP was constructed essentially as TRBO-G-5′gGFP but TRBO-G was instead linearized using the *Not*I site 3′ to the GFP stop codon. Additionally, the reverse primer for gGFP amplification consisted of a spacer sequence (GGCCGC) between the TRBO-G overlapping region and the gGFP complementary region to repair the *Not*I site within the final TRBO-G-3′gGFP construct.
The Cas9 expression construct was assembled using the human codon-optimized Cas9 nuclease (HcoCas9) (Addgene: 42230) (Cong et al. 2013). HcoCas9 was PCR amplified using a forward primer with a *Bam*HI site extension and a reverse primer carrying an extension with an *Xho*I site. A modified pRTL22 (Restrepo et al. 1990) carrying the 3′ *Tobacco etch virus* (TEV) sequence upstream of the CaMV 35S terminator sequence was used as the sub-cloning vector. The 35S-Cas9-term cassette was then transferred into the binary destination plasmid pBINPLUS-sel through the *Hin*dIII site.

Agroinfiltrations

Agrobacterium tumefaciens strain GV3101 (pMP90RK) was used for agroinfiltration of all binary vectors used, as described previously (Odokonyero et al. 2015). In brief, cultures were

grown overnight (16-20 hrs) under constant shaking (250 rpm) at 28 ºC in LB media with 50 mg/L Kanamycin. Cells were pelleted by centrifugation and resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES pH 5.7, and 200 µM acetosyringone). TRBO-based cultures were resuspended to a final infiltration concentration of OD_{600} 0.4 and pBINPLUS-sel Cas9 expression vector at OD_{600} 0.5. Co-infiltrations were carried out by mixing equal volumes of resuspended TRBO and Cas9 *N. benthamiana* cultures to the above mentioned final concentrations. Four week old 16c plants were infiltrated with *Agrobacterium* suspensions on the abaxial side of the leaf and returned to normal growth conditions.

DNA and indel assays

Single plant DNA samples for indel assays were carried using 50 mg of leaf tissue from three infiltrated leaves, totaling 150 mg of tissue, to avoid tissue-dependent effects as well as to create a pooled biological replicate. DNA extractions were then carried out using the ZR Plant/Seed DNA Miniprep kit (Zymo Research). For the *Bsg*I restriction enzyme resistance assay and the Surveyor Nuclease (Integrated DNA Technologies), 100 ng of genomic DNA was used for PCR amplification of either *mgfp5, NbAGO1-H* or *NbAGO1-L*. Amplicons were then cleaned using DNA Clean & Concentrator -5 (Zymo Research) kit and resuspended in DNase and RNase-free water. Amplicon concentrations were measured using a Nanodrop and 200-400 ng DNA used for final Surveyor Nuclease and 400-500 ng DNA for *Bsg*I digestions. Amplicons of genomic *mgfp5* were subject to *Bsg*I restriction enzyme and incubated at 37ºC overnight. *NbAGO1-H* and *NbAGO1-L* amplicons were treated with Surveyor Nuclease a DNA mismatch endonuclease. Surveyor Nuclease digestions were carried out using the manufacturer's instructions (IDT). Both the *Bsg*I restriction enzyme resistance assay and the Surveyor Nuclease

reactions were visualized using 2% agarose gel electrophoresis stained with ethidium bromide. Image files (.tif) were uploaded in the image analysis software ImageJ (NIH). The background signal was subtracted from gel images and band intensities were measured using standard gel peak analysis workflow.

RNA extractions and (RT)-PCR

Plant RNA extractions were carried out using the Direct-zol RNA Miniprep kit (Zymo Research) following the manufacturer's instructions. cDNA was synthesized with equal volumes of total RNA using the SuperScript III first stand synthesis kit (Invitrogen) along with a TMV 3′-UTR-specific reverse primer, RT-PCR was carried out using the TMV 3′-UTR specific reverse primer and a TMV specific forward primer 5′ to the TRBO *Pac*I restriction cite using *Taq* polymerase (NEB). RT-PCR bands were gel extracted using Zymoclean Gel DNA Recovery kit (Zymo Research) and cloned into the pGEM-T Easy vector system (Promega) for sequencing.

GFP imaging

N. benthamiana 16c plants or leaves of TRBO-G infiltrated plants were visualized under a handheld UV mercury lamp as previously described (Odokonyero et al. 2015; Everett et al. 2010). TRBO-gGFP and RM-gGFP treated leaf images were visualized using an Olympus SZX7 stereomicroscope with a GFP filter essentially as described previously (Gao et al. 2013).

Protein extractions and western blotting

For Cas9 protein detection, extractions were carried out using 50 mg of 3 dpi 16c leaf tissue infiltrated with both pFGC-pcoCas9 (Addgene plasmid: 52256) and pHcoCas9. Tissue was

ground in liquid nitrogen followed by suspension in 250 µl of 1:1 1x TE buffer and 5x Laemmli SDS buffer. Samples were boiled for 5 minutes before centrifugation for 2 min at 10,000xg at room temperature. The cleared lysate (supernatant) was kept for SDS-PAGE where 20 µl of each treatment was electrophoresed through a 4-20% gradient ExpressPlus PAGE gel (GenScript) in 1x NuPAGE MOPS SDS running buffer (ThermoFisher scientific) at 100V until adequate protein ladder separation occurred (~3 hrs). Proteins were then transferred onto a nitrocellulose membrane in Tris-glycine transfer buffer (25 mM Tris, 192 mM glycine) at 100 V for 60 min. The membrane was then blocked in 5% non-fat milk in TBS-Tween with gentle agitation for 30 minutes at room temperature followed by an overnight incubation in DYKDDDDK tag mouse mAb, (GenScript) at a dilution of 1:2,000 at 4ºC. Membranes were washed with TBS-Tween three times for 5 minutes each wash. Secondary goat anti-mouse IgG conjugated with horseradish peroxidase (Sigma) was incubated for 1 hr at a dilution of 1:2,000. Membranes were then washed briefly with TBS-Tween before three final wash steps in TBS for 5 min each. Colorimetric detection of Cas9 protein was achieved by dissolving 30 mg of 4-chloro-1-naphthol in 10 ml of cold methanol and adding 30 µl of 30% hydrogen peroxide to 50 ml of 1x TBS. The two substrates were mixed and added to the nictrocellulose membrane. This was followed by a dark incubation period of at least 15 min with gentle agitation for signal development. Reactions were stopped by washing the membrane with distilled water, blotted dry, and then photographed.

GFP expression was measured essentially as stated above. However, protein samples after extraction were diluted 1:1 with equal quantities of 1x TE buffer and 5x Laemmli SDS buffer for increased downstream detection sensitivity. Following dilution, 20 µl of each sample was loaded and electrophoresed. GFP signal was detected as described previously except rabbit pAb to GFP (GenScript) at a dilution of 1:1,000 was used as the primary antibody followed by

using an IgG goat anti-rabbit horseradish peroxidase conjugate (Sigma) at a dilution of 1:2,000. Protein polyacrylamide gel staining was carried out by adding 50 ml of QC Colloidal Coomassie stain (BioRadiation) following manufacture protocol.

Ribozyme efficiency assay

In vitro ribozyme efficiency activity assays for the RGR, gHDV, and gGFP constructs were carried out using the final TRBO destination cloning vector as the template for PCR extension encompassing the cloning sites. A forward primer including the T7 promoter sequence was designed in the TMV MP ORF followed by the TMV 3′ UTR reverse primer. *In planta* ribozyme activity for the corresponding TRBO-sgRNA constructs was assessed by cDNA synthesis followed by an RT-PCR step using the T7-TMV complementary forward primer and TMV 3′ UTR reverse primer. PCR amplicons were then used as a template for T7 *in vitro* transcription reactions (New England BioLabs). Following transcript synthesis, RNA was visualized using 2% agarose gel electrophoresis stained with ethidium bromide.

Cas9/sgRNA *in vitro* cleavage assays

Subgenomic sgRNAs and gGFP RNAs were synthesized using T7 RNA synthesis methods as described above. Briefly, a forward primer starting at the CP subgenomic promoter transcriptional start site of TRBO caring a 5′ T7 promoter was used to PCR amplify each of the TRBO-gGFP, TRBO-gHDV, and TRBO-RGR constructs along with a reverse primer within the 3′ UTR. 150 ng of each of the PCR amplicons (sub-gGFP, sub-gHDV, and sub-RGR respectively) was then used as a template for T7 *in vitro* transcription reactions (New England

BioLabs). RNA synthesis reactions were verified using 1% agarose gel electrophoresis stained with ethidium bromide and quantified using a Nanodrop.

A PCR *mgfp5* fragment was amplified from untreated 16c genomic DNA and TA cloned into pGEM-T Easy (Promega). TA-*mgfp5* was used as the DNA template for each of the cleavage reactions. 100 nM of purified Cas9 Nuclease (New England Biolabs) was first incubated in Cas9 Nuclease reaction buffer (New England Biolabs) and 100 nM with either gGFP, sub-gGFP, sub-gHDV, or sub-RGR RNAs at room temperature for 5 minutes. 100 ng of purified TA-*mgfp5* plasmid and 1 µL of *Sca*I restriction enzyme (New England Biolabs) was then added to each reaction (20 µL total volume) and incubated for 30 minutes at 37ºC. Reactions were visualized using 1% agarose gel electrophoresis stained with ethidium bromide.

CHAPTER III

IN PLANTA 5′ PROCESSING OF GUIDE RNAS CREATES CATALYTICALLY ACTIVE CAS9 COMPLEXES

INTRODUCTION

The CRISPR/Cas9 platform, native found *Streptococcus pyogenes,* has been developed into a diverse set of functional genetic tools, including gene editing technology (Mali et al. 2013; Cong et al. 2013) and transcriptional control through gene activation (Gilbert et al. 2013) and repression (Qi et al. 2013). These technologies rely on the single guide RNA (sgRNA) programmable endonuclease Cas9 for specificity. One central restriction on the deployment of these CRISPR tools in eukaryotic organisms is the ability to deliver both the Cas9 nuclease and sgRNAs in living cells. To circumvent this problem, viral vectors have been used in a wide variety of organisms for delivery of the protein and RNA products in adequate concentrations for a phenotypic response (Baltes et al. 2014; Malina et al. 2013; Platt et al. 2014; Cody et al. 2017). However, the long-standing assumption that sgRNA delivery requires specific engineering there are not 5′ and 3′ nucleotide overhanging sequences unrelated to the functional sgRNA sequence has handicapped the use of the technology.

Previously we created an optimal sgRNA delivery system using the *Tobacco mosaic virus* based vector, TRBO, where we examined different sgRNA delivery strategies aimed to eliminate nucleotide overhangs 5′ proximal to the spacer sequence and 3′ proximal to sgRNA scaffolding through the use of RNA catalytic ribozymes (Cody et al. 2017). Contrary to typical CRISPR dogma, the sgRNA without ribozymes harboring large nucleotide overhangs upstream and downstream of the sgRNA sequence performed optimally in comparison to the ribozyme

containing construct. Following this observation, we challenged the TRBO vector system to co-deliver a protein coding region and a sgRNA (TRBO-G-3′gGFP), from a single transcript, resulting in genomic indels and high concentration of viral-produced protein. These results, in addition to results from another study (Mikami et al. 2017) and other serendipitous findings (Ali et al. 2015; Cong et al. 2013), contradict the consensus in the field which suggests that sgRNAs containing nucleotide overhangs inactivate the Cas9-sgRNA complex catalytic activity.

Complications resulting from sgRNAs harboring 5′ and 3′ overhangs has been an area of concern for researchers since the inception of CRISPR as a gene editing technology (Cong et al. 2013; Jinek et al. 2012; Mali et al. 2013). Predicted *in planta* produced subgenomic RNAs from the TRBO vector carry both 5′ and 3′ overhangs leading to one of three activities that could explain Cas9 mediated DSBs for this delivery system: 1) Cas9 tolerates large non-complementary overhangs 5′-proximal to the spacer and 3′-proximal to the scaffold RNA; 2) Cas9 has the ability to cleave sgRNA overhangs *in planta*; or 3) endogenous ribonucleases in *Nicotiana benthamiana* cleave sgRNA *in vivo*. The latter point has been the most popular implication, which others have suggested based on studies in bacterial systems harboring native CRISPR systems (Deltcheva et al. 2011; Karvelis et al. 2013). However, sgRNA 5′ overhang processing has not been supported through the identification of a CRISPR associated protein (Cas) or a protein in complex with Cas9 in the bacterial models harboring native CRISPR systems (Karvelis et al. 2013; Deltcheva et al. 2011; van der Oost et al. 2014). The lack of knowledge of how 5′-ends of sgRNAs are processed has yielded a considerable amount of focus on engineering methods for 5′-proximal and 3′-proximal nucleotide removal of sgRNAs in eukaryotic organisms (Xie et al. 2015; Cermak et al. 2017). Due to TRBO being an efficient protein and sgRNA co-delivery tool in *N. benthamiana*, and the overall lack of knowledge of

native 5′ CRISPR RNAs (crRNA) processing currently in the literature—synonymous to 5′ sgRNA processing used in our models here—we investigated what is occurring to the 5′ end of sgRNAs *in vivo*, specifically in the plant *N. benthamiana*.

Here, we use *in vitro* assays with Cas9 and sgRNA transcripts containing nucleotide (nt) overhangs (nucleotides not corresponding/aligning to the 100 nt chimera sgRNA sequence) of either or both of the 5′ and 3′ ends. In doing so, we concluded that 5′ overhangs completely inhibit the ability of the Cas9-sgRNA complex *in vitro*. Following these results we hypothesized, and conclusively find, that upon co-infiltration of a Cas9 and the TRBO protein-sgRNA co-expression constructs in *N. benthamiana* that Cas9 bound transcripts were enriched for sgRNAs which lacked the additional 5′ transcript sequence expected using the viral delivery vector. Further analyses determined that the 5′ nucleotide overhang structure on sgRNA transcripts were removed (processed) in the plant cytosol. Next, we delivered a transcript which mimicked the expected viral-sgRNA delivery transcript using a nuclear promoter, and demonstrated that these transcripts also formed catalytically active Cas9-sgRNA complexes, albeit at levels lower to that of the positive control which only delivered sgRNAs without 5′ overhang structures. Finally, we elected to interrogate two cytosolic RNase pathways for involvement in the 5′ processing event responsible for activating the Cas9-sgRNA complex, the first being RNA silencing and the second a 5′ to 3′ exoribonuclease pathway. While these assays lacked a conclusive result for 5′ processing it did allow us to develop a tentative model for creating catalytically active Cas9/sgRNA complexes *in vivo*. Ultimately, the results from these experiments will have far reaching impacts on the development of CRISPR technology in future applications, and, of equal importance, this system provides a model for understanding fundamental biology of the native CRISPR/Cas9 5′ processing system.

RESULTS

Predicted *in planta* subgenomic RNAs with 5′ overhangs negate catalytic activity of Cas9 *in vitro*

To test whether Cas9 can cleave a DNA template harboring a protospacer sequence using a sgRNA containing overhangs on either the 5′, 3′, or 5′ and 3′ ends, we elected to use the viral-based protein and sgRNA overexpression tool we previously developed, TRBO-G-3′gGFP (**Figure 3.1A**), as a template to conduct *in vitro* Cas9 cleavage assays (Cody et al. 2017). T7 promoter-carrying forward primers were designed at the subgenomic RNA transcription start site (T7-F1) and at the start of the spacer sequence of gGFP (T7-F2) to replicate both 5′ overhang carrying sgRNA and "clean" sgRNA (lacking extraneous nucleotides), respectively (**Figure 3.1B**). To evaluate 3′ overhang effects on Cas9 nuclease activity, reverse primers were designed both in the 3′ TMV UTR and on the 3′ most end of the sgRNA scaffolding, to replicate both 3′ overhang carrying sgRNA and "clean" sgRNA, respectively (**Figure 3.1B**). PCR amplifications of 5′ overhang carrying (T7/F1-R2), 3′ (T7/F2-R1), 5′ and 3′ overhang containing (T7/F1-R1), and "clean" gGFP (T7/F2-R2) were used as a DNA template for T7 transcription reactions. The resulting T7 produced RNA transcripts were used for loading into purified Cas9 protein. Genomic DNA from the *mgfp5* harboring transgenic *N. benthamiana* 16c plants served as a template for amplification of *mgfp5* and amplicons were subsequently used as the DNA target for *in vitro* assays. A successful Cas9 cleavage is merited by the presence of digested DNA template that only occurred, in this case, when using a gGFP transcript without 5′ overhangs (**Figure 3.1C**). Surprisingly, Cas9 still cleaved target DNA with a large 3′ gGFP overhang *in vitro* (**Figure 3.1C**). These results indicate that while 3′ sgRNA overhangs can be present and still

allow for Cas9 dependent DSBs, sgRNAs carrying 5′ spacer sequence adjacent overhangs cannot be present for Cas9 DNA cleavage under these conditions.

Concentrations of sgRNAs in the previous Cas9 cleavage assays (**Figure 3.1C**) were used at levels suited for optimal non-overhang containing sgRNAs. Perhaps sgRNAs containing 5′ overhangs do allow for Cas9 programed DNA cleavage, but the efficiency of the Cas9-sgRNA complex catalytic activity is reduced. To mimic the TRBO-sgRNA delivery system, which should produce an abundance of sgRNAs *in planta*, and to rule out the possibility of sgRNA dosage-dependent Cas9 DNA catalysis events, we further examined catalytic activity using increasing transcript concentrations (30 nM-1500 nM) of 5′ overhang carrying sgRNAs (T7/F1-R2) with the Cas9 *in vitro* cleavage assay system. Even with large concentrations of 5′ overhang gGFP template available for Cas9 loading, there was no evidence of DNA cleavage (**Figure 3.1D**). These results suggest that the increased concentrations of gGFP expected through TRBO delivery *in planta* could not be the source of efficient Cas9 editing without native 5′ sgRNA processing capabilities.

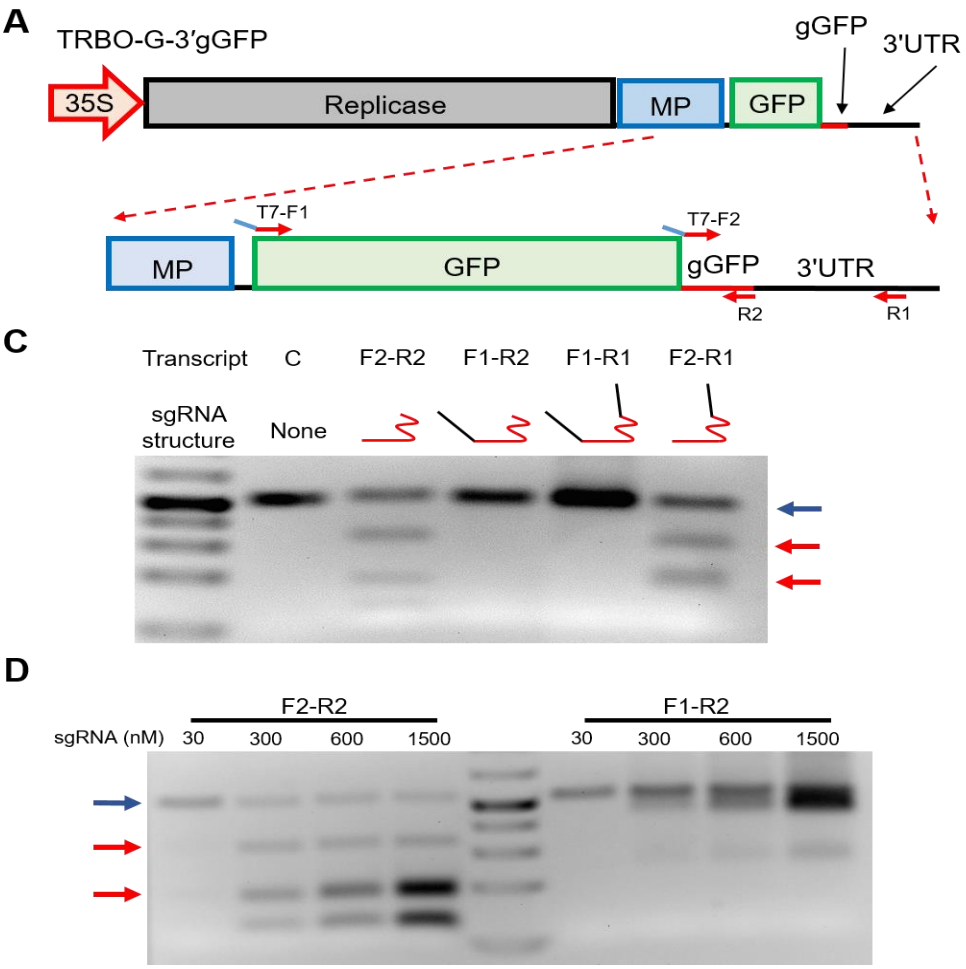


Figure 3.1. ***In vitro*** **Cas9 cleavage using sgRNA with varying 5′ and 3′ nucleotide overhangs.** **A)** Genomic depiction of the TRBO protein and sgRNA delivery tool TRBO-G-3′gGFP. **B)** A zoomed in view of the TRBO-G-3′gGFP genome to illustrate T7 promoter carrying primers (blue lines with red arrows) used for T7 transcription reactions. Primers (red arrows) were designed to test if *in vitro* RNA products can direct Cas9 based DSBs. Blue lines upstream of the "T7" marked promoters represent the T7 promoter sequence used for transcription reactions. **C)** T7 transcription reactions using the corresponding primers as a template. Arrows indicate DNA bands. The blue arrow is undigested *mgfp5* DNA template. Red arrows indicate Cas9 digested *mgfp5* DNA template. Lanes are as follows: *mgfp5* DNA template and Cas9 nuclease only (C), gGFP without overhangs (F2-R2; red squigle), gGFP with a 5′ overhang (F1-R2; black line), gGFP with both 5′ and 3′ overhangs (F1-R1), and gGFP with a 3′ overhang (F2-R1). Cas9 DNA cleavage only occurs using gGFP transcripts without 5′ overhangs. **D)** sgRNA dosage-dependent cleavage of *mgfp5* DNA template using increased concentrations of gGFP without overhangs (F2-R2) and gGFP with a 5′ overhang (F1-R2). Increased concentrations of gGFP transcript do not yield measurable levels of Cas9 DSBs in F1-R2 transcripts. Bands that do not correspond to either undigested DNA or digested DNA (red and blue arrows) are sgRNA transcripts used in the assay.

Cas9 bound sgRNAs have processed 5′ ends *in planta*

Previously we established that co-delivery of pHcoCas9 (**Figure 3.2A**) and TRBO-G-3′gGFP (**Figure 3.1A**) delivers catalytically competent Cas9-sgRNA complexes *in planta* (Cody et al. 2017). However, *in vitro* results indicate that full length, TRBO generated, subgenomic RNA transcripts could not form catalytically active Cas9-sgRNA complexes due to the inability of the complex to cleave the DNA template. These results suggest that sgRNA 5′ ends are being processed *in planta*, presumably by host factors, to produce catalytically competent sgRNAs. To better understand the structure of sgRNAs bound to Cas9 *in planta* and to understand if 5′ processing is occurring, immunoprecipitation of Cas9 from *N. benthamiana* 16c plants infiltrated with pHcoCas9, TRBO-G-3′gGFP, or pHcoCas9 and TRBO-G-3′gGFP were performed followed by RNA extractions. Additionally, RT-PCR amplification was performed using three primer sets to detect an enrichment of TRBO-G-3′gGFP derived RNA product with a particular emphasis on shortened gGFP fragments (**Figure 3.2B**). Forward primers were designed in the genome of TMV within the movement protein (MP) coding segment (F1), at the GFP start codon (F2) and at the 5′-end of the gGFP spacer sequence (F3). Due to our previous determination that 3′ sgRNA overhangs do not impede Cas9-sgRNA ability to induce DSBs (**Figure 3.1C**), we elected to amplify sgRNA fragments using a reverse primer located within the sgRNA scaffolding (R2) to enable us to focus on the newly discovered biological relevance 5′ proximal to the spacer sequence.

Since we previously established that the majority of editing events occur 2-3 days post inoculation (dpi) (Cody et al. 2017), 3 dpi samples were assayed from each treatment for downstream analysis. Cas9 protein was isolated by immunoprecipitation (IP) using a Cas9 specific antibody followed by protein A/G agarose bead pull-down, and proper Cas9 protein

isolation from the protein A/G agarose beads was detected using a western blot (**Figure 3.2C**). RNA extractions were carried out using all three Cas9-IP samples as well as from total RNA samples for each tissue. RT reactions were performed using the sgRNA scaffold specific (R2) primer. RT-PCR amplifications showed a clear enrichment of gGFP specific amplicons (F3-R2) in the pHcoCas9 and TRBO-G-3′gGFP co-infiltrated tissue compared to the predicted subgenomic RNA product (F2-R2) and genomic/first subgenomic containing RNA product (F1-R2) (**Figure 3.2D**). The two negative controls, pHcoCas9 and TRBO-G-3′gGFP single construct infiltrated samples, did not amplify products for the expected molecular weight for each primer set, as expected. Notably, total RNA RT-PCR amplifications showed approximately equal quantities of product between the TRBO, and the pHcoCas9 and TRBO-G-3′gGFP co-infiltrated samples (**Figure 3.2D**) which also held true over 3, 5, and 7 dpi (**A-7**).

We next tested if there is processing specificity to the sgRNAs loaded in Cas9 by examining if the sgRNAs were specifically cleaved at the 5′ complementary region (devoid of overhangs). Forward primers were designed from the gGFP (F3) spacer sequence moving upstream of the subgenomic RNA in increments (**Figure 3.2B**). RT-PCR indicated a clear reduction in band intensity with primers used upstream and 5′ proximal to the gGFP spacer sequence (**Figure 3.2E**). These data suggest that the 5′ end of gGFP is being processed (cleaved) *in planta* to eliminate the nucleotide overhang produced during subgenomic RNA production and the resulting processed sgRNAs are either preferentially bound by Cas9, or are reliant on Cas9 binding for proper 5′ nucleotide removal.

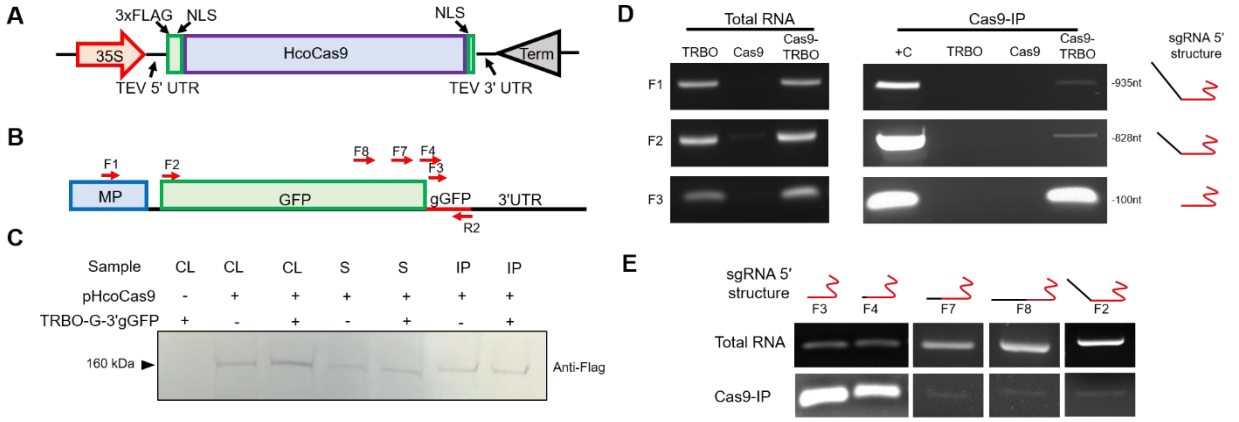


Figure 3.2. *In planta* sgRNA transcript processing, and Cas9 loading. A) The plant Cas9 expression construct pHcoCas9. This construct contains an N-terminal triple FLAG-tag (3xFLAG), nuclear localization signals (NLS), and a human codon-optimized Cas9 nuclease. Transcription is initiated by a 35S promoter and terminated by a 35S terminator. Transcripts also contain the *Tobacco etch virus* (TEV) 5′ and 3′ UTR for increased translation efficiency. **B)** The 3′ genomic organization and corresponding primers (red arrows) used for TRBO-G-3′gGFP RT-PCR experiments. The primers are designed to detect the 5′ condition of *in vivo* delivered gGFP products (presence of overhangs or not). **C)** Western blot using anti-Flag antibody to detect Cas9 in cellular lysate (CL), Cas9-IP (IP), and supernatant from Cas9-IP before washing (S) upon infiltration with pHcoCas9 and/or TRBO-G-3'gGFP, as indicated by + and -. **D)** RNA isolation of Cas9-IP samples for TRBO-G-3′gGFP (TRBO), pHcoCas9 (Cas9), and pHcoCas9/TRBO-G-3′gGFP (Cas9-TRBO). RT-PCR was performed using primers depicted in **A**, using cDNA from total RNA and Cas9-IP (also shown in **C**), to examine presence of *in vivo* gGFP 5′ overhangs. Enrichment of gGFP RNAs that do not encompass the predicted subgenomic RNAs as shown by ample amplification using F3-R2 primers and not F2-R2 for Cas9-TRBO. The positive control (+C) was carried out using TRBO-G-3′gGFP purified plasmid. **E)** RT-PCR using total RNA and Cas9 bound RNA (Cas9-IP) from 16c leaf tissue infiltrated with pHcoCas9 and TRBO-G-3′gGFP. Primers F4, F7, and F8 are located in increasing distance upstream to gGFP, respectively.

sgRNA *in planta* 5′ processing occurs without a protospacer

Due to the overwhelming majority of sgRNAs undergoing processing while loaded in Cas9, we next aimed to understand which host enzymes were responsible for 5′ sgRNA processing events *in planta*. One possible mechanism for processing that we hypothesized could be responsible is the host RNase H enzyme. Due to the "target" scanning demonstrated previously using *in vitro* experiments, where the Cas9-sgRNA complex scans host DNA for the presence of a protospacer adjacent motifs (PAM) and a complimentary 10 nt 3′ spacer "seed" sequence creating an R-loop structure (Sternberg et al. 2014). R-loop structures have been demonstrated to be responsible for RNase H RNA catalysis (Ohle et al. 2016). When using TRBO-G-3′gGFP as the sgRNA delivery tool, in this system, there would be large RNA overhangs located upstream of the 5′-most nucleotide of the protospacer complementary region, creating an R-loop structure which could potentially be recognized by RNase H enzymes.

To test if sgRNA processing was reliant on a host protospacer DNA region with 100% complementarity we co-infiltrated pHcoCas9 and TRBO-G-3′gGFP either into 16c *N. benthamiana* (contain *mgfp5*) plants, or wild-type (wt) *N. benthamiana* (not containing the genomic target) plants. Following infiltration, 16c and wt plants were harvested at 3 dpi. Cas9 was immunoprecipitated (IP) and RNA was sampled from both wt and 16c Cas9-IP and the total lysate used for the IP reactions. Western blot assays were used to detect expression of Cas9 in wt and 16c total lysate used for the Cas9-IP as well as the Cas9-IP reactions to ensure Cas9 precipitation (**Figure 3.3A**). RT-PCR products for total RNA lysate of 16c and wt plants indicated no discrepancies in band intensities when using previously designed primer sets used in **Figure 3.2A**. However, both 16c and wt Cas9-IP showed only 5′ processed sgRNAs bound to Cas9 through RT-PCR (**Figure 3.3B**). Following these results it can be concluded that sgRNA 5′

processing was not reliant on a protospacer being present and is occurring through another RNA degradation pathway.

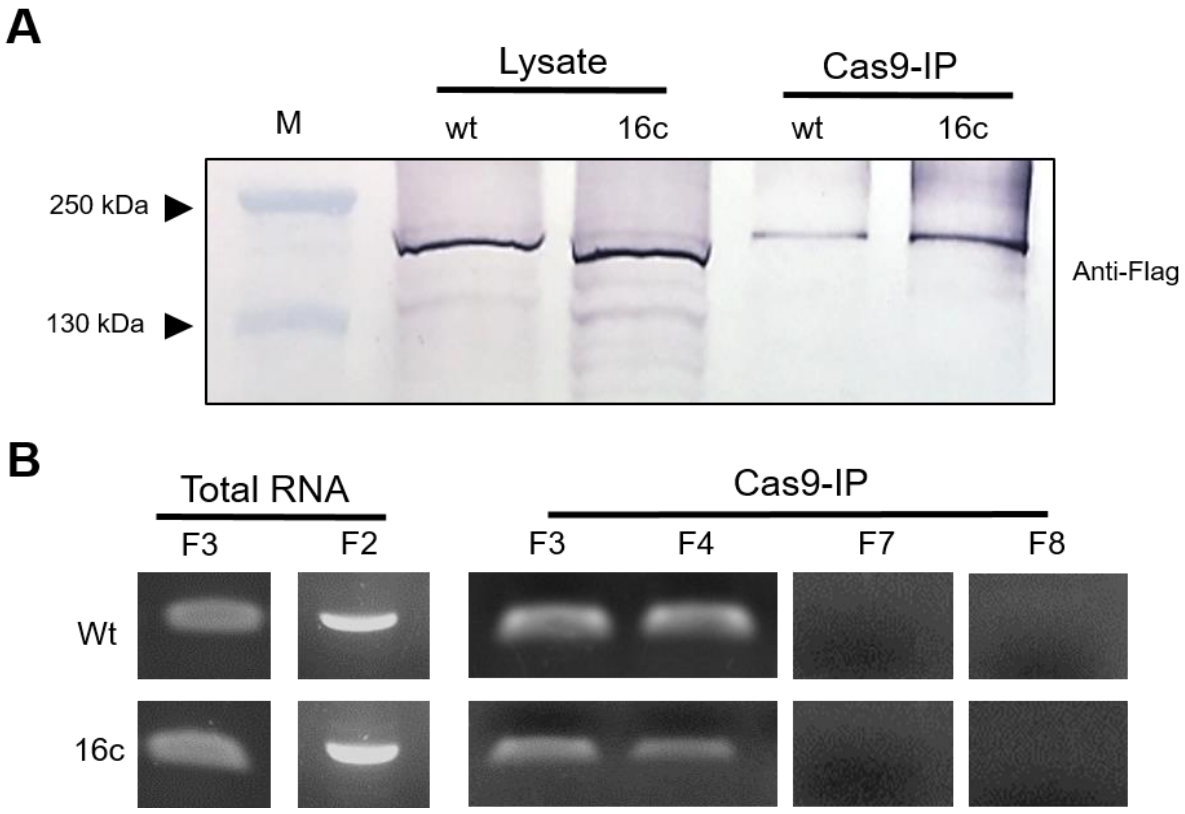


Figure 3.3. 5′ sgRNA processing with and without a protospacer sequence. *N. benthamiana* 16c plants were used as a protospacer containing system and wt plants as non-protospacer containing system for detecting gGFP 5′ processing. **A)** The wt and 16c plants were infiltrated with pHcoCas9 and TRBO-G-3′gGFP and probed for Cas9 expression in total protein lysate and Cas9-IP samples. Anti-Flag primary antibody was used for Cas9 detection. M represents the lane containing the protein marker. **B)** Total RNA and Cas9-IP RNA from samples in **A**, of both wt and 16c plants, were used for RT-PCR assays to detect the 5′ structure of gGFP containing transcripts from TRBO-G-3′gGFP. The primer sets are as presented in **Figure 3.2B**.

TRBO synthesized sgRNAs bound to Cas9 are processed in the cytoplasm

We next interrogated the subcellular localization of Cas9 protein and bound sgRNAs to identify the cellular location of 5′ sgRNA processing (nucleus or cytosol). Using the previously developed RT-PCR scheme (**Figure 3.2A**), we determined the ratio of unprocessed full-length subgenomic RNAs to 5′ processed gGFP both with and without cellular production of Cas9 protein and also the cellular localization of the processing events. Following cellular fractionations, equal cellular fractions were analyzed for the presence of Cas9 protein through western blotting which indicated that Cas9 preferentially localizes to the nucleus, albeit there remains a population of Cas9 protein in the cytosol (**Figure 3.4A**). Both nuclear and cytosol isolations from pHcoCas9 and TRBO-G-3′gGFP co-infiltrated tissue were then used for Cas9-IP (**Figure 3.4A**), followed by RNA extractions. Total RNA was also extracted from pHcoCas9 and TRBO-G-3′gGFP total cellular lysate as well as from the cytosol and nuclear lysate fractions. RT-PCR analysis from the total nuclear lysate and the Cas9-IP isolated from the nuclear fraction indicated that sgRNAs were being processed prior to translocation into the nucleus (**Figure 3.4B**). Additionally, there was a clear enrichment for specific gGFP 5′ processed forms in the Cas9-IP cytosolic fraction reactions as compared to the reactions from the total RNA in cytosolic fraction (**Figure 3.4B**). While the cytosolic total fraction lysate showed no discrepancies between the gGFP processing forms, and was indistinguishable from the total lysate control, the Cas9-IP RNA contained mostly 5′ processed forms of sgRNAs.

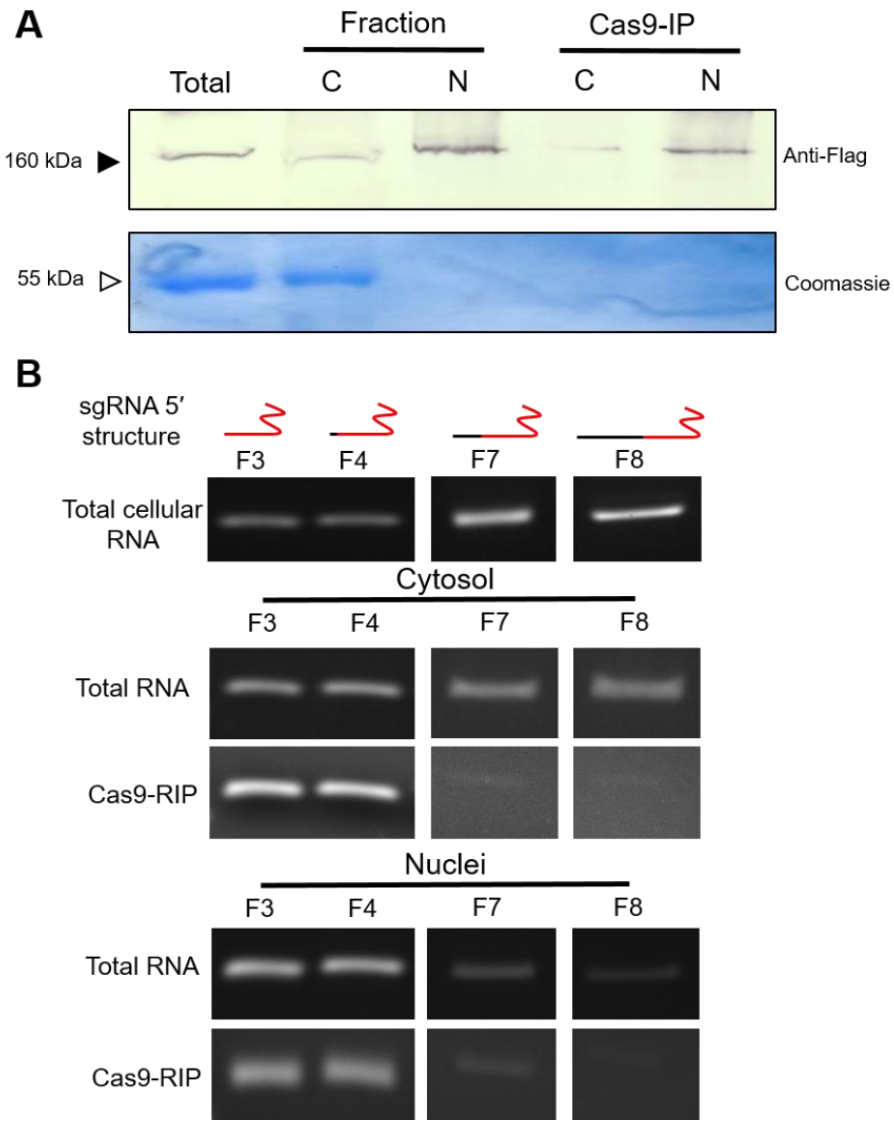


Figure 3.4. 5′ processing of sgRNA transcripts of pHcoCas9 and TRBO-G-3′gGFP co-infiltrated 16c plants upon examination of total RNA and Cas9-IP RNA in nuclear and cytosolic fractions. A) Protein lysate and Cas9 immunoprecipitations (Cas9-IP) from total lysate (Total), cytosolic fraction (C) and nuclear fraction (N) were used for detection of Cas9. The top panel is a western blot to detect Cas9 (solid arrow 160 kDa) using an anti-Flag primary antibody. The bottom panel is a Coomassie stained gel used to detect Rubisco (open arrow 55 kDa). **B)** Samples assayed for Cas9 expression in **A** were used for detection of 5′ gGFP processing through RNA extractions followed by RT-PCR using forward and reverse primer sets described in **Figure 2A**. Total lysate RNA, total cytosolic RNA, Cas9-IP cytosolic RNA, total nuclear RNA, and Cas9-IP nuclear RNA were used to detect the cellular location of 5′ sgRNA processing.

To ensure that sgRNA processing occurs in the cytosol, and to remove the possibility that Cas9-sgRNA complexes in the Cas9-IP cytoplasm fraction (**Figure 3.4B**) are not from previously localized complexes to the nucleus, we removed the nuclear localization signals (NLS) from Cas9 and constructed p-NLSCas9 (**Figure 3.5A**). Protein expression from p-NLSCas9 was verified by western blot (**Figure 3.5B**). 16c plants were then infiltrated with TRBO-G-3′gGFP as well as co-infiltrated with either the NLS lacking p-NLSCas9 construct or the NLS containing pHcoCas9 vector. To confirm a lack of nuclear localization of the p-NLSCas9 encoded protein, 7 dpi DNA was assayed for verification of indel formation following each treatment. As expected the p-NLSCas9 construct lacked the ability to produce DSBs from 16c genomic DNA (**Figure 3.5C**). Following these results, tissue was sampled from 4 dpi 16c plants and used for Cas9-IPs followed by RNA extractions as well as total lysate RNA extractions. Total RNA and Cas9 bound RNA from both NLS containing and p-NLSCas9 vector treatments was then used as a template for an RT reaction. Following RT-PCR reactions it was confirmed that sgRNA 5′ processing occurred in both the NLS lacking and NLS containing construct (**Figure 3.5D**). This data reinforces that 5′ sgRNA processing is occurring in the cytosol *in planta*.

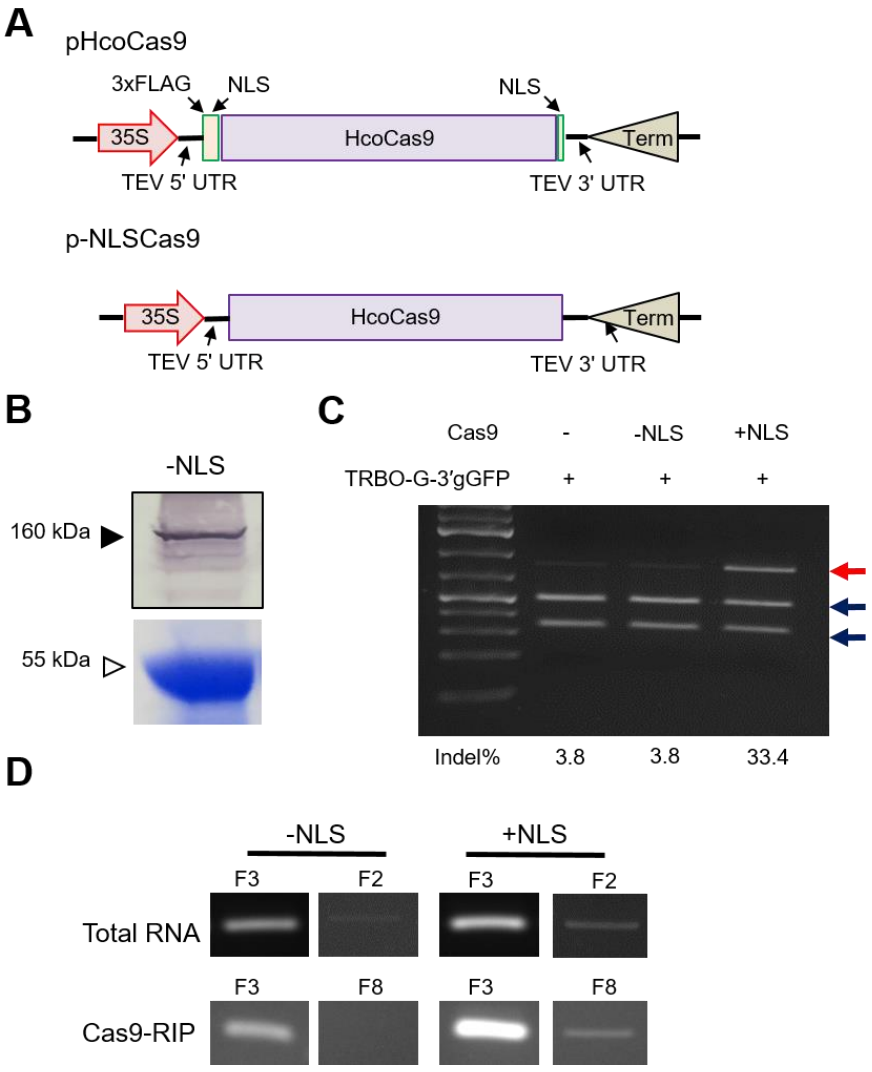


Figure 3.5. Assays of 5′ sgRNA transcript processing of nuclear localized and non-nuclear localized Cas9. A) Cas9 expression constructs containing a nuclear localization signal (pHcoCas9) and without an NLS (p-NLSCas9) that were used for agroinfiltration of 16c plants both. The construct p-NLSCas9 does not contain the two NLS and the 3xFLAG epitope tag, in comparison to pHcoCas9. **B)** Cas9 expression detection from p-NLSCas9 construct using a western blot probed with anti-Cas9 antibody in the upper panel. The lower panel is a Coomassie stained gel used for Rubisco detection. **C)** *Bsg*I digest of samples collected at 7 dpi used to detect indels from 16c plant tissue infiltrated with TRBO-G-3′gGFP and no Cas9 construct, p-NLSCas9 or pHcoCas9. Arrows to the right of the gel image depict indel containing bands (red) and wt *mgfp5* amplified DNA (blue). **D)** RT-PCR analysis of 5′ sgRNA transcript processing in 16c plant tissue infiltrated with TRBO-G-3′gGFP and either p-NLSCas9 or pHcoCas9. Total RNA from total cellular lysate and Cas9-IP RNA from Cas9 constructs either with NLS (+NLS) or without NLS (-NLS). Primer sets are as described in **Figure 3.2B**.

Nuclear transcribed protein-sgRNA fusion transcripts create catalytically active Cas9 complexes

Next we wanted to understand if the 5′ sgRNA processing events are in some way associated with a response to virus infection, or due to TRBO (viral) replication and gene expression in the cytosol. If cytosolic localization or the virus infection response is responsible for sgRNA 5′ processing events, then nuclear transcribed transcripts carrying nucleotide overhangs should not create catalytically active Cas9-sgRNA complexes. To determine if host viral responses or cytosolic localization are responsible for processing events, we used the GFP-gGFP fusion transcript, which is delivered by the second subgenomic RNA in the TRBO-G-3′gGFP construct, as a template and created a new construct where transcription was driven by the *Arabidopsis thaliana* Pol III U6 nuclear promoter. The U6 promoter based expression of sgRNAs should localize expression of transcripts to the nucleus, which is potentially the reason other researchers initially used the U6 promoter to deliver sgRNAs *in vivo* in some of the first CRISPR/Cas9 assays. However, in this case we will use the nuclear localization of transcripts produced from the U6 promoter to discern the effect overhangs and specifically nuclear sgRNA overhangs has on indel percentages which occurs upon successful host non-homologous end joining (NHEJ) double stranded break repair from a catalytically competent Cas9-sgRNA complex.

To separate cytosolic transcript expression and potential viral host responses, the protein-sgRNA fusion transcript, U6-GFP-gGFP, was constructed along with a transcript producing only "clean" (no 5′ nucleotide overhangs) sgRNA, U6-gGFP, to serve as the positive control. Both U6-GFP-gGFP and U6-gGFP were inserted into the pHcoCas9 expression vector to produce pHco-U6-GFP-gGFP and pHco-U6-gGFP, respectively (**Figure 3.6A**). 16c plants were used for

half-leaf assays using pHco-U6-GFP-gGFP and pHco-U6-gGFP to test for *in planta* catalytic activity (**Figure 3.6B**). Tissue was taken at 7 dpi from three assayed plant samples and subjected to PCR amplification followed by a *Bsg*I digest. Leaves infiltrated with pHco-U6-GFP-gGFP surprisingly showed a high quantity of indels 17%-30%, while pHco-U6-gGFP was considerably higher at 33%-40% (**Figure 3.6C**). Each half-leaf assay consistently measured lower percentages of indels when infiltrated with pHco-U6-GFP-gGFP as compared to the pHco-U6-gGFP infiltrated portion of the leaf (**Figure 3.6C**). One possible explanation for the lower indel percentages using the pHco-U6-GFP-gGFP construct would be the length of the transcript (~850nts) which is much longer than a typical Pol III transcribed RNA (100-150nts). Thus, gGFP expression could be lower in the pHco-U6-GFP-gGFP due to decreased Pol III fidelity at the 3′ end of the transcript. To test if the discrepancy of indel mutation percentages between these two constructs was due to lower expression levels of the pHco-U6-GFP-gGFP transcripts or due to 5′ sgRNA overhangs impairing catalytic activity, 5 dpi half-leaf assays were used for RT-PCR expression analysis (**Figure 3.6D**). Ultimately, there was no difference in expression levels of gGFP between pHco-U6-GFP-gGFP and pHco-U6-gGFP, indicating that the lower indel percentages from the pHco-U6-GFP-gGFP is due to a reduction in 5′ processing efficiency in host cells. One explanation for this phenomenon is due to the nuclear localization of transcripts synthesized from U6 promoters, while we have demonstrated that 5′ processing events occur in the cytosol. Perhaps more importantly, these assays demonstrate that, in fact, pHco-U6-GFP-gGFP is capable of delivering sgRNAs with considerable 5′ overhangs that are clearly capable of producing indels in the presence of Cas9. These findings contradict the assumptions currently made in the literature on sgRNA delivery specificity in regards to sgRNA nucleotide overhangs (Cermak et al. 2017; Xie et al. 2015; Tsai et al. 2014).

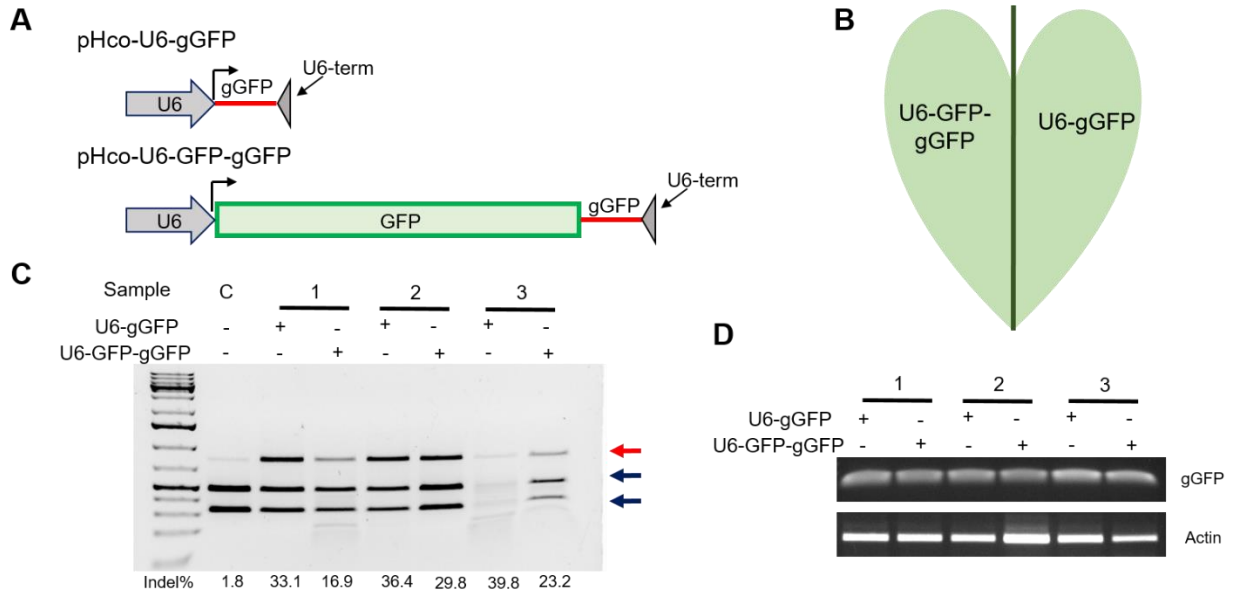


Figure 3.6. U6 nuclear promoter transcription of a sgRNA with a 5′ overhang *in planta*. A) Two sgRNA in planta delivery constructs cloned into pHcoCas9. pHco-U6-gGFP contains a polymerase III U6 promoter followed by the gGFP sgRNA and a U6 transcription terminator. The pHco-U6-GFP-gGFP construct is same as pHco-U6-gGFP but contains the GFP protein coding region directly 5′ to the gGFP sequence and 3′ to the U6 promoter. **B)** Depiction of the experimental setup of half-leaf assays used. Constructs shown in A were agroinfiltrated into one side of the leaf with the leaves midrib (dark green line) serving to separate the treatments. **C)** *Bsg*I digest from mgfp5 amplified PCR products of three replicates (Sample 1-3) half-leaf assays depicted in **B** and sampled at 7 dpi. The negative control (C) represents pHcoCas9 infiltrated 16c plants. The red arrow indicates the *Bsg*I resistant bands containing indels and the blue arrows indicate digested, or wt *mgfp5* sequences. Indel percentages quantified using ImageJ image analysis software for each treatment are shown under the corresponding lane. **D)** RT-PCR analysis of half-leaf assays depicted in B used to compare the expression levels of gGFP from both the pHco-U6-gGFP and pHco-U6-GFP-gGFP. RNA was extracted at 7 dpi from the same leaves used in **C**. The top panel, gGFP, primers specifically amplifying gGFP expression were used. The bottom panel, Actin, primers specifically amplifying *N. benthamiana* Actin expression was used as the loading control.

Targeting of *N. benthamiana* ribonuclease pathways using biochemical inhibitors

Since 5′ processing of the sgRNAs occurs in the cytosol, we next attempted to inhibit two cytosolic ribonuclease pathways that could be responsible for the RNA cleavage events leading to the activation of Cas9-sgRNA complexes. One possible pathway responsible for this occurrence being the 5′ to 3′ exoribonuclease RNA degradation, which is carried out by a family of conserved proteins, XRN. In this scenario the exoribonuclease would sheer sgRNA 5′ nucleotide overhangs flush with the 5′-most nucleotide of the sgRNA bound to Cas9 (**Figure 3.7A**). The RNA silencing pathway is another possible mechanism responsible for the 5′ sgRNA processing event. This pathway creates site-specific endoribonucleases through sequence homology with an "invading" RNA. Considering RNA silencing is a conserved host response to control plant virus infection (Alvarado and Scholthof 2009) and is also a mechanism which has been shown to decrease the expression of transgenes (Saxena et al. 2011). In this scenario, the nuclease would target transcripts containing large sgRNA 5′ overhangs, such as those produced by either TRBO-G-3′gGFP or pHco-U6-GFP-gGFP (**Figure 3.7B**). Due to the complicated genetics (allotetraploid) of *N. benthamiana*, and the overall diversity and functional complementarity of the *xrn* gene family and RNA silencing associated gene families we elected to avoid traditional reverse genetics approaches, such as a viral induced gene silencing (VIGS) screen. Instead we inhibited the respective pathways by co-delivery of biochemical inhibitors or suppressors to specifically target each pathway.

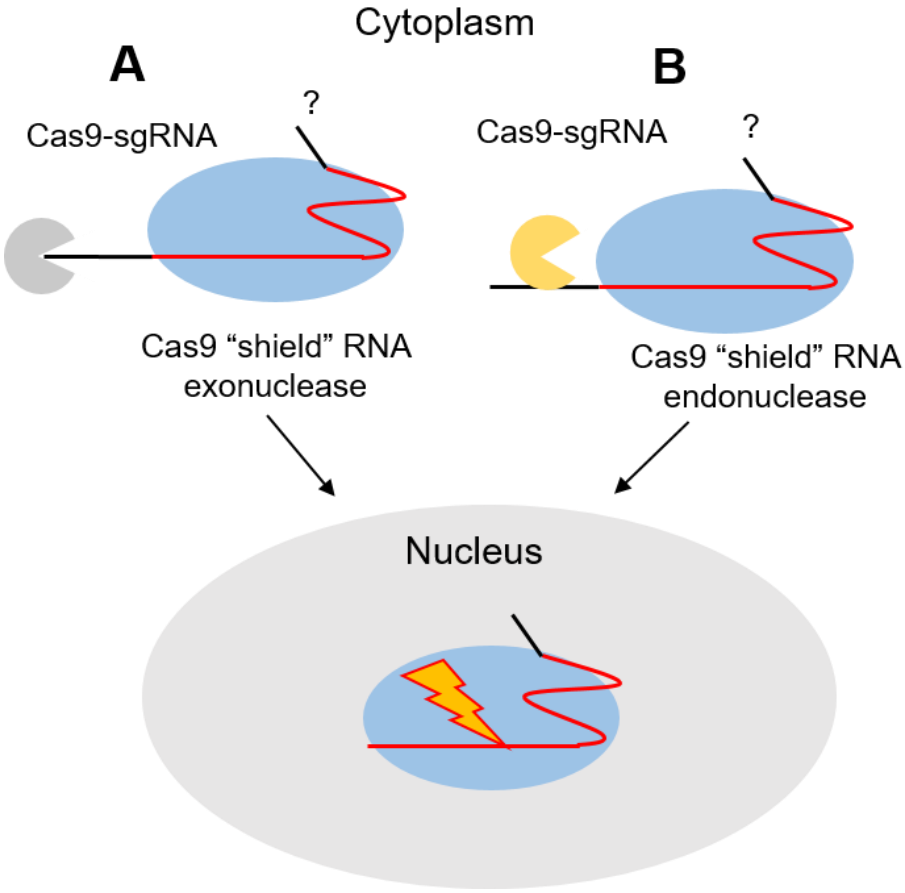


Figure 3.7. *In planta* 5′ sgRNA processing model. Cas9 binds to sgRNAs containing 5′ overhangs in the cytoplasm to create a Cas9-sgRNA that is not catalytically active. The Cas9-sgRNA complex is then processed by one of two catalytic events that occurs to remove or process the 5′ end of the sgRNA to make a catalytically active complex. **A**) An RNA exonuclease is responsible for 5′ to 3′ degradation of nucleotide sequences which are not shielded by Cas9. **B**) An RNA endonuclease cleaves the 5′ nucleotide overhang while the sgRNA is shielded by Cas9. In either scenario, **A** or **B**, Cas9 acts as a "shield" to protect the integrity of the spacer sequence of the sgRNA from degradation.

In order to test if the *xrn* degradation pathway is responsible for 5′ processing of sgRNAs *in vivo*, we developed an expression vector that selectively inhibited the ability of the pathway to degrade RNAs. Previously it was shown that a 152 nt RNA segment from Zika virus (subgenomic flaviviral RNA or sfRNA) serves as a barrier against the human XRN1 protein based degradation *in vitro* and *in vivo* due to the secondary RNA structure which shields the RNA from degradation (Akiyama et al. 2016). To inhibit XRN processing *in planta* the sfRNA sequence was inserted 5′ proximal to the gGFP sequence in pHco-U6-GFP-gGFP to create pHco-U6-GFP-sfRNA-gGFP (**Figure 3.8A**). Three replicates were conducted on 16c plants with one half leaf infiltrated with pHco-U6-GFP-sfRNA-gGFP and the other with pHco-U6-GFP-gGFP (**Figure 3.8B**). At 7dpi, leaf tissue was then sampled and mgfp5 was amplified and tested for indels using a *Bsg*I digest. Indel percentages were calculated for each treatment with no significant difference between the pHco-U6-GFP-sfRNA-gGFP and pHco-U6-GFP-gGFP constructs (**Figure 3.8C**). Considering the indel percentages were indistinguishable between these two treatments it cannot be concluded that XRN proteins are involved in sgRNA processing. Rather, these results indicate one of the following: 1) XRN proteins are not involved in the processing of sgRNAs, 2) sfRNAs do not retain their function in plants or plant XRN genes are not susceptible to sfRNAs, or 3) there are compounding ribonuclease pathways involved in the processing events.

We next devised a method to dissect the RNA silencing pathways possible involvement in the removal of 5′ sgRNA overhangs *in planta*. For this, half-leaf agroinfiltration assays of 16c plants were performed using pHco-U6-GFP-gGFP construct was used with and without the RNA silencing suppressor P19 (Saxena et al. 2011) (**Figure 3.8D**). 7dpi leaf tissue was sampled and tested for indels using a *Bsg*I digest on three replicate treatments. Image analysis revealed that

while the indel percentages between individual plants varied significantly, the addition of P19 consistently decreased the overall percentage of indel rates by 3.5-13.2% (**Figure 3.8E**). Indeed sample 1 had considerably higher indel percentages (45.1-52.6%) as compared to the other two samples (14-27.2%) (**Figure 3.8E**). However, in general there is a decreased indel percentage for pHco-U6-GFP-gGFP and P19 co-infiltrated tissue compared to the singularly delivered pHco-U6-GFP-gGFP. Perhaps the differences noted between the samples is more of a testament to the importance of running half-leaf assays to reduce any confounding results due to physiological differences between individuals or within different tissues of the same individual, even in near isogenic lines. In either case, it would be expected that if the RNA silencing pathway was solely responsible for 5′ sgRNA processing that the overall percentage of indels would be closer to the pHcoCas9 infiltrated negative control (~0%). However, these experiments do not eliminate the possibility that the RNA silencing pathway is involved in the processing of Cas9 bound sgRNAs, but rather hint toward a mechanism were multiple ribonuclease pathways being responsible for processing events in *N. benthamiana*.

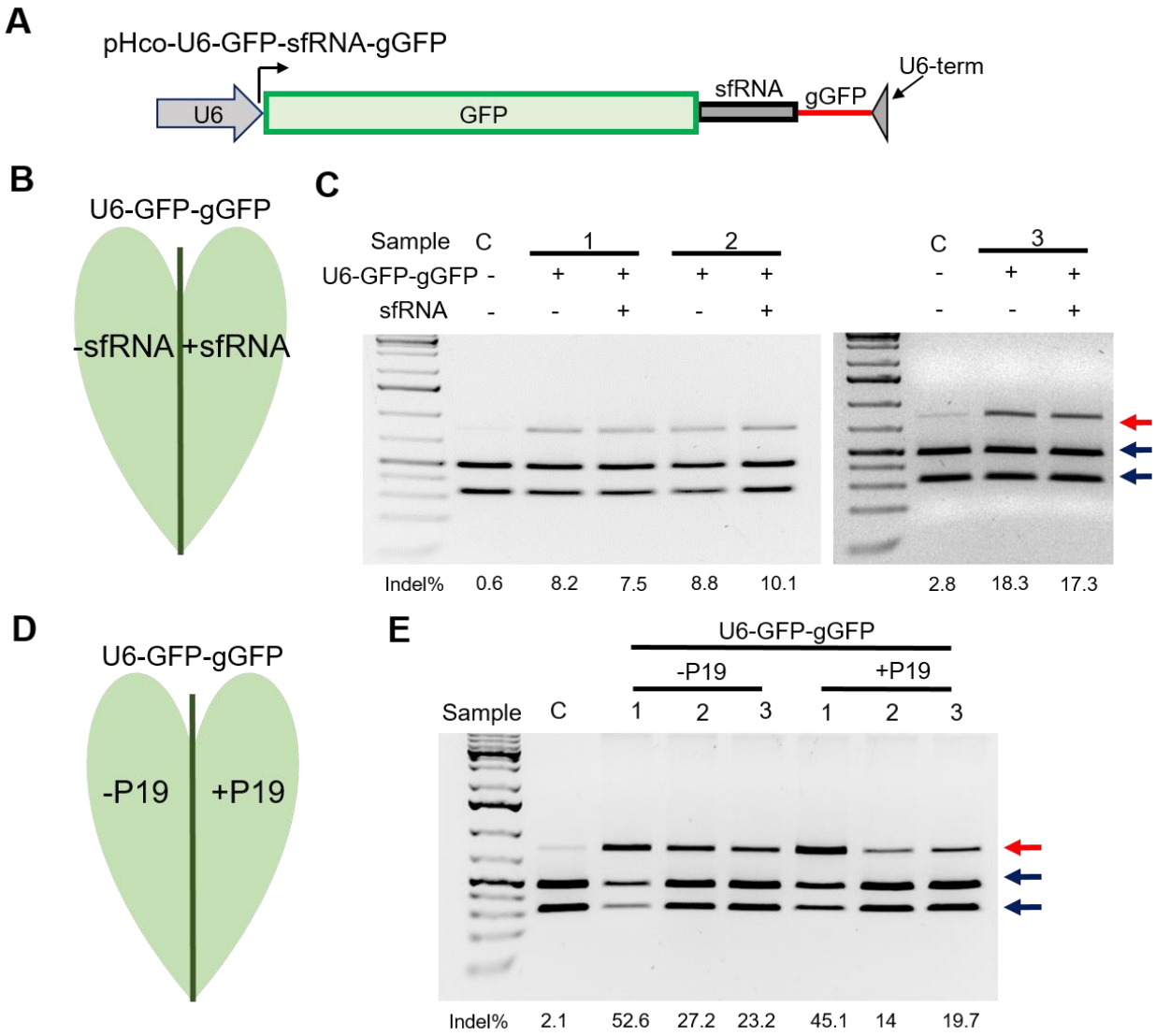


Figure 3.8. Inhibiting *N. benthamiana* 5′ to 3′ exonuclease and RNA silencing pathways. **A)** Depiction of the gGFP expression construct pHco-U6-GFP-sfRNA-gGFP which contains the XRN1 resistant sfRNA sequence 5′ proximal to the gGFP sequence and 3′ proximal to the GFP coding sequence. **B)** Half-leaf assays used for analysis of pHco-U6-GFP-gGFP with and without the sfRNA sequence. For both **C** and **E** *Bsg*I digest analysis was performed on *mgfp5* amplicons where the red arrow indicates indel containing bands and the blue arrows represent DNA sequences that were not modified. Lanes containing C serve as the *Bsg*I digest positive (non-edited) control and were harvested from tissue that was infiltrated only with pHcoCas9. Indel percentages for each sample and treatment are indicated below the corresponding lane. **C)** Half-leaf assays harvested from 16c plants 7 dpi following the treatment depicted in **B**. **D)** Experimental set up of half-leaf assays infiltrated with pHco-U6-GFP-gGFP and both with and without the RNA silencing suppressor P19. **E)** Half-leaf assays harvested from 16c plants 7 dpi following the treatment depicted in **D**.

Silencing the decapping enzyme DCP2 increases gene editing events in *N. benthamiana*

Previously we constructed the pHco-U6-GFP-sfRNA-gGFP expression system in an attempt to use the XRN degradation resistant sfRNA to inhibit the ability the Cas9-sgRNA complex to create DSBs. This method was not successful for interpretation of the XRN degradation pathways potential involvement in sgRNA processing due to three possible explanations: 1) XRN proteins are not involved in the processing of sgRNAs, 2) sfRNAs do not retain their function in plants or, 3) there are multiple ribonuclease pathways involved in the processing events. By examining the typical 5′ to 3′ exoribonuclease pathway for degradation of mRNAs, in eukaryotic organisms, we discovered a diverse set of XRN proteins in our experimental model *N. benthamiana*. However, considering XRN proteins require a single phosphate group on the 5′-most end of the transcript, we found a potential bottleneck of the pathway in that XRN proteins cannot degrade mRNA unless the 5′-ends of transcripts have the cap structure removed by a decapping complex. The catalytic component of the decapping complex, DCP2, (Xu et al. 2006) from *A. thaliana* contains homologs in *N. benthamiana*, and while *N. benthamiana* contains two copies of the gene (Niben101Scf01105g01020.1 and Niben101Scf26315g00002.1), both can be targeted simultaneously using VIGS. While the sfRNA construct did not allow for us to conclusively determine that the 5′ to 3′ exonuclease pathway of mRNA turnover, by knocking down DCP2 we could potentially determine if XRN proteins are involved in sgRNA processing due to the proteins dependence on a monophosphate group on the 5′ end of the RNA (Jinek et al. 2011).

To test if inhibiting the decapping pathway would inhibit the ability of capped sgRNA transcripts to induce DSBs, a 400 nt region aligning with both of the *N. benthamiana* DCP2 transcripts was amplified from cDNA and cloned into the TRV VIGS vector to create TRV-

DCP2. Since the U6 promoter does not produce capped transcripts, it was necessary to move back into our initial viral-sgRNA delivery platform, TRBO-G-3′gGFP, which produces transcripts that contain 5′ cap structures. Three week old 16c plants were infiltrated with the empty TRV vector control (TRV-00), TRV-DCP2, or a *phytoene desaturase 3* silencing vector (TRV-PDS3), used as the systemic silencing experimental control. At 17 dpi plants demonstrated systemic silencing on the new growth for TRV-PDS3 (**Figure 3.9A**). However, there was not noticeable difference in phenotype observed for TRV-00 and TRV-DCP2 treated plants (**Figure 3.9A**). Then, 3-4 leaves were selected at the apical portion of four TRV-00 and TRV-DCP2 infected plants for co-infiltration of pHcoCas9 and TRBO-G-3′gGFP. At 6 dpi leaf tissue of co-infiltrated pHcoCas9 and TRBO-G-3′gGFP was sampled and DNA extracted from the replicates of TRV-00 and TRV-DCP2 treated plants. Following a PCR amplification of *mgfp5, Bsg*I digests were performed on amplicons and indel containing bands calculated. TRV-00 treated plants consistently had lower indel percentages calculated (13.8%-31.6%) while TRV-DCP2 treated plants contained much higher indel values (29.8%-42.2%) (**Figure 3.9B**). This is a rather shocking result considering the expected results would be that the TRV-DCP2 treated plants would have much lower indel percentages than TRV-00 if the pathway is involved in sgRNA processing. If the DCP2 genes are not involved you would expect comparable indel percentages between TRV-00 and TRV-DCP2 plants. However, these analysis do not take into consideration the possible effect that silencing a decapping enzyme has on a capped virus, such as TMV, and its replication efficiency. Indeed, it has been reported that silencing of decapping enzymes increases viral replication (Ma et al. 2015). However this assay does not consider the possibility of multiple ribonuclease pathways responsibility in the processing of sgRNAs. If the XRN family of proteins is involved in the sgRNA processing, which we do not demonstrate in these

assays, there would certainly be other ribonucleases involved. With this in mind, it would not be responsible to eliminate the possibility of the 5′ to 3′ exoribonuclease family of proteins from the potential of processing 5′-ends of sgRNAs without further exploration. One potential explanation is that another pathway that is not sensitive to 5′ cap structures, such as RNA silencing endoribonucleases reported earlier, could be acting upstream of *xrn* genes involvement. Thus, RNA silencing associated genes could be responsible for producing the monophosphate group at the 5′-end of RNAs.

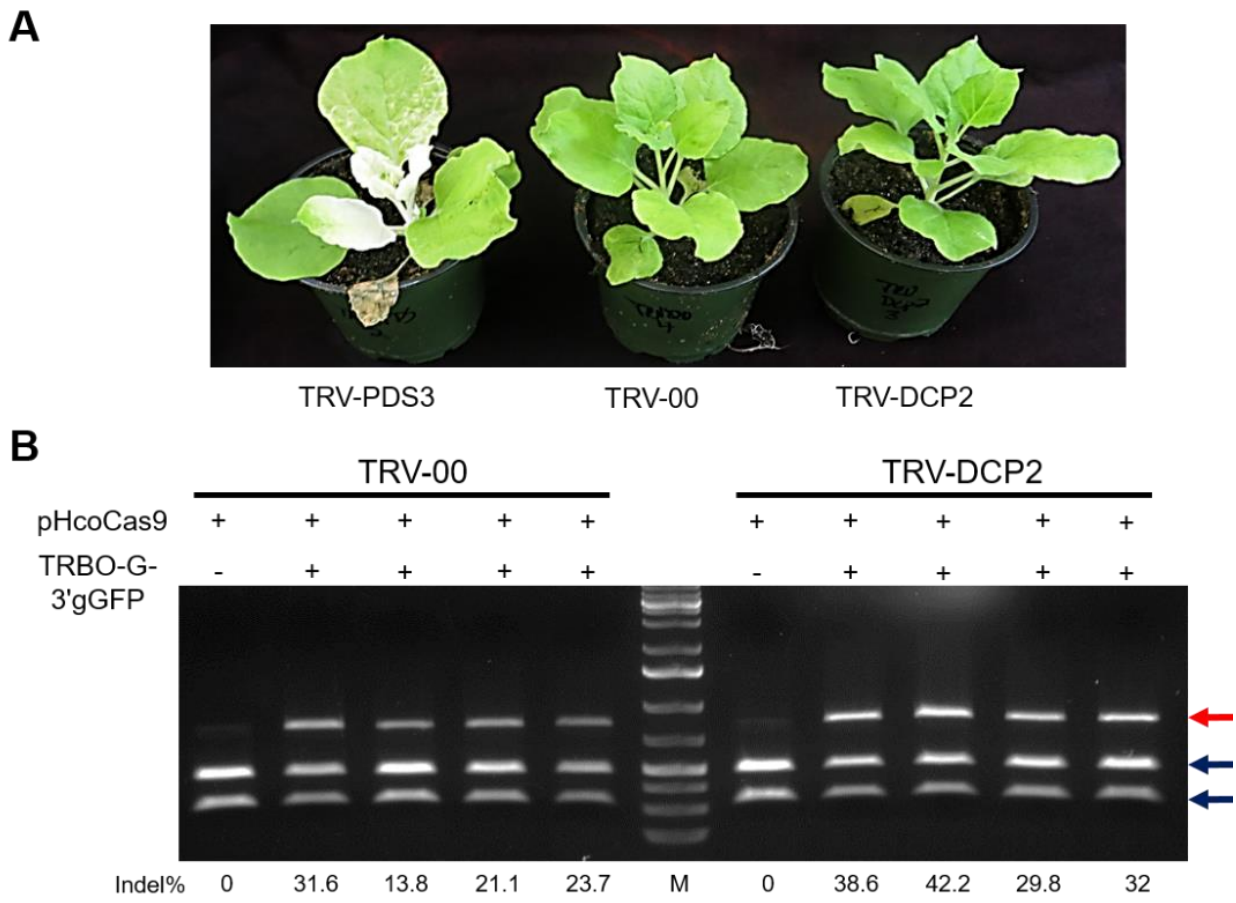


Figure 3.9. VIGS of the decapping enzyme DCP2 which is the catalytic portion of the decapping complex. A) 16c plants silenced for PDS, DCP2, and an empty vector control (TRV-00) using a TRV VIGS screen. Images are taken 17 days after TRV-00 and TRV-DCP2 were infiltrated with pHcoCas9 and TRBO-3′gGFP. TRV-PDS3 served as the screen's positive control for systemic silencing. **B)** Systemic tissue of TRV-DCP2 and empty vector TRV-00 treated plants were used for co-infiltrations of pHcoCas9 and TRBO-G-3′gGFP. At 6 days after co-infiltration of pHcoCas9 and TRBO-G-3′gGFP leaves were used for DNA extraction followed by PCR amplification of *mgfp5* and treated with *Bsg*I enzyme. *BsgI* resistant bands (indels) are indicated by the red arrow and wt *mgfp5* digested bands are signified by the blue arrows to the right of the gel. Indels are calculated for each treatment and are indicated at the bottom of the panel. pHcoCas9 infiltrated leaves serve as the negative control to ensure a complete digestion reaction.

DISCUSSION

While studies have been conducted analyzing the 3′ processing of CRISPR RNAs and (crRNAs) and trans-activating CRISPR RNAs (tracrRNAs) in both native and non-native CRISPR systems by RNAse III enzymes (Deltcheva et al. 2011; Karvelis et al. 2013), these studies have not yet discovered the mechanism of 5′ processing of crRNA/tracrRNAs or sgRNAs (van der Oost et al. 2014). While it was highlighted that a secondary processing step focused on 5′ overhang removal must be taking place (van der Oost et al. 2014; Deltcheva et al. 2011) there has not been much insight into the mechanism for sgRNA processing, and certainly no implication that it could be conserved across phylogenetic kingdoms. Furthermore, there has been a considerable amount of focus on nucleotide mismatches in the 20 nt complementary spacer sequence of sgRNAs and the loss of catalytic capabilities of the subsequent Cas9-sgRNA complexes on protospacer sequences (Jinek et al. 2012; Zheng et al. 2017). However, the interrogation of specifically extraneous 5′ sgRNA nucleotides and the biological effect on complex cleavage has not been examined to date.

In this study, we systematically examined the effect of 5′ and 3′ overhangs on Cas9-sgRNA complexes using *in vitro* assays, and determined that specifically overhangs 5′ proximal to the sgRNA sequence inhibited the catalysis of protospacer carrying DNA (the target for editing). Following these results, we hypothesized that for our previously developed TRBO-G-3′gGFP construct to be catalytically active *in planta*, processing of nucleotides 5′ proximal to the sgRNA sequence must occur. We then determined that transcripts produced from TRBO-G-3′gGFP contained sgRNA that were processed at the 5′-ends (5′ nucleotides not corresponding to the sgRNA were removed) when bound by Cas9 in leaf tissue infiltrated with both pHcoCas9 and TRBO-G-3′gGFP. Additional experimentation localized the site of the Cas9 bound sgRNA

processing events to the cytosol. We next inquired if the sgRNA processing events were based on the cytosolic expression of TRBO subgenomic RNAs and if this was a viral dependent RNA processing event. The development of U6 promoter driven expression of the protein-sgRNA fusion transcript, GFP-gGFP, corroborates that cytosolic expression is optimal, however, possibly not necessary for sgRNA transcript processing events, and that processing events are not unique to TRBO delivery of sgRNAs. In an attempt to identify the ribonuclease pathway responsible for *in vivo* "activation" of Cas9-sgRNA complexes by sgRNA 5′ processing, we elected to inactivate the RNA silencing pathway using the suppressor P19, and also inhibit the 5′ to 3′ exoribonuclease gene family, *xrn*, through the insertion of the XRN1 resistant sfRNA sequence directly upstream of a sgRNA sequence. Co-delivery of the nuclear transcribed protein-sgRNA fusion transcript with P19 decreased in indel percentages compared to the control without P19, which indicates a link between the RNA silencing pathway and 5′ sgRNA processing events. Targeting of the 5′ to 3′ exoribonuclease RNA degradation pathway using the sfRNA-containing construct as well as *dcp2*-silenced plants did not lead to a reduction in indel events, which would be expected from the production of catalytically inactive Cas9-sgRNA complexes. However, these experiments do not conclusively determine that the 5′ to 3′ exoribonuclease pathway is not involved in the removal of 5′ overhang nucleotides. Rather, it appears more likely that there are multiple ribonuclease pathways involved in the processing and subsequent activation of Cas9-sgRNA complexes *in planta*.

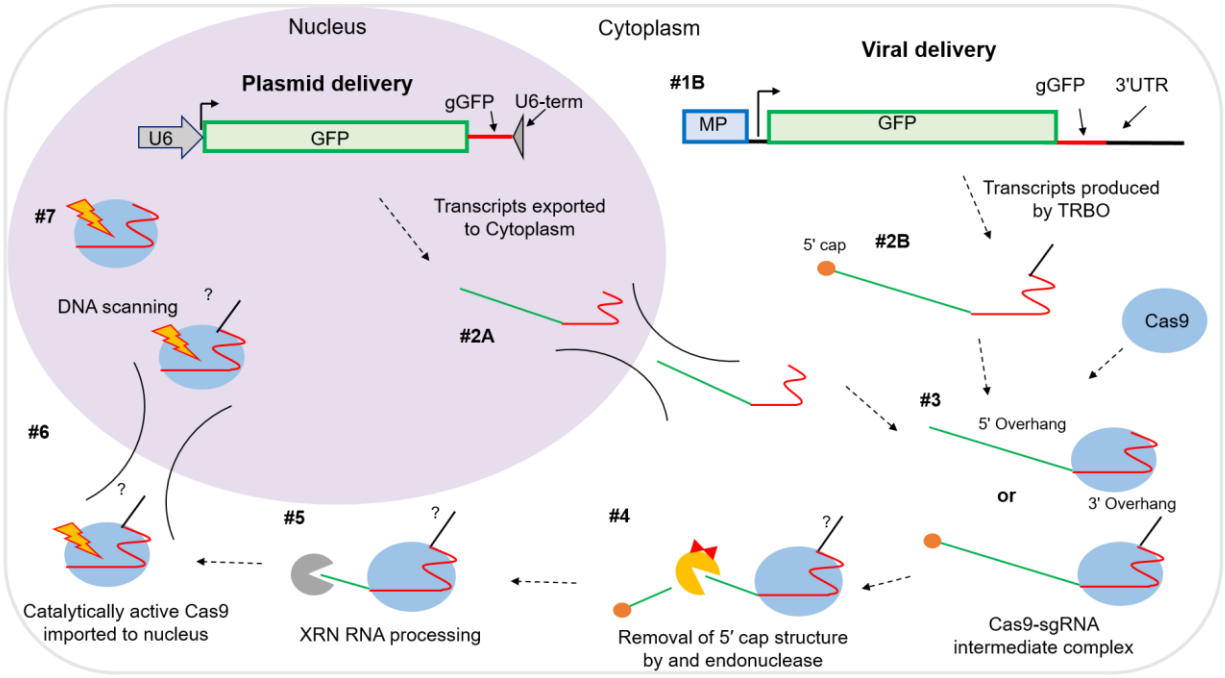


Figure 3.10. A model for 5′ sgRNA processing *in planta*. Two independent sgRNA expression systems based on nuclear transcription (**1A**) and cytosolic viral expression (**1B**). Following nuclear transcription (**1A**) transcripts are exported from the nucleolus to the cytosol (**2A**). In contrast, viral delivery of sgRNA expression (**1B**) is carried out through the production of subgenomic RNAs by the viral replicase protein (**2B**). In **2A** and **2B** both transcripts contain excess nucleotides 5′ proximal to the sgRNA sequence, but in this case **2B** also contains a 3′ sgRNA proximal overhang sequence and also a 5′ cap. In either case, Cas9 protein recognition of the sgRNA scaffolding sequence allows for binding of the Cas9 and sgRNA creating a temporary complex that contains the entire sgRNA containing transcript sequence (**3**). The 5′ cap structure is removed by either an unknown endoribonuclease or the decapping complex (**4**). Following cap removal, sgRNAs are processed by a cellular 5′ to 3′ exoribonuclease until it reaches the Cas9-sgRNA complex where Cas9 serves as a shield and protects the sgRNA from being degraded (**5**). The exonuclease activity serves as mechanism for processing the sgRNA to create a catalytically active Cas9-sgRNA complex (**6**). The new catalytically active complex is then imported into the nuclease (**7**) where it can begin to scan the genomic DNA for matching protospacer sequences (**8**).

In **Figure 3.10** we suggest a model for 5′ sgRNA processing for both nuclear transcribed and cytosolic transcribed (viral) RNAs in *N. benthamiana* based on our experimental results. For nuclear and cytosolic expressed transcripts that contain 5′ sequences that do not correspond to the sgRNA sequence, cytosolic localization is critical for processing after being bound by Cas9. In order for sgRNA transcripts to be processed to the correct length Cas9 binding is paramount, as also suggested by others (Mikami et al. 2017). Altogether, our data suggests that Cas9 binding is important for proper processing (**Figure 3.2 C** and **D**). This is in agreement with the structural analysis of the Cas9-sgRNA complex (Nishimasu et al. 2014; Anders et al. 2014), which demonstrated that the 5′-most end of the sgRNA sequence is harbored within the active site of Cas9. In essence, the inclusion of the sgRNA sequence within the protein would protect the RNA from degradation by host ribonucleases. This leads to our prediction that Cas9 "shields" the sgRNA sequence from degradation by exo- or endoribonucleases (**Figure 3.7 A** and **B**). However, upon further examination of the RNA silencing pathway as well as the 5′ to 3′ exonuclease pathways effects on the creation of catalytically active Cas9-sgRNA complexes, it appears that an amalgamation of RNA degrading pathways is responsible for these processing events.

The most likely scenario for the processing events, at least using our transcriptional models (TRBO-G-3′gGFP and pHco-U6-GFP-gGFP), is an initial RNA cleavage event by an endoribonuclease which would provide two essential steps for further processing by exoribonucleases: 1) removal of the 5′ cap structure and exposure of a 5′ monophosphate group which is necessary for exoribonuclease activity (Jinek et al. 2011); and, 2) shortening of the sgRNA molecule to be processed further by XRN proteins. The importance of at least one endoribonuclease pathway, post-transcriptional RNA silencing, was demonstrated here, but this

certainly does not eliminate the possibility of other pathways involvement in processing, rather, it suggests the opposite. The importance of the exoribonuclease degradation of sgRNAs, was only supported by insinuations, primarily based on structural evidence (Nishimasu et al. 2014; Anders et al. 2014). The possibility of endoribonucleases processing the sgRNAs to the specificity seen in **Figure 3.2D** would seem highly unlikely. However, a 5′ to 3′ exoribonuclease progressively cleaving towards to the sgRNA-Cas9 binding site can easily be envisioned.

The translational impact of this report reaches beyond basic CRISPR biology or plant biology. It has been shown previously that there are native mechanisms for processing CRISPR arrays in human (Cong et al. 2013) and plant cells (Mikami et al. 2017; Cody et al. 2017). However, these findings seem to have been overlooked in a multitude of studies which has led to extravagant engineering of *in vivo* sgRNA delivery platforms; these have perhaps been developed based on the false premise that, outside of the native *S. pyogenes* system, CRISPR-Cas9 delivery must be supplemented with specialized sgRNA delivery tools (Xie et al. 2015; Tsai et al. 2014; Gao and Zhao 2014; Cermak et al. 2017). Perhaps one explanation for this oversight is the inherent focus on the 3′ processing of crRNAs by the RNase III enzyme in bacteria (Deltcheva et al. 2011; Sapranauskas et al. 2011) and human cells (Cong et al. 2013; Hsu et al. 2014) that appear to have a functional overlap among prokaryotes and eukaryotes. However, similar discussions about the secondary processing step of 5′ ends of crRNAs or sgRNAs *in vivo* are not evident. Nevertheless, further evidence found in the native CRISPR type II-C system where the crRNA synthesis system demonstrates the dispensable nature of the RNase III enzyme for creating catalytic Cas9-crRNA/tracrRNA complexes in *Neisseria meningitides* (Zhang et al. 2013). The ability of *N. meningitides* and *Campylobacter jejuni* to produce crRNAs that do not need 5′ processing based on the promoter specificity of the array,

we believe, further support the importance of the 5′ processing step for the catalytic or interference activity of the complex (Dugar et al. 2013). Indeed, RNase III serves to separate individual crRNA/tracrRNA duplexes from others on a single transcript carrying multiple crRNAs, but to be catalytically "activated" processing of the crRNA 5′ end must occur. The relevance of RNase III on activating complexes is purely circumstantial, solely based on the position of the crRNA on the overall crRNA transcriptional array. For example, the first crRNA in a crRNA transcriptional array can still be catalytically active in the absence of an RNase III enzyme, but the downstream crRNAs cannot. In either case, the importance of 5′ modifications appears to have large implications on catalytic events, as shown in this study, but also demonstrated in the native CRISPR II system (Zetsche et al. 2015). Perhaps the excitement of the potential uses of CRISPR systems has exceeded our own knowledge of its basic biology.

In summary, as modeled in **Figure 3.10**, our results demonstrate in either viral or nuclear promoter based transcriptional events, Cas9 bound sgRNA transcripts 5′-ends are processed in the cytosol. Upon being exported or produced in the cytosol, sgRNAs are bound by Cas9 which serves as a "shield" to the sgRNA against host ribonucleases. After binding by Cas9, exposed nucleotides on the sgRNA transcript are degraded by a combination of endo- and exoribonucleases creating a catalytically competent Cas9-sgRNA complex. Following "activation", the Cas9-sgRNA complex is imported into the nucleus and begins "scanning" DNA sequences for a sgRNA complementary sequence.

MATERIALS AND METHODS

Cloning and construct development

The constructs pHcoCas9 and TRBO-G-3′gGFP were described in detail previously (Cody et al. 2017). The p-NLSCas9 plasmid was constructed using the human codon-optimized Cas9 nuclease (HcoCas9) (Addgene: 42230) (Cong et al. 2013) as a template. For this, the Cas9 encoding sequence without the NLS was amplified using a forward primer designed downstream of the NLS sequence with a *Bam*HI site and a start codon (ATG). The reverse primer was designed upstream of the C-terminal NLS sequence followed by a stop codon (TAA) and an *Xho*I site. The PCR amplicon was cloned into a modified pRTL22 (Restrepo et al. 1990) sub-cloning vector, as described by Cody et al. (2017). The 35S-Cas9-term cassette was then transferred into the binary destination plasmid pBINPLUS-sel using the *Hin*dIII site to create p-NLSCas9.

pHco-U6-gGFP and pHco-U6-GFP-gGFP were constructed using the pChimera subcloning vector from Fauser et al. (2014) as a template for amplifying the U6 promoter and terminator for Gibson assembly into *Pac*I linearized pHcoCas9. TRBO-G-3′gGFP was used to amplify gGFP for pHco-U6-gGFP and GFP-gGFP for pHco-U6-GFP-gGFP. For construction of pHco-U6-GFP-sfRNA-gGFP, pHco-U6-GFP-gGFP was used as the destination vector for Gibson assembly. Following the GFP stop codon and before gGFP, sequence primers were designed and used to amplify the plasmid. The sfRNA sequence was synthesized (Genscript) and used as a template for amplification using primers that overlapped both the sfRNA sequence and the pHco-U6-GFP-gGFP linearized backbone. Sequences for these constructs can be seen in **A-8**.

TRV-DCP2 was constructed by linearizing TRV-00 (Ratcliff et al. 2001) with *Sma*I. The VIGS best prediction tool (Fernandez-Pozo et al. 2015) was used to designa a 400 nt segment

overlapping the two *dcp2* transcripts, *dcp2-1* (Niben101Scf01105g01020.1), and *dcp2-2* (Niben101Scf26315g00002.1) (**A-8**). Using cDNA from a wt *N. benthamiana* plant, primers corresponding to the *dcp2* genes were constructed in the negative complement orientation along with overhangs corresponding to the overlapping sequence of TRV-00 for Gibson assembly.

Cas9/sgRNA *in vitro* cleavage assays

TRBO-G-3′gGFP RNA templates containing either 5′ and 3′, 5′ or 3′, or non-overhang nucleotides flanking gGFP templates were synthesized using T7 RNA synthesis (New England BioLabs). T7 RNAs were synthesized using 150 ng of each PCR template amplified from TRBO-G-3′gGFP. Forward primers T7-F1 and T7-F2 contained a T7 promoter followed by either the start sequence of GFP or gGFP, respectively. Reverse primers R1 and R2 correspond to sequences in the TMV 3′ UTR and the 3′ most end of the sgRNA scaffolding sequence, respectively. RNA synthesis reactions were verified for proper RNA synthesis by visualization using 1% agarose gel electrophoresis stained with ethidium bromide and the concentration for each reaction was quantified using a NanoDrop (ThermoFisher Scientific).

A PCR *mgfp5* fragment was amplified from untreated 16c genomic DNA followed by a cleanup step (DNA Clean & Concentrator -5, Zymo Research). Subsequently, 100 nM (final concentration) of purified Cas9 Nuclease (New England Biolabs) was first incubated in Cas9 Nuclease reaction buffer (New England Biolabs, Cambridge, MA) and 30 nM (final concentration) of each T7 synthesized gGFP containing transcripts. For assays using varying concentrations of gGFP template the corresponding final nM concentrations indicated was added to each reaction. The Cas9 and sgRNA mixture was then incubated at room temperature for 5 minutes. Then, 3 nM (final concentration) of purified *mgfp5* PCR template was added to each

reaction and incubated for 60 minutes at 37ºC. Reactions were visualized using 1.5 % agarose gel electrophoresis stained with ethidium bromide.

Agroinfiltration

Agrobacterium tumefaciens strain GV3101 (pMP90RK) was used for agroinfiltration of all binary vectors used, as described previously (Odokonyero et al. 2015). In brief, cultures were grown overnight (16-20 hrs) under constant shaking (250 rpm) at 28 ºC in LB media with 50 mg/L kanamycin. Cells were pelleted by centrifugation and resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES pH 5.7, and 200 µM acetosyringone). TRBO-based cultures were resuspended to a final infiltration concentration of OD_{600} 0.4 and pBINPLUS-sel Cas9 expression vectors at OD_{600} 0.5. pKYLX7-wt-p19 (Saxena et al. 2011)cultures were resuspended to a final concentration of OD_{600} 0.4 for the co-infiltration assays. Four week old 16c plants were agroinfiltrated with on the abaxial side of the leaf and returned to normal growth conditions. TRV RNA1 and RNA2 cultures used at a concentration of OD_{600} 0.25 for agroinfiltration of 3 week old 16c plants for VIGS assays.

DNA and indel assays

Single plant DNA samples for indel assays were carried out using leaf tissue from three infiltrated leaves, totaling 100-150 mg of tissue, to avoid tissue-dependent effects as well as to create a pooled biological replicate. DNA extractions were then carried out using *Quick* DNA Miniprep kit (Zymo Research). A total of 100 ng of genomic DNA was used for PCR amplification of *mgfp5* gene from 16c plants. Amplicons were cleaned (DNA Clean & Concentrator -5, Zymo Research) and 250-400 ng DNA was digested with *Bsg*I at 37ºC

overnight. *Bsg*I restriction enzyme resistance assays were then visualized using 1.2% agarose gel electrophoresis stained with ethidium bromide. Image files (.tif) were uploaded to ImageJ (NIH) and band intensities were measured using a standard gel peak analysis workflow.

Cas9-gRNA immunoprecipitation assays

At 3 dpi *N. benthamiana* tissue was ground in liquid nitrogen and resuspended in RIPA buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.5% Na-deoxycholate, and 150 mM NaCl) at a ratio of 1g of tissue to 3 ml of buffer. The tissue was centrifuged at 10,000 *g* for 20 min, then the supernatant was filtered through Miracloth (EMD Millipore) pre-soaked in RIPA buffer. Then, 4 ml of lysate was incubated by end over end agitation in 4ºC for 4 hrs with 1:400 anti-Cas9 antibody (Biolegend). Following, 200 µl of protein G agarose slurry (Thermo Scientific) was added and the mixture was incubated for an additional hour at 4ºC. The Cas9-protein G beads were collected by centrifugation at 2,500 *g* for 3 min. The supernatant was decanted and the agarose slurry was washed 5 times with 500 µl of RIPA buffer. Following the wash steps some of the resuspended slurry was used for western blot analysis to detect Cas9 was isolated using the Cas9-IP technique. The remainder of the Cas9-IP sample was used for RNA extractions.

Nuclei isolation

Leaf tissue was isolated at 3 dpi and ground in liquid nitrogen and resupended at a ratio of 1g to 5 ml of nuclei isolation buffer (0.25 M, sucrose, 15mM PIPES pH 6.8, 5 mM $MgCl_2$, 60 mM KCl, 15 mM NaCl, 1 mM $CaCl_2$, 0.9% Triton X-100). The ground tissue and buffer solution was then incubated on ice for 30 min, followed by centrifugation at 10,000 *g* for 20 min at 4ºC.

The supernatant (cytosol fraction) was then used directly for assays. The pellet (nuclei fraction) was washed with 500 µl of nuclei isolation buffer followed by re-isolation of nuclei through centrifugation 3 additional times. The final pellets (cytosol devoid nuclei) was then resuspended in 100 µl of RIPA buffer and used for downstream assays.

RNA extractions and (RT)-PCR

Plant RNA extractions were carried out using the Direct-zol RNA Miniprep kit (Zymo Research) following the manufacturer's instructions. cDNA was synthesized with equal volumes of total RNA using the M-MLV Reverse Transcriptase (Invitrogen) and gene specific primers. RT-PCR was carried out using Q5 High Fidelity Polymerase (New England Biolabs).

CHAPTER IV

CONCLUSION: DEVELOPING A VIRAL-DELIVERED GENE EDITING PLATFORM FOR FUNCTIONAL GENETICS IN DIVERSE *NICOTIANA* HOSTS

INTRODUCTION

The CRISPR/Cas9 gene editing platform has been developed as a diverse tool set for functional genetic studies (Cong et al. 2013; Gilbert et al. 2013; Ran et al. 2013). The hallmark of the CRISPR system is its simplicity of design. CRISPR consists of two main deliverable genetic parts, a "programmable" DNA nuclease (Cas9) and a sequence specific single guide RNA (sgRNA), which when bound to Cas9, localizes to a genomic DNA target of interest (gene) through complementary base pairing and catalyzes a DNA double stranded break (DSB) (Jinek et al. 2012; Cong et al. 2013; Mali et al. 2013). DSBs can be repaired by two host DNA repair mechanisms: homologous recombination and non-homologous end-joining (NHEJ). These repair mechanisms have been used for gene insertions and gene knockout events, respectively. While the CRISPR/Cas9 system has been used extensively among plant biologists as a tool for creating plant lines with gene knockouts and gene insertions (Schiml et al. 2014; Baltes et al. 2014), it has yet to be used as a transient screening tool for functional genetic screens in plants, as has been accomplished in other biological systems (Shalem et al. 2014). A system where native gene function can be assayed through transcriptional repression or through the production of missense transcripts through NHEJ-based nucleotide insertion and deletions (indels), would be of great utility to the plant community for functional genetic screens.

Plant-virus interaction studies have been primarily limited in their breadth by the phenotype which viruses inflict on the host (Wang 2015). This is perhaps best illustrated by plant

virus names, which typically consist of the host they were initially discovered infecting and the symptoms for which they cause on the host (e.g., *Tobacco mosaic virus*; TMV). However, these names, and ultimately most studies, do not take into consideration non-pathogenic (symptomless) interactions with a host or, more broadly, of virus infection in plants where they were not originally discovered. However, the development of viral vectors expressing reporter genes such as green fluorescent protein (GFP) have greatly aided our understanding of movement of virus cell-to-cell and systemic spread. In addition to development of viral vectors, the virology community has used many different techniques to understand host genes' effects on viral replication and movement, such as creating knockout or knockdown plants (Odokonyero et al. 2015), or transiently silencing genes using techniques such as viral induced gene silencing (VIGS) (Odokonyero et al. 2017). These techniques can be time consuming and can also introduce other variables to the screen which can be problematic, such as using *Agrobacterium* and multiple viruses to deliver silencing constructs. Since we have relatively little knowledge about how viruses cause disease, or in general how viruses infect hosts, at the molecular level, a tool that can analyze both viral replication and movement while simultaneously assaying for host gene function in various plants would be of great experimental utility.

Previously we developed a TMV based viral vector, TRBO, as a tool to transiently co-deliver a sgRNA and a GFP coding sequence (Cody et al. 2017). When co-delivered with a Cas9 expression vector, viral replication and movement can be assayed through the localization of GFP fluorescence while simultaneously knocking out a host factor (gene). While previous chapters have focused on the initial development of this technology and understanding the native mechanism in *Nicotiana benthamiana* which allow for co-delivery of a sgRNA and a protein coding segment on a single transcript, here I focus on the application of the tool for functional

genetic studies. My intent was to develop a virus vector as a tool to transiently assay for gene function in alternative *Nicotiana* host species. Here, I present demonstrate the proof-of-concept for multi-host gene targeting and the potential to target novel host factors using a single viral vector. I also give evidence for the viability of this screening tool in nine *Nicotiana* species. These experiments aim to stimulate the further development of these tools for more advanced applications in the future.

RESULTS

Replication of the TMV viral vector, TRBO-G, in nine *Nicotiana* species

To better understand TMV interactions within a diverse set of hosts, we selected 9 *Nicotiana* species (*N. attenuata, N. benthamiana, N. clevelandii, N. edwardsonii, N. glauca, N. otophora, N. sylvestris, N. tabacum,* and *N. tomentosiformis*) and infiltrated leaves of these plants with the TMV coat protein GFP replacement mutant, TRBO-G, and documented green fluorescence under a UV lamp over the course of 6 days by collecting images of individual leaves (**Figure 4.1**). It appeared that approximately half of the host plant species exhibited near saturation of GFP expression in infiltrated leaf tissue at 3 days post infiltration (3 dpi), excluding *N. sylvestris, N. edwardsonii, N. glauca* and *N. attenuata*. However, movement of TRBO-G to the vascular tissue occurred at 4 dpi in all hosts, except for *N. benthamiana* which had signs of TRBO-G replication and movement at 3 dpi. Perhaps the most intriguing viral replication and movement patterns are found in species which did not become saturated with TRBO-G replication, such as *N. glauca* and *N. edwardsonii*. The *N. edwardsonii* response represents a known resistance interaction among TMV and the TMV resistance (R) gene (*N*) that can be identified by the hypersensitive (necrotic) response (HR), seen in **Figure 4.1**, which has a purple

fluorescence under UV light (Balaji et al. 2007). While these images illustrate known interactions, such as the one that we observe in *N. edwardsonii*, it also brings to light interactions that are not well understood, such as the isolated replication "islands" of TRBO-G seen in *N. glauca*. These expression studies demonstrate the use of TRBO-G in *Nicotiana* hosts will be useful to examine the intricacies of virus-plant interactions.

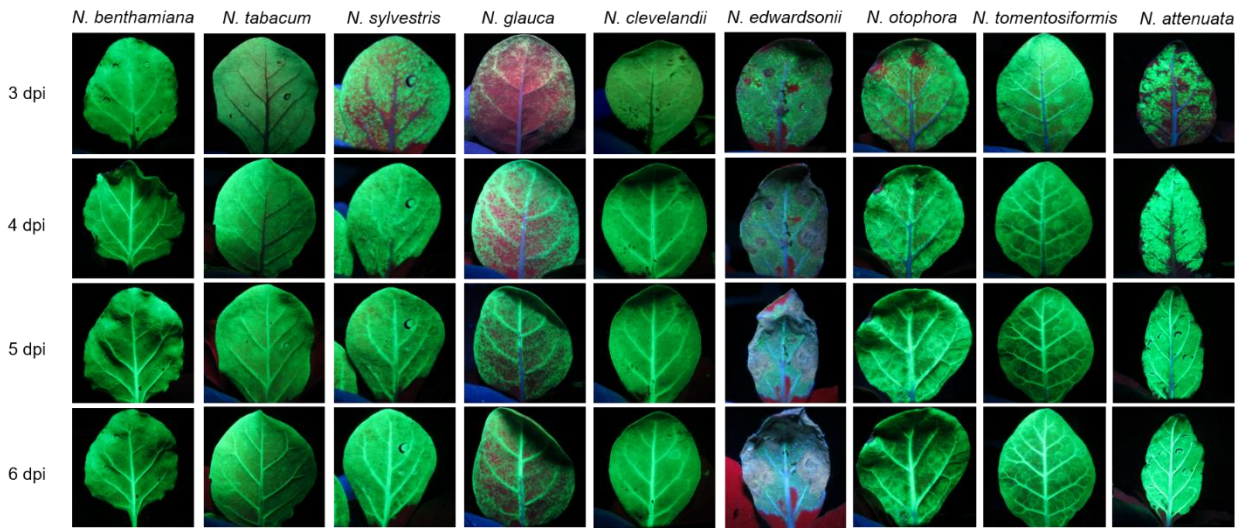


Figure 4.1. TRBO-G replication in *Nicotiana* species. The TRBO-G viral vector was infiltrated into 9 *Nicotiana* species and analyzed under UV light for GFP expression, representative of TMV replication and movement. Each column represents one of the nine *Nicotiana* species surveyed. The rows indicate images taken at 3, 4, 5, and 6 dpi from top to bottom, respectively.

Using *N. glauca* as a proof-of-concept for the TRBO-sgRNA delivery tool in alternative hosts

Following the TRBO-G expression experiments, the nature of the replication "islands" in *N. glauca* was investigated using the TRBO-sgRNA delivery vector to both measure TRBO replication and movement (as a model for TMV movement) through GFP fluorescence while simultaneously knocking out a host factor (gene). Prior to instigating species-wide screenings, we needed to demonstrate that the Cas9 and TRBO-sgRNA delivery system works in hosts other than *N. benthamiana*. Additionally, to reduce the number of host genes to be screened, we wanted to focus on specific pathways that may be responsible for the phenotype in *N. glauca*. The experiments in this chapter focus on comparisons between *N. benthamiana* and *N. glauca*.

Through previous experience, we hypothesized that the RNA silencing pathway is responsible for the TRBO-G replication pattern in *N. glauca*. To test this, we used the RNA silencing suppressor P19 from *Tomato bushy stunt virus* (TBSV) to inactivate the pathway and observe the resulting phenotype. For this purpose, *N. benthamiana* and *N. glauca* were either infiltrated with TRBO-G alone, or co-infiltrated with TRBO-G and the P19 expression vector pKYLX7-wt-p19 (Saxena et al. 2011) (**Figure 4.2A**). GFP expression was monitored over the course of 17 days through imaging under UV light. If the RNA silencing pathway is responsible for the aberrant TRBO-G replication phenotype in *N. glauca*, the addition of P19 should alter the observed phenotype to elevated levels of GFP expression when co-infiltrated with TRBO-G. Indeed, when TRBO-G and pKYLX7-wt-p19 were co-infiltrated in *N. glauca*, GFP expression saturated the leaves at 7 days post infiltration (dpi) while the tissue expressing TRBO-G only exhibited the typical isolated viral replication pattern previously observed in **Figure 4.1** (**Figure 4.2B**). Further support for the involvement of the RNA silencing pathway is demonstrated by the

reversion of *N. glauca* TRBO-G infiltrated leaves to the auto-fluorescent (red in color) phenotype observed from 10-17 dpi under UV light. A similar phenotype was observed in *N. benthamiana* at 17 dpi, as well, in TRBO-G infiltrated tissue, but not when co-infiltrated with the P19 expression vector.

Next we aimed to demonstrate the efficacy of the pHcoCas9/TRBO-sgRNA transient delivery tool in an alternate *Nicotiana* host to establish a proof-of-concept to analyze host gene function during viral infection (**Figure 4.2C**). We selected a previously developed TRBO-sgRNA expression vector which targets the conserved *Nicotiana* homolog ARGONAUTE 1 (TRBO-gAGO1 as seen in **Figure 4.2C**) (Cody et al. 2017). The *N. glauca* genome has been sequenced, but the data is not publicly available at this time (Khafizova et al. 2017). In an effort to work around this, an *AGO1* primer set used for *N. benthamiana* was tested on *N. glauca* genomic DNA, which yielded amplicons at the expected molecular weight for *AGO1*. Additionally, the gAGO1 "spacer" sequence (20 nt genomic complementary sequence) was analyzed for sequence complementarity using NCBI BLAST against the *Nicotiana* genus (taxid: 4085). The BLAST search verified that the sequence was conserved in most *Nicotiana* species, therefore, increasing the possibility of successful DNA catalytic events, "gene editing", in *N. glauca*. Next, *N. benthamiana* and *N. glauca* plants were infiltrated with TRBO-gAGO1 or both pHcoCas9 and TRBO-gAGO1. At 7 dpi, infiltrated leaf tissues were sampled and genomic DNA extracted. Following *AGO1* amplification using gene specific primers, a Surveyor Nuclease assay was carried out to detect the presence of indels. The Surveyor Nuclease assays yielded indels for both *N. benthamiana* and *N. glauca* leaves co-infiltrated with pHcoCas9 and TRBO-gAGO1, but not when infiltrated with TRBO-gAGO1 alone (**Figure 4.2D**). These demonstrate

the potential use of this tool for screens in species other than *N. benthamiana*, and the ability to use a single sgRNA to assay gene function in highly conserved homologs in multiple hosts.

Demonstrating the potential of TRBO-sgRNA delivery tool in alternate *Nicotiana* hosts

The experiments conducted with *N. glauca* and *N. benthamiana* demonstrate the viability of the pHcoCas9/TRBO-sgRNA delivery system for targeting of homologous genes in two closely related species. However, can this tool be used in other *Nicotiana* hosts? We identified four crucial steps which must be present for a good screen using these tools. When assaying for homologous gene function across closely related hosts (one gene, multiple hosts), it must be ensured that the sgRNA selected is present in each of the species. The selection of sgRNAs is entirely dependent on the host factors (genes) for which are being assayed. For instance, if targeting novel host factors (one host phenotype multiple genes to screen) in a host of interest—as would be expected to identify the *N. glauca* phenotypic response to TRBO-G—there must be genetic information (be it genomic or transcriptomic) present for gene identification and sgRNA design. In any case, the importance of sgRNA selection cannot be overstated as it is an integral part of the screen, but due to design parameters specificity to individual experiment we will focus on other key factors in each of the nine *Nicotiana* hosts in **Figure 4.1**. First, do the species have any sensitivity to P19? Second, does pHcoCas9 express Cas9 protein in each of the species? Third, can the TRBO-G vector replicate in selected *Nicotiana* species? The latter point is covered in **Figure 4.1**.

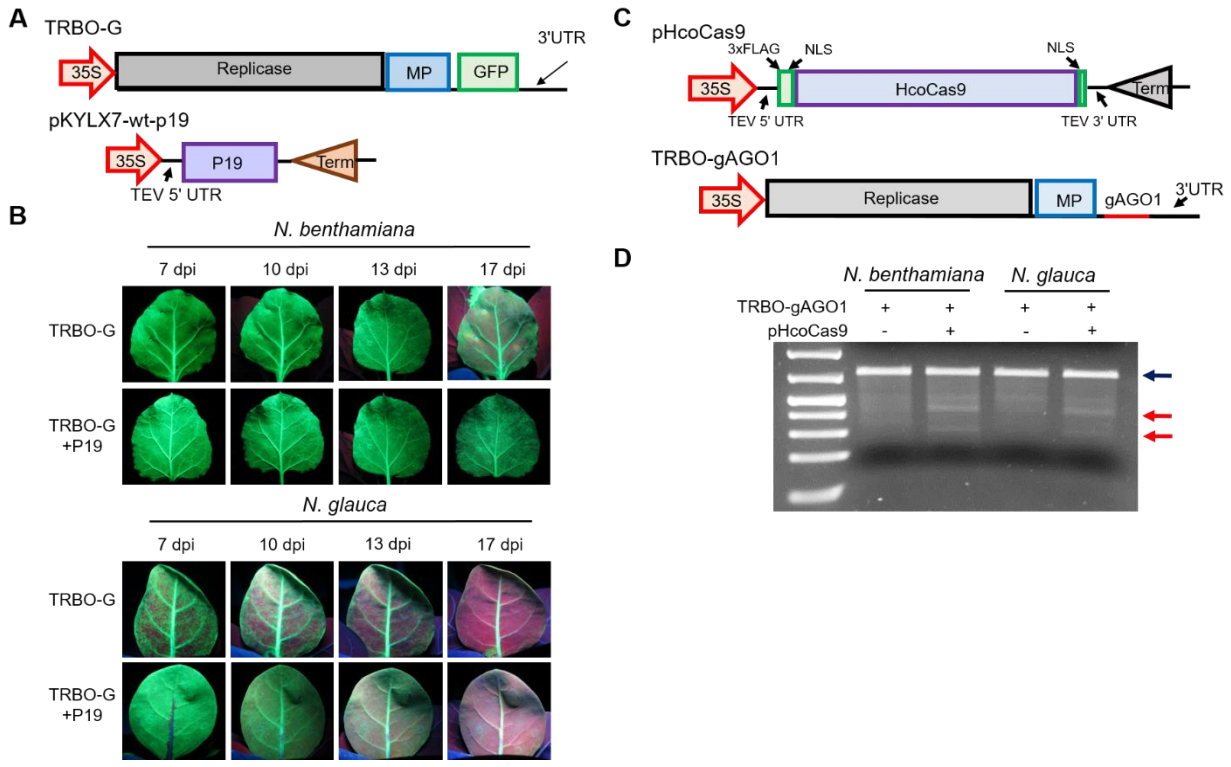


Figure 4.2. Using the pHcoCas9 and TRBO-sgRNA delivery tool in *N. benthamiana* and *N. glauca* to target a homologous gene. A) The binary vectors used for infiltration in *N. benthamiana* and *N. glauca*. TRBO-G is a coat protein (CP) replacement mutant which expresses green fluorescence protein (GFP). pKYLX7-wt-p19 is a transient P19 expression construct which transcription is driven by a 35S promoter and terminated by the nopaline synthase terminator. **B)** Image panels of TRBO-G alone or co-infiltrated with P19 under UV light of both *N. benthamiana* and *N. glauca*. Images were taken at 7, 10, 13, and 17 dpi and analyzed for GFP expression, indicative of TRBO-G replication and movement. **C)** pHcoCas9 is a binary vector which expresses Cas9 protein *in planta*. The human codon optimized Cas9 gene transcription is driven by a 35S promoter and terminated by the 35S terminator sequence. Transcripts also carry the *Tobacco etch virus* (TEV) 5′ and 3′ untranslated sequence (UTR) for increased production of Cas9 protein. The protein contains an N and C terminus nuclear translation signal (NLS) and a 3xFLAG epitope tag at the N terminus. TRBO-gAGO1 consists of the TRBO backbone with a sgRNA targeting the ARGONAUTE 1 gene (gAGO1) homologs in multiple *Nicotiana* species expressed under the CP subgenomic RNA promoter. **D)** Surveyor Nuclease assays from *N. benthamiana* and *N. glauca* tissue infiltrated with TRBO-gAGO1 or pHcoCas9 and TRBO-gAGO1 (indicated by + or – above the gel). *AGO1* genomic amplicons from each treatment were subjected to the Surveyor Nuclease enzyme. Bands which contain indels are indicated by red arrows and wild-type AGO1 sequence is highlighted by the blue arrow.

From our previous results we observed increased indel percentage when the P19 expression vector was co-infiltrated with pHcoCas9 and TRBO-sgRNA (Chiong 2018). While the reason for increased indel percentages when using P19 is still elusive—it is not necessarily due to an increase in Cas9 expression—nevertheless, it does potentially increase the potency of the screen when co-delivered with the two integral CRISPR parts, Cas9 and sgRNA. For this reason we elected to survey the *Nicotiana* species for P19 sensitivity. While previously there was a terrific survey of *Nicotiana* species for hypersensitive responses (HR) to P19 (Angel and Schoelz 2013), we wanted to recapitulate these results in our own lab environment. Following a survey of the nine *Nicotiana* species through infiltrating leaf tissue using pJL3:P19 (**Figure 4.2A**), HR for both *N. sylvestris* and *N. tabacum* was documented. Our findings were in agreement with Angel and Schoelz (2013) (**Table 4.1**).

Next we aimed to test the ability of pHcoCas9 expression vector to produce Cas9 protein transiently in each of the species. To test if *Nicotiana* species can transiently express Cas9, 7 of the 9 plant species were infiltrated with a combination of pHcoCas9 or pHcoCas9 and pKYLX7-wt-p19, and sampled at 3, 5, and 7. Western blots of cellular lysates were analyzed for each of the samples and rated based on band intensities, with 0 being no expression and 3 indicating similar expression levels to that of *N. benthamiana* (**Table 4.1**). Notably, all seven sampled *Nicotiana* plants demonstrated some level of Cas9 protein expression, thus allowing for the possibility of running screens using the pHcoCas9/TRBO-sgRNA delivery tool. However, the diversity of transient Cas9 protein expression could be a limiting factor for the screening system in some of the hosts described.

Nicotiana species	P19 HR	pHcoCas9 expression	TRBO-G replication
N. benthamiana	No	+++	+++
N. sylvestris	HR	+	++
N. tabacum	HR	++	+++
N. clevelandii	No	N/A	+++
N. edwardsonii	No	+++	HR
N. glauca	No	+	+
N. otophora	No	+++	+++
N. tomentosiformis	No	+++	+++
N. attenuata	No	N/A	++

Table 4.1. *Nicotiana* species profiling for the use of the Cas9/TRBO-sgRNA delivery tool. From left to right, the first column indicated each of the 9 species assayed for P19 hypersensitivity, pHcoCas9 binary vector-based Cas9 protein expression, and TRBO-G replication. P19 HR is indicated by either a "No" for no hypersensitivity or "HR" for hypersensitivity response. The pHcoCas9 expression was analyzed using western blots and given a score of 0-3 (0, +, ++, or +++) with 0 indicating no expression and +++ indicating similar expression levels to *N. benthamiana*. Missing data points are denoted by "N/A". TRBO-G replication was measure by UV lamp images taken in Figure 1. Scores (0, +, ++, or +++) were given for each species with 0 indicating no replication and +++ representing similar expression patterns to that of *N.benthamiana*. HR represents a hypersensitive response from the plant due to TRBO-G or P19.

DISCUSSION

CRISPR/Cas9 use as a gene editing tool in a diverse group of organisms has been widely adopted by molecular biologist. One aspect of CRISPR technology which has aided the excitement of the technique is its potential use for functional genetic studies in "non-model" organisms. While this idea is a provocative one for plant biologists who are not working on *Arabidopsis thaliana*, it has yet to be implemented in a diverse set of plant species on a large scale. Additionally, many of the other models have adapted the technology into a high-throughput screening technique for targeting of entire genomes (Wang et al. 2014; Boettcher et al. 2018). In plant biology, most studies have settled on using tedious plant transformation techniques to generate desired plant lines using CRISPR gene editing tools (Schiml et al. 2014; Gao et al. 2015). While exciting, it is far removed from being able to be used in "non-model" species as the pivotal functional genetics tool which initially intrigued the field. In fact, there are a variety of gene editing tools which could have been used, and were, in combination with transformation technology over a decade ago to obtain similar results as using CRISPR technology using zinc finger nucleases (Lloyd et al. 2005; Wright et al. 2005). In short, since the initial proof-of-concept studies using the CRISPR gene editing technique in plants in 2013 (Li et al. 2013; Nekrasov et al. 2013; Xie and Yang 2013), little has been done to change how plant biologist think about using the tool.

Here, I demonstrate the use of CRISPR technology by adapting a viral vector to deliver sgRNAs to a diversity of *Nicotiana* species, opening up the possibility for surveys of host factors (genes) responsible for specific observed phenotypes in viral-host interactions. While it has been known that there are a variety of responses to TMV in *Nicotiana* hosts (Holmes 1946), little has been done to molecularly understand the diversity of responses of the virus in alternative hosts. I

observed a unique phenotype when *N. glauca* was infected with TRBO-G, an effect I describe as green replication "islands". After co-infiltrating the RNA silencing suppressor P19 with TRBO-G, it was observed that the isolated replication "islands" were no longer present and that the entire leaf was saturated with GFP, indicative of vigorous viral replication and movement. It has been previously established that the TMV replicase protein 126-kDa has RNA silencing suppression activity (Ding et al. 2004). While there is known suppressor activity of 126-kDa, it has also been reported that TMV increases the expression of its gene products when co-infected with TBSV (Mendoza et al. 2017). The synergism between TMV and TBSV is thought to be mostly due to the potent RNA silencing activity of P19 expressed by TBSV. Perhaps the differences in phenotypes observed in *N. glauca* upon infiltration of TRBO-G compared to that of TRBO-G with P19 is due to the suppressor's activity are not acting on the same mechanism of the RNA silencing pathway. In this case, we might be observing an intricate evolutionary interaction between TMV and *N. glauca*, where *N. glauca* has actually evolved a mechanism to evade the suppression activity of 126-kDa protein. Further evidence for this specialized reaction with TMV could be explained by differences in the 126-kDa protein of TMV and *Tobacco mild green mosaic virus* (TMGMV), which might be the reason the latter is the primary TMV species in *N. glauca* in the field (Fraile et al. 1997).

Co-infiltration experiments with P19 and TRBO-G almost certainly indicate that the RNA silencing pathway is responsible for the phenotype observed when TRBO-G alone was infiltrated in *N. glauca*, it does not yield a specific host factor that might be responsible for the response. To demonstrate that our previously developed pHcoCas9 and TRBO-sgRNA transient delivery technique works in other *Nicotiana* hosts (Cody et al. 2017), and to potentially screen for the host factor which causes these disease symptoms in the future, we tested the system in *N.*

glauca. Following targeting of the highly conserved *AGO1* gene in both *N. benthamiana* and *N. glauca*, I confirmed that the tool does generate indels in both species at the targeted location. It is also noteworthy that the editing events that were measured are generated entirely through transient expression of the toolset, not through plant transformation and regeneration.

The experiments and tools described in this chapter demonstrate that our virus toolset can be used for efficient gene expression and gene editing across *Nicotiana* species. While simple in nature, the results point towards a transformation in how plant biologists and virologists alike might approach future research to understand complicated interactions between hosts and viruses. The cis-expression of a sgRNA from the viral genome implicates the CRISPR system—specifically the genomic indels resulting from the system—into the fitness of the "virus". In particular, we can be select for a virus carrying sgRNAs that "aid" the spread of the virus by targeting and mutating key genes involved in host resistance to the virus. With the increase use of whole genome screening using sgRNA libraries being an effective tool in other areas of research, why is this not being done in plants? Furthermore, can this be an addendum to natural selection processes that already exists between the host and pathogen? Similar approaches have been carried out in phage, albeit for purpose of producing molecules for specific synthetic purposes (Esvelt et al. 2011). Instead of creating an "artificial" selection process based on a new molecule we would like to create, as is done in phage-assisted continues evolution or PACE, perhaps we can change the host in which we assay or the environmental conditions in which the host and viral interactions are occurring. In doing this we are not assaying virus-host interactions as black and white, but, rather, as the dynamic interactions that are occurring in environmental situations. New technology such as CRISPR should not just increase our throughput of currently established experimental procedures (such as screening), but it should also serve to widen our

lens of biological perspective when running assays. Advancements, such as the TRBO-sgRNA delivery tool described in the current and previous chapters, could change how we understand viruses and evolutionary fitness in hosts where disease symptoms are not necessarily observed. Perhaps our disease and host centric perspectives have caused us to overlook the larger question of the importance of viruses in the environment.

MATERIALS AND METHODS

Agroinfiltration and UV imaging

Agrobacterium tumefaciens strain GV3101 (pMP90RK) was used for agroinfiltration of all binary vectors used. Methods for cell preparation are as described previously (Cody et al. 2017). In brief, cultures were grown overnight (16-20 hrs) under constant shaking (250 rpm) at 28 ºC in LB media with 50 mg/L kanamycin. Cells were pelleted by centrifugation and resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES pH 5.7, and 200 µM acetosyringone). TRBO-G, TRBO-gAGO1 and pJL3:P19 were resuspended to a final infiltration concentration of OD_{600} 0.4 and the pHcoCas9 vector at OD_{600} 0.5. Co-infiltrations were carried out by mixing equal volumes of resuspended TRBO-G and P19, or TRBO-gAGO1 and pHcoCas9 cultures to the above mentioned final concentrations. Four- six week old *Nicotiana* plants were infiltrated with *Agrobacterium* suspensions on the abaxial side of the leaf and returned to normal growth conditions. Following infiltration of TRBO-G, or TRBO-G and pKYLX7-wt-p19, plants were visualized under a handheld UV mercury lamp as previously described (Odokonyero et al. 2015; Everett et al. 2010).

Genomic DNA extractions and Surveyor Nuclease assay

Single plant DNA samples for *N. benthamiana* and *N. glauca* indel assays were carried out using 50 mg of leaf tissue from three infiltrated leaves, totaling 150 mg of tissue, to avoid tissue-dependent effects as well as to create a pooled biological replicate. DNA extractions were then carried out using the ZR Plant/Seed DNA Miniprep kit (Zymo Research, Irvine, CA). Then, 100 ng of genomic DNA was used for PCR amplification of the *AGO1* alleles. Amplicons were then cleaned (DNA Clean & Concentrator -5, Zymo Research,) and resuspended in DNase and RNase-free water. Amplicon concentrations were measured using a NanoDrop (ThermoFisher Scientific), and 200-400 ng DNA used for final Surveyor Nuclease assays. *AGO1* amplicons were treated with Surveyor Nuclease a DNA mismatch endonuclease. Surveyor Nuclease digestions were carried out using the manufacturer's instructions (Integrated DNA Technologies). Surveyor Nuclease reactions were visualized using 1.2% agarose gel electrophoresis stained with ethidium bromide.

REFERENCES

Aguero J, Ruiz-Ruiz S, Del Carmen Vives M, Velazquez K, Navarro L, Pena L, Moreno P, Guerri J (2012) Development of viral vectors based on *Citrus leaf blotch virus* to express foreign proteins or analyze gene function in citrus plants. Mol Plant Microbe In. **25**:1326-1337.

Ahlquist P, French R, Janda M, Loesch-Fries LS (1984) Multicomponent RNA plant virus infection derived from cloned viral cDNA. Proc Natl Acad Sci USA. **81**:7066-7070.

Akiyama BM, Laurence HM, Massey AR, Costantino DA, Xie X, Yang Y, Shi P-Y, Nix JC, Beckham JD, Kieft JS (2016) Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. Science. **354**:1148-1152.

Ali Z, Abul-Faraj A, Li L, Ghosh N, Piatek M, Mahjoub A, Aouida M, Piatek A, Baltes NJ, Voytas DF, Dinesh-Kumar S, Mahfouz MM (2015) Efficient virus-mediated genome editing in plants using the CRISPR/Cas9 system. Mol Plant. **8**:1288-1291.

Alvarado V, Scholthof HB (2009) Plant responses against invasive nucleic acids: RNA silencing and its suppression by plant viral pathogens. Semin Cell Dev Biol. **20**:1032-1040.

Anandalakshmi R, Marathe R, Ge X, Herr JM, Jr., Mau C, Mallory A, Pruss G, Bowman L, Vance VB (2000) A calmodulin-related protein that suppresses posttranscriptional gene silencing in plants. Science. **290**:142-144.

Anandalakshmi R, Pruss GJ, Ge X, Marathe R, Mallory AC, Smith TH, Vance VB (1998a) A viral suppressor of gene silencing in plants. Proc Natl Acad Sci USA. **95**:13079-13084.

Anandalakshmi R, Pruss GJ, Ge X, Marathe R, Mallory AC, Smith TH, Vance VB (1998b) A viral suppressor of gene silencing in plants. Proc Natl Acad Sci USA. **95**:13079-13084.

Anders C, Niewoehner O, Duerst A, Jinek M (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature. **513**:569-573.

Angel CA, Schoelz JE (2013) A survey of resistance to *Tomato bushy stunt virus* in the genus *Nicotiana* reveals that the hypersensitive response is triggered by one of three different viral proteins. Mol Plant Microbe In. **26**:240-248.

Balaji B, Cawly J, Angel C, Zhang Z, Palanichelvam K, Cole A, Schoelz J (2007) Silencing of the N family of resistance genes in *Nicotiana edwardsonii* compromises the hypersensitive response to tombusviruses. Mol Plant Microbe In. **20**:1262-1270.

Baltes NJ, Gil-Humanes J, Cermak T, Atkins PA, Voytas DF (2014) DNA replicons for plant genome engineering. Plant Cell. **26**:151-163.

Barrangou R, Birmingham A, Wiemann S, Beijersbergen RL, Hornung V, Smith A (2015) Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference. Nucleic Acids Res. **43**:3407-3419.

Basak J, Nithin C (2015) Targeting non-coding RNAs in plants with the CRISPR-Cas technology is a challenge yet worth accepting. Front Plant Sci. **6**:1001.

Baulcombe DC (1999) Fast forward genetics based on virus-induced gene silencing. Curr Opin Plant Biol. **2**:109-113.

Baulcombe DC, Chapman S, Cruz S (1995) Jellyfish green fluorescent protein as a reporter for virus infections. Plant J. **7**:1045-1053.

Bevan M (1984) Binary *Agrobacterium* vectors for plant transformation. Nucleic Acids Res. **12**:8711-8721.

Boettcher M, Tian R, Blau JA, Markegard E, Wagner RT, Wu D, Mo X, Biton A, Zaitlen N, Fu H, McCormick F, Kampmann M, McManus MT (2018) Dual gene activation and knockout screen reveals directional dependencies in genetic networks. Nat Biotechnol. **36**:170-178.

Brooks C, Nekrasov V, Lippman ZB, Van Eck J (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. Plant Physiol. **166**:1292-1297.

Buck KW (1999) Replication of *Tobacco mosaic virus* RNA. Philos Trans R Soc Lond B Biol Sci. **354**:613-627.

Burch-Smith TM, Anderson JC, Martin GB, Dinesh-Kumar SP (2004) Applications and advantages of virus-induced gene silencing for gene function studies in plants. Plant J. **39**:734-746.

Cermak T, Curtin SJ, Gil-Humanes J, Cegan R, Kono TJY, Konecna E, Belanto JJ, Starker CG, Mathre JW, Greenstein RL, Voytas DF (2017) A multipurpose toolkit to enable edvanced genome engineering in plants. Plant Cell. **29**:1196-1217.

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. **39**:e82.

Chiong K (2018) *Tobacco mosaic virus* as a gene editing platform. Texas A&M University.

Cody WB, Scholthof HB, Mirkov TE (2017) Multiplexed gene editing and protein overexpression using a *Tobacco mosaic virus* viral vector. Plant Physiol. **175**:23-35.

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

Culver JN, Lehto K, Close SM, Hilf ME, Dawson WO (1993) Genomic position affects the expression of *Tobacco mosaic virus* movement and coat protein genes. Proc Natl Acad Sci USA. **90**:2055-2059.

Dahlman JE, Abudayyeh OO, Joung J, Gootenberg JS, Zhang F, Konermann S (2015) Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. Nat Biotechnol. **33**:1159-1161.

Dawson WO, Bar-Joseph M, Garnsey SM, Moreno P (2015) *Citrus tristeza virus*: Making an ally from an enemy. Ann Rev Phytopathol. **53**:137-155.

Dawson WO, Beck DL, Knorr DA, Grantham GL (1986) cDNA cloning of the complete genome of *Tobacco mosaic virus* and production of infectious transcripts. Proc Natl Acad Sci USA. **83**:1832-1836.

Dawson WO, Lewandowski DJ, Hilf ME, Bubrick P, Raffo AJ, Shaw JJ, Grantham GL, Desjardins PR (1989) A *Tobacco mosaic virus*-hybrid expresses and loses an added gene. Virology. **172**:285-292.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. **471**:602-607.

Ding XS, Liu J, Cheng N-H, Folimonov A, Hou Y-M, Bao Y, Katagi C, Carter SA, Nelson RS (2004) The *Tobacco mosaic virus* 126-kDa protein associated with virus replication and movement suppresses RNA silencing. Mol Plant Microbe In. **17**:583-592.

Dolja VV, Koonin EV (2013) The closterovirus-derived gene expression and RNA interference vectors as tools for research and plant biotechnology. Front Microbiol. **4**:83.

Dong C, Beetham P, Vincent K, Sharp P (2006) Oligonucleotide-directed gene repair in wheat using a transient plasmid gene repair assay system. Plant Cell Rep. **25**:457-465.

Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science. **346**.

Dugar G, Herbig A, Förstner KU, Heidrich N, Reinhardt R, Nieselt K, Sharma CM (2013) High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. PLOS Genet. **9**:e1003495.

Esvelt KM, Carlson JC, Liu DR (2011) A system for the continuous directed evolution of biomolecules. Nature. **472**:499-503.

Everett AL, Scholthof HB, Scholthof KB (2010) Satellite *Panicum mosaic virus* coat protein enhances the performance of plant virus gene vectors. Virology. **396**:37-46.

Fauser F, Schiml S, Puchta H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. Plant J. **79**:348-359.

Fernandez-Pozo N, Rosli Hernan G, Martin Gregory B, Mueller Lukas A (2015) The SGN VIGS tool: User-friendly software to design Virus-Induced Gene Silencing (VIGS) CONSTRUCTS FOR FUNCTIONAL GENOMICS. Mol Plant. **8**:486-488.

Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature. **260**:500.

Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suva ML, Bernstein BE (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. **529**:110-114.

Forment J, Gadea J, Huerta L, Abizanda L, Agusti J, Alamar S, Alos E, Andres F, Arribas R, Beltran JP, Berbel A, Blazquez MA, Brumos J, Canas LA, Cercos M, Colmenero-Flores JM, Conesa A, Estables B, Gandia M, Garcia-Martinez JL, Gimeno J, Gisbert A, Gomez G, Gonzalez-Candelas L, Granell A, Guerri J, Lafuente MT, Madueno F, Marcos JF, Marques MC, Martinez F, Martinez-Godoy MA, Miralles S, Moreno P, Navarro L, Pallas V, Perez-Amador MA, Perez-Valle J, Pons C, Rodrigo I, Rodriguez PL, Royo C, Serrano R, Soler G, Tadeo F, Talon M, Terol J, Trenor M, Vaello L, Vicente O, Vidal C, Zacarias L, Conejero V (2005) Development of a citrus genome-wide EST collection and cDNA microarray as resources for genomic studies. Plant Mol Biol. **57**:375-391.

Fraile A, Escriu F, Aranda MA, Malpica JM, Gibbs AJ, García-Arenal F (1997) A century of tobamovirus evolution in an Australian population of *Nicotiana glauca*. J Virol. **71**:8316-8320.

Franck A, Guilley H, Jonard G, Richards K, Hirth L (1980) Nucleotide sequence of *Cauliflower mosaic virus* DNA. Cell. **21**:285-294.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. **32**:279-284.

Gallie DR, Sleat DE, Watts JW, Turner PC, Wilson TMA (1987) The 5'-leader sequence of *Tobacco mosaic virus* RNA enhances the expression of foreign gene transcripts *in vitro* and *in vivo*. Nucleic Acids Res. **15**:3257-3273.

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

Gao SJ, Damaj MB, Park JW, Beyene G, Buenrostro-Nava MT, Molina J, Wang X, Ciomperlik JJ, Manabayeva SA, Alvarado VY, Rathore KS, Scholthof HB, Mirkov TE (2013) Enhanced transgene expression in sugarcane by co-expression of virus-encoded RNA silencing suppressors. PLOS One. **8**:e66046.

Gao Y, Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. J Integr Plant Biol. **56**:343-349.

Gardner RC, Howarth AJ, Hahn P, Brown-Luedi M, Shepherd RJ, Messing J (1981) The complete nucleotide sequence of an infectious clone of *Cauliflower mosaic virus* by M13mp7 shotgun sequencing. Nucleic Acids Res. **9**:2871-2888.

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. **154**:442-451.

Gleba Y, Klimyuk V, Marillonnet S (2007) Viral vectors for the expression of proteins in plants. Curr Opin Plant Biol. **18**:134-141.

Goelet P, Lomonossoff GP, Butler PJ, Akam ME, Gait MJ, Karn J (1982) Nucleotide sequence of *Tobacco mosaic virus* RNA. Proc Natl Acad Sci USA. **79**:5818-5822.

Grimsley N, Hohn B, Hohn T, Walden R (1986) "Agroinfection," an alternative route for viral infection of plants by using the Ti plasmid. Proc Natl Acad Sci USA. **83**:3282-3286.

Gursinsky T, Pirovano W, Gambino G, Friedrich S, Behrens SE, Pantaleo V (2015) Homeologs of the *Nicotiana benthamiana* antiviral ARGONAUTE1 show different susceptibilities to microRNA168-mediated control. Plant Physiol. **168**:938-952.

Hajeri S, Killiny N, El-Mohtar C, Dawson WO, Gowda S (2014) *Citrus tristeza virus*-based RNAi in citrus plants induces gene silencing in *Diaphorina citri*, a phloem-sap sucking insect vector of citrus greening disease (Huanglongbing). J Biotechnol. **176**:42-49.

Haseloff J, Siemering KR, Prasher DC, Hodge S (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. Proc Natl Acad Sci USA. **94**:2122-2127.

Holmes FO (1946) A comparison of the experimental host ranges of Tobacco-etch and Tobacco-mosaic viruses. Phytopathology. **36**:643-659.

Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell. **157**:1262-1278.

Jia H, Wang N (2014) Targeted genome editing of sweet orange using Cas9/sgRNA. PLOS One. **9**:e93806.

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

Jinek M, Coyle SM, Doudna JA (2011) Coupled 5' nucleotide recognition and processivity in Xrn1-mediated mRNA decay. Mol Cell. **41**:600-608.

Jones L, Keining T, Eamens A, Vaistij FE (2006) Virus-induced gene silencing of argonaute genes in *Nicotiana benthamiana* demonstrates that extensive systemic silencing requires *Argonaute1*-like and *Argonaute4*-like genes. Plant Physiol. **141**:598-606.

Karvelis T, Gasiunas G, Miksys A, Barrangou R, Horvath P, Siksnys V (2013) crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. RNA Biol. **10**:841-851.

Kay R, Chan A, Daly M, McPherson J (1987) Duplication of CaMV 35*S* promoter sequences creates a strong enhancer for plant genes. Science. **236**:1299-1302.

Khafizova G, Dobrynin P, Polev D, Matveeva T (2017) Whole-genome sequencing of *Nicotiana glauca*. bioRxiv.

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. **517**:583-588.

Kumagai MH, Donson J, della-Cioppa G, Harvey D, Hanley K, Grill LK (1995) Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. Proc Natl Acad Sci USA. **92**:1679-1683.

Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc. **8**:2180-2196.

Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and *Nicotiana benthamiana* using guide RNA and Cas9. Nat Biotechnol. **31**:688-691.

Lindbo JA (2007) TRBO: a high-efficiency *Tobacco mosaic virus* RNA-based overexpression vector. Plant Physiol. **145**:1232-1240.

Lindbo JA, Silva-Rosales L, Proebsting WM, Dougherty WG (1993) Induction of a highly specific antiviral state in transgenic plants: Implications for regulation of gene expression and virus resistance. Plant Cell. **5**:1749-1759.

Liu Y, Schiff M, Dinesh-Kumar SP (2002) Virus-induced gene silencing in tomato. Plant J. **31**:777-786.

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

Ma X, Nicole M-C, Meteignier L-V, Hong N, Wang G, Moffett P (2015) Different roles for RNA silencing and RNA processing components in virus recovery and virus-induced gene silencing in plants. J Exp Bot. **66**:919-932.

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

Malina A, Mills JR, Cencic R, Yan Y, Fraser J, Schippers LM, Paquet M, Dostie J, Pelletier J (2013) Repurposing CRISPR/Cas9 for in situ functional assays. Genes Dev. **27**:2602-2614.

Mandadi KK, Scholthof KB (2013) Plant immune responses against viruses: how does a virus cause disease? Plant Cell. **25**:1489-1505.

Mendoza MR, Payne AN, Castillo S, Crocker M, Shaw BD, Scholthof HB (2017) Expression of separate proteins in the same plant leaves and cells using two independent virus-based gene vectors. Front Plant Sci. **8**.

Mikami M, Toki S, Endo M (2017) *In planta* processing of the SpCas9–gRNA complex. Plant Cell Physiol. **58**:1857-1867.

Morgens DW, Deans RM, Li A, Bassik MC (2016) Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. Nat Biotechnol.

National Academies of Sciences E, and Medicine (2016) Genetically engineered crops: Experiences and prospects. The National Academies Press, Washington, DC. doi:10.17226/23395

Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. Nat Biotechnol. **31**:691-693.

Nishimasu H, Ran FA, Hsu Patrick D, Konermann S, Shehata Soraya I, Dohmae N, Ishitani R, Zhang F, Nureki O (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. **156**:935-949.

Odokonyero D, Mendoza MR, Alvarado VY, Zhang J, Wang X, Scholthof HB (2015) Transgenic down-regulation of ARGONAUTE2 expression in *Nicotiana benthamiana* interferes with several layers of antiviral defenses. Virology. **486**:209-218.

Odokonyero D, Mendoza MR, Moffett P, Scholthof HB (2017) *Tobacco rattle virus* (TRV)-mediated silencing of *Nicotiana benthamiana* ARGONAUTES (NbAGOs) reveals new antiviral candidates and dominant effects of TRV-NbAGO1. Phytopathology. **107**:977-987.

Ohle C, Tesorero R, Schermann G, Dobrev N, Sinning I, Fischer T (2016) Transient RNA-DNA hybrids are required for efficient double-strand break repair. Cell. **167**:1001-1013.e1007.

Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F (2014) CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell. **159**:440-455.

Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. Nat Biotechnol. **23**:967-973.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. **152**:1173-1183.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. **154**:1380-1389.

Ratcliff F, Martin-Hernandez AM, Baulcombe DC (2001) Technical Advance: *Tobacco rattle virus* as a vector for analysis of gene function by silencing. Plant J. **25**:237-245.

Restrepo MA, Freed DD, Carrington JC (1990) Nuclear transport of plant potyviral proteins. Plant Cell. **2**:987-998.

Ruiz MT, Voinnet O, Baulcombe DC (1998) Initiation and maintenance of virus-induced gene silencing. Plant Cell. **10**:937-946.

Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res. **39**:9275-9282.

Saxena P, Hsieh YC, Alvarado VY, Sainsbury F, Saunders K, Lomonossoff GP, Scholthof HB (2011) Improved foreign gene expression in plants using a virus-encoded suppressor of RNA silencing modified to be developmentally harmless. Plant Biotechnol J. **9**:703-712.

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

Scholthof HB, Alvarado VY, Vega-Arreguin JC, Ciomperlik J, Odokonyero D, Brosseau C, Jaubert M, Zamora A, Moffett P (2011) Identification of an ARGONAUTE for antiviral RNA silencing in *Nicotiana benthamiana*. Plant Physiol. **156**:1548-1555.

Scholthof HB, Scholthof KB, Jackson AO (1996) Plant virus gene vectors for transient expression of foreign proteins in plants. Annu Rev Phytopathol. **34**:299-323.

Scholthof KB, Mirkov TE, Scholthof HB (2002) Plant virus gene vectors: biotechnology applications in agriculture and medicine. In: Genet. Eng., vol 24. 2002/11/06 edn., pp 67-85

Sean C, Tony K, David B (1992) *Potato virus X* as a vector for gene expression in plants. Plant J. **2**:549-557.

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. **343**:84-87.

Shalem O, Sanjana NE, Zhang F (2015) High-throughput functional genomics using CRISPR-Cas9. Nat Rev Genet. **16**:299-311.

Siegel A (1985) Plant-virus-based vectors for gene transfer may be of considerable use despite a presumed high error frequency during RNA synthesis. Plant Mol Biol. **4**:327.

Stanley J, Markham PG, Callis RJ, Pinner MS (1986) The nucleotide sequence of an infectious clone of the geminivirus *Beet curly top virus*. EMBO J. **5**:1761-1767.

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. **507**:62-67.

Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat Biotechnol. **32**:569-576.

van der Oost J, Westra ER, Jackson RN, Wiedenheft B (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Micro. **12**:479-492.

Voinnet O, Lederer C, Baulcombe DC (2000) A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. Cell. **103**:157-167.

Wang A (2015) Dissecting the molecular network of virus-plant interactions: The complex roles of host factors. Annu Rev Phytopathol. **53**:45-66.

Wang T, Wei JJ, Sabatini DM, Lander ES (2014) Genetic screens in human cells using the CRISPR-Cas9 system. Science. **343**:80-84.

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

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

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

Xu J, Yang JY, Niu QW, Chua NH (2006) Arabidopsis DCP2, DCP1, and VARICOSE form a decapping complex required for postembryonic development. Plant Cell. **18**:3386-3398.

Xu Q, Chen L-L, Ruan X, Chen D, Zhu A, Chen C, Bertrand D, Jiao W-B, Hao B-H, Lyon MP, Chen J, Gao S, Xing F, Lan H, Chang J-W, Ge X, Lei Y, Hu Q, Miao Y, Wang L, Xiao S, Biswas MK, Zeng W, Guo F, Cao H, Yang X, Xu X-W, Cheng Y-J, Xu J, Liu J-H, Luo OJ, Tang Z, Guo W-W, Kuang H, Zhang H-Y, Roose ML, Nagarajan N, Deng X-X, Ruan Y (2013) The draft genome of sweet orange (*Citrus sinensis*). Nat Genet. **45**:59-66.

Yin K, Han T, Liu G, Chen T, Wang Y, Yu AY, Liu Y (2015) A geminivirus-based guide RNA delivery system for CRISPR/Cas9 mediated plant genome editing. Sci Rep. **5**:14926.

Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS, Lim WA (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell. **160**:339-350.

Zetsche B, Gootenberg Jonathan S, Abudayyeh Omar O, Slaymaker Ian M, Makarova Kira S, Essletzbichler P, Volz Sara E, Joung J, van der Oost J, Regev A, Koonin Eugene V, Zhang F (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. **163**:759-771.

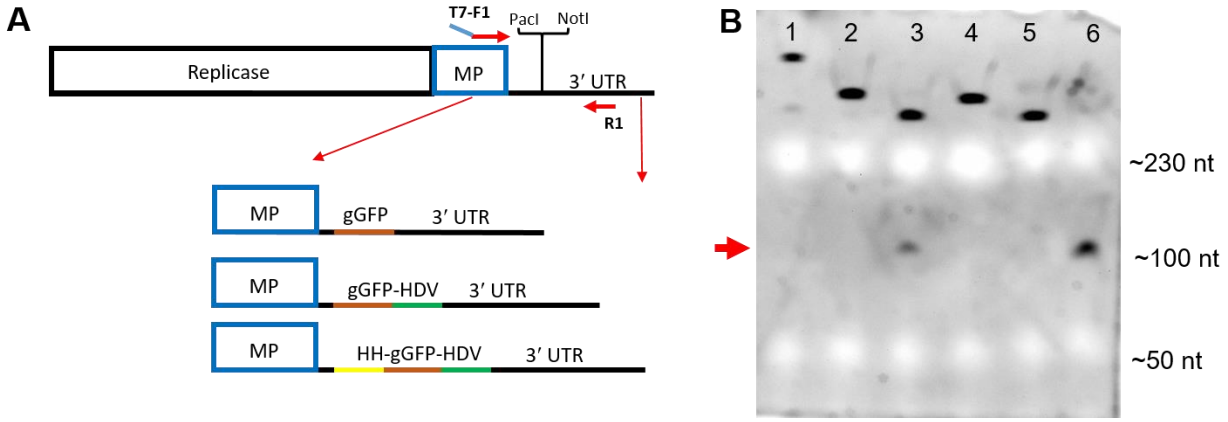
Zhang X, Choi PS, Francis JM, Imielinski M, Watanabe H, Cherniack AD, Meyerson M (2015) Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. Nat Genet.

Zhang Y, Heidrich N, Ampattu Biju J, Gunderson Carl W, Seifert HS, Schoen C, Vogel J, Sontheimer Erik J (2013) Processing-Independent CRISPR RNAs Limit Natural Transformation in Neisseria meningitidis. Mol Cell. **50**:488-503.

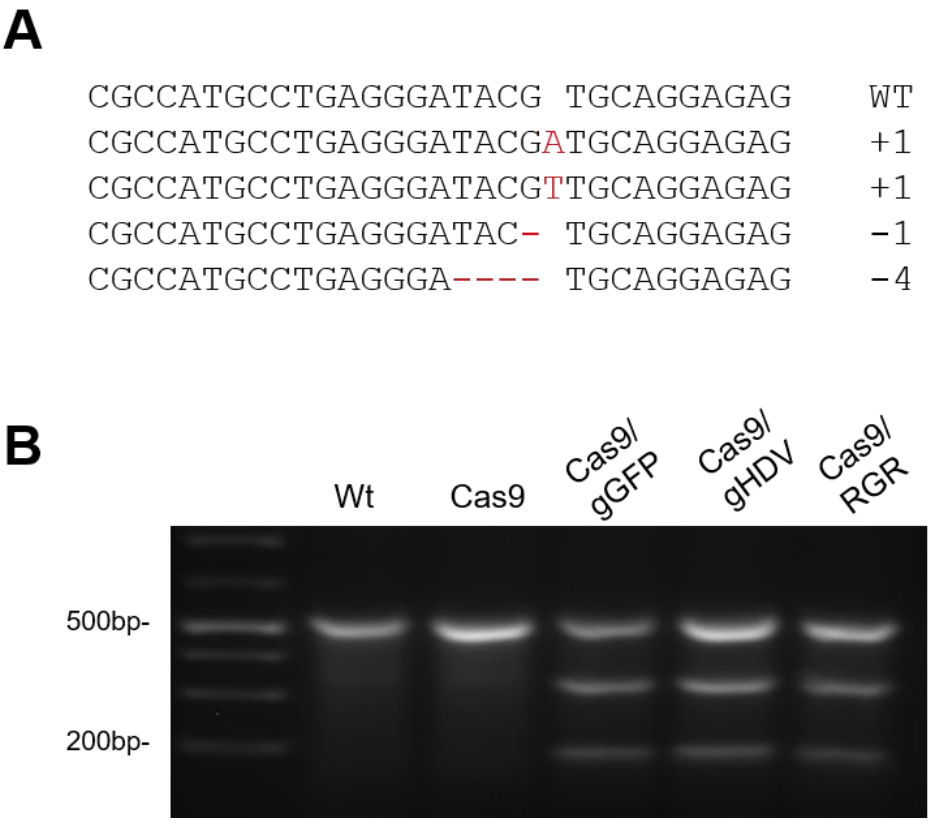
Zheng T, Hou Y, Zhang P, Zhang Z, Xu Y, Zhang L, Niu L, Yang Y, Liang D, Yi F, Peng W, Feng W, Yang Y, Chen J, Zhu YY, Zhang L-H, Du Q (2017) Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. Sci Rep. **7**:40638.

APPENDIX

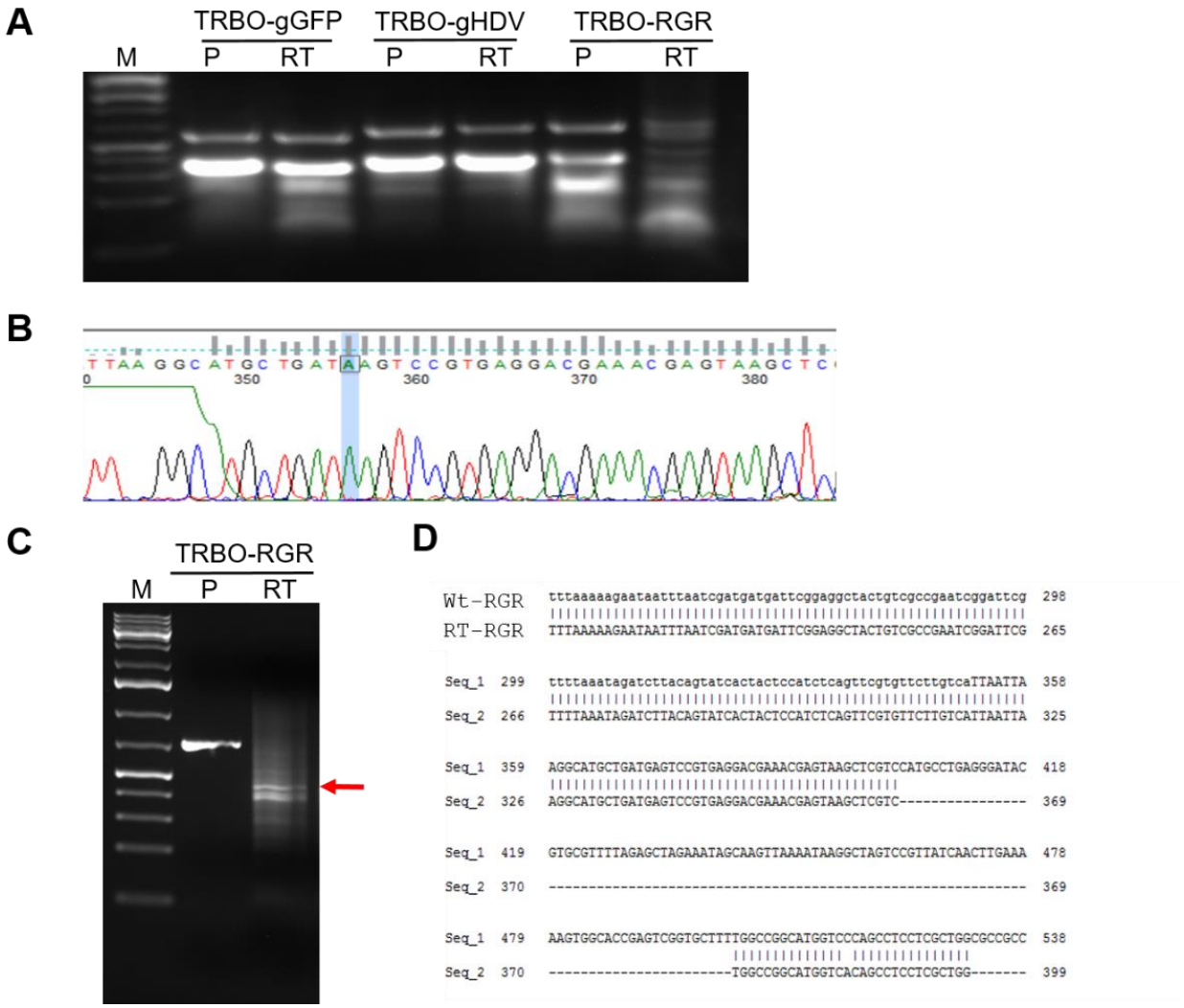
FIGURES



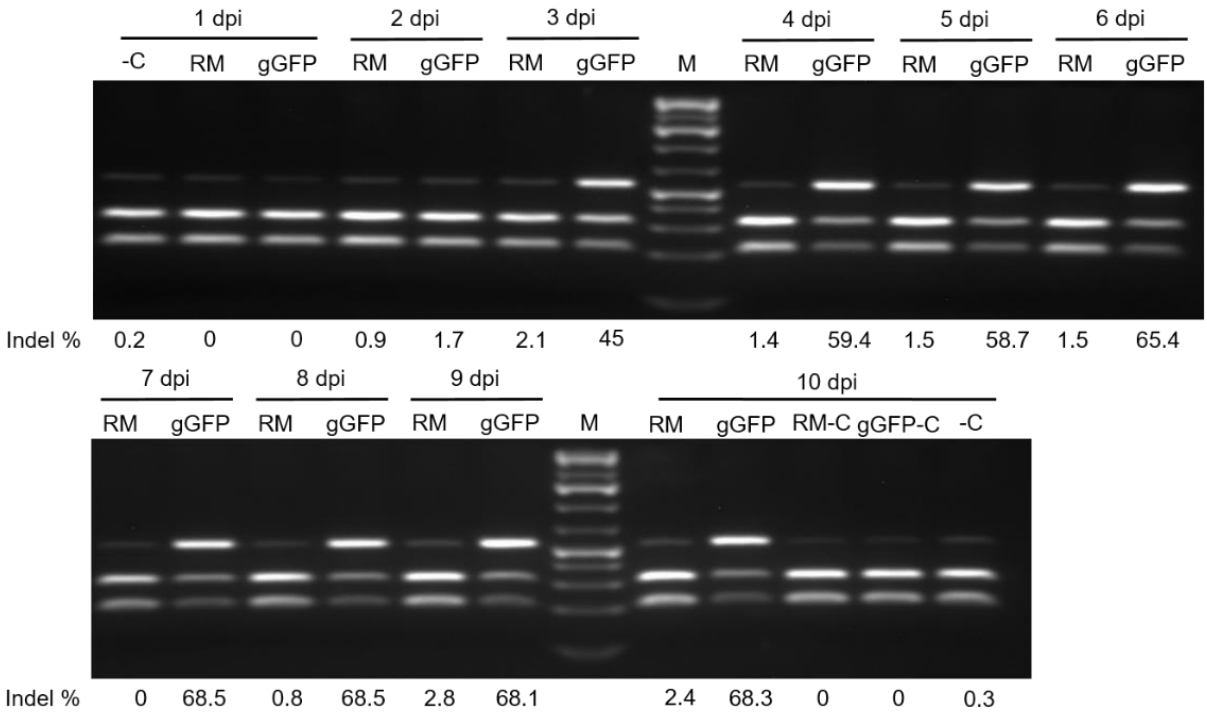
A-1. Predicted subgenomic RNAs used for *in vitro* Cas9 nuclease activity assays. A) T7 primer amplification genomic location for Plasmid and Reverse transcriptase PCR reactions. **B)** Denaturing PAGE urea RNA gel. 1) GFP-3′gGFP RNA template from TRBO-G-3′gGFP 2) subgenomic RNA gGFP-gAGO1 3) subgenomic RGR 4) subgenomic HDV 5) subgenomic gGFP 6) gGFP clean guide (+ control). 100 nt sgRNA red arrow.



A-2. Cas9/TRBO-gGFP indel conformation. A) Sanger sequencing from *Bsg*I undigested *mgfp5* amplicons taken from Cas9/TRBO-gGFP infiltrated plants. **B)** Surveyor nuclease assay of *mgfp5* amplicons from 16c *N. benthamiana* infiltrated leaves used in **Figure 2.1D**.



A-3. TRBO-sgRNA delivery *in planta*. A) *In vitro* T7 transcription reactions using plasmid (P) and RT-PCR products (RT) from 14 dpi 16c *N. benthamiana* plants co-infiltrated with pHcoCas9 and either TRBO-gGFP, TRBO-gHDV, or TRBO-RGR**.** Each RT-PCR band was sequenced to verify the retention of the initial construct. **B)** A sequenced clone from a RT-PCR amplicon from TRBO-RGR. Highlighted is a point mutation within the HH ribozyme of RGR. **C)** TRBO-RGR only infiltrated tissue RT-PCR (RT) and plasmid template used for the PCR control (P). Red arrow points to the RT product that was cloned and sequenced. **D)** Sequence cloned amplicon from **C** (RT-RGR) aligned to the original TRBO-RGR plasmid sequence (wt-RGR)



A-4. Indel percentage measured over 10 days. 16c plants were co-infiltrated with pHcoCas9 and either RM-gGFP (RM) or TRBO-gGFP (gGFP). RM-GFP and TRBO-gGFP are as described in **Figure 2.4A**. pHcoCas9 infiltrated leaves were sampled at 1 and 10 dpi as the negative control (-C). Additionally infiltrated leaves using only RM-gGFP (RM-C) and TRBO-gGFP (gGFP-C) were sampled at 10 dpi and used as negative controls.

A-5. *NbAGO1* paralogs *NbAGO1-H* (KR942296) and *NbAGO1-L* (KR942297) sequence alignment. The red box indicates the spacer target sequence used for gAGO1-containing constructs.

sgRNA sequences (5′→3′)

RGR (HH-gGFP-HDV)

GGCATGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCATGCCTGAGGGATACGTGCGTTTTAGAGCTAGAAATAGCAAG
TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTCCCAGCCTCCTCGCTGG
CGCCGCCTGGGCAACATGCTTCGGCATGGCGAATGGGAC

HDV (gGFP-HDV)

CATGCCTGAGGGATACGTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
GAGTCGGTGCTTTTGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGCCTGGGCAACATGCTTCGGCATGGCGAATGGGAC

gGFP

CATGCCTGAGGGATACGTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
GAGTCGGTGCTTTT

gAGO1

CTCATCGTGGAAATATGCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
GAGTCGGTGCTTTT

gGFPgAGO1

CATGCCTGAGGGATACGTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
GAGTCGGTGCTTTTGCCTCATCGTGGAAATATGCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
TTGAAAAAGTGGCACCGAGTCGGTGCTTTT

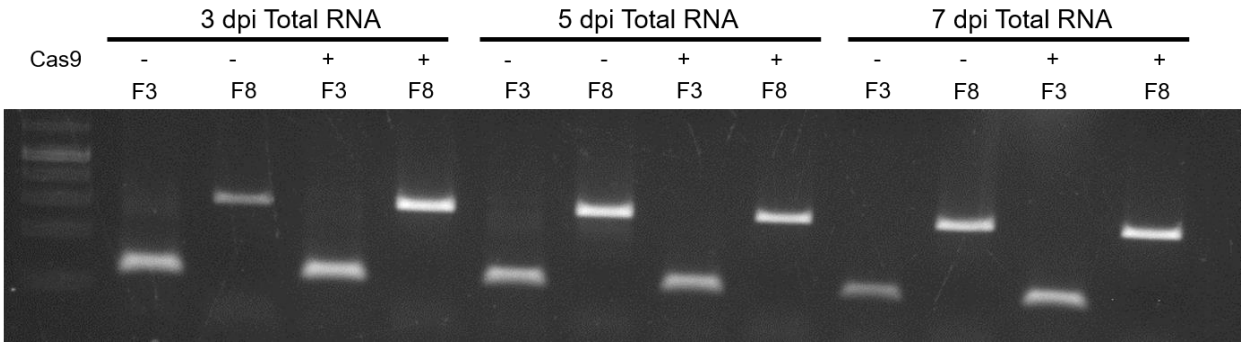
TRBO-G-5′gGFP

CATGCCTGAGGGATACGTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
GAGTCGGTGCTTTTTTAATATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATG
TTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAAGCTTACACTTAAATTTATTTGCACTAC
TGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGA
AACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAG
ACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACAT
TCTCGGACACAAACTCGAGTACAACTATAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACT
TCAAAATTCGCCACAACATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTC
CTTTTACCAGACAACCATTACCTGTCGACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGA
GTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAATAAACTAG

TRBO-G-3′gGFP

ATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTC
TGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAAGCTTACACTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCA
TGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAAACGGCATGACTTTTTCAA
GAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGT
TTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTCGAG
TACAACTATAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACAT
TGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATT
ACCTGTCGACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGG
ATTACACATGGCATGGATGAGCTCTACAAATAAACTAGGCCATGCCTGAGGGATACGTGCGTTTTAGAGCTAGAAATAGCAAGTTA
AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT

A-6. Sequences used for guide RNA delivery tools used in this study. The red text indicates the hammerhead ribozyme sequence. The blue letters represent the hepatitis delta virus ribozyme sequence, and black letter represent the sgRNA sequence. The purple text indicate “scar” nucleotides from the restriction digest to linearize TRBO-G before Gibson assembly. The green text indicates the *gfpc3* sequence.



A-7. 16c plants total RNA expression time course from TRBO-G-3′gGFP infiltrated leaves with and without Cas9. Total RNA was sampled at 3, 5 and 7 dpi from 16c tissue infiltrated with TRBO-G-3′gGFP both with and without pHcoCas9. Total RNA was assayed for processed and unprocessed sgRNAs by using the F3 and F8 forward primer sets, respectively to detect different portions of the sgRNA containing TRBO transcripts. Primer binding locations are previously discussed in **Figure 3.2B**.

>U6-gGFP
CTTTTTTTCTTCTTCTTCGTTCATACAGTTTTTTTTTGTTTATCAGCTTACATTTTCTTGAACCGTAGCTTTCGTTTTCTTCTTTTTAACTTT
CCATTCGGAGTTTTTGTATCTTGTTTCATAGTTTGTCCCAGGATTAGAATGATTAGGCATCGAACCTTCAAGAATTTGATTGAATAAAACATC
TTCATTCTTAAGATATGAAGATAATCTTCAAAAGGCCCCTGGGAATCTGAAAGAAGAGAAGCAGGCCCATTTATATGGGAAAGAACAATAGTA
TTTCTTATATAGGCCCATTTAAGTTGAAAACAATCTTCAAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCTGAGTTTATATACAGCTAGA
GTCGAAGTAGTGATTGCATGCCTGAGGGATACGTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTA*

>U6-GFP-gGFP
CTTTTTTTCTTCTTCTTCGTTCATACAGTTTTTTTTTGTTTATCAGCTTACATTTTCTTGAACCGTAGCTTTCGTTTTCTTCTTTTTAACTTT
CCATTCGGAGTTTTTGTATCTTGTTTCATAGTTTGTCCCAGGATTAGAATGATTAGGCATCGAACCTTCAAGAATTTGATTGAATAAAACATC
TTCATTCTTAAGATATGAAGATAATCTTCAAAAGGCCCCTGGGAATCTGAAAGAAGAGAAGCAGGCCCATTTATATGGGAAAGAACAATAGTA
TTTCTTATATAGGCCCATTTAAGTTGAAAACAATCTTCAAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCTGAGTTTATATACAGCTAGA
GTCGAAGTAGTGATTGATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCA
CAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAAGCTTACACTTAAATTTATTTGCACTACTGGAAAACTACCTGTTC
CATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAAACGGCATGACTTTTTCAAGAGTG
CCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATA
CCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTCGAGTACAACTATAACTCACACA
ATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATTGAAGATGGATCCGTTCAACTAGCAG
ACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCCCTTTCGAAAGATC
CCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAATAAGCCATGC
CTGAGGGATACGTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTT
TTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTA*

>U6-GFP-sfRNA-gGFP
CTTTTTTTCTTCTTCTTCGTTCATACAGTTTTTTTTTGTTTATCAGCTTACATTTTCTTGAACCGTAGCTTTCGTTTTCTTCTTTTTAACTTT
CCATTCGGAGTTTTTGTATCTTGTTTCATAGTTTGTCCCAGGATTAGAATGATTAGGCATCGAACCTTCAAGAATTTGATTGAATAAAACATC
TTCATTCTTAAGATATGAAGATAATCTTCAAAAGGCCCCTGGGAATCTGAAAGAAGAGAAGCAGGCCCATTTATATGGGAAAGAACAATAGTA
TTTCTTATATAGGCCCATTTAAGTTGAAAACAATCTTCAAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCTGAGTTTATATACAGCTAGA
GTCGAAGTAGTGATTGATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCA
CAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAAGCTTACACTTAAATTTATTTGCACTACTGGAAAACTACCTGTTC
CATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAAACGGCATGACTTTTTCAAGAGTG
CCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATA
CCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTCGAGTACAACTATAACTCACACA
ATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATTGAAGATGGATCCGTTCAACTAGCAG
ACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCCCTTTCGAAAGATC
CCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAATAAGCTGTCA
GGCCTGCTAGTCAGCCACAGTTTGGGGAAAGCTGTGCAGCCTGTAACCCCCCCAGGAGAAGCTGGGAAACCAAGCTCATAGTCAGGCCGAGAA
CGCCATGGCACGGAAGAAGCCATGCTGCCTGTGAGCCCCTCAGAGGACACTGAGCATGCCTGAGGGATACGTGCGTTTTAGAGCTAGAAATAG
CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGG
CATTA*

>VIGS tool best prediction(Niben101Scf26315 and Niben101Scf01105)
TGTCAGGGCTTCATCGATCTTCTAGCGCTCCTTCGAAAAATGGCCTTCCTCCTCAGGAACTCCTTGACGATCTTTGCAGTCGGTTTGTTTTGA
ACGTGCCAAAAGAAGATCAGCAATCATTTGAGAGGATTTTATTTCTTGTGGAGTATGCACATTGGTTTTATGAAGACAATTCTGTAGAAAAAA
ATCCATCATTGAAGTCTTTCACTCTCAAGGAGTTTACATCCTTAATGTTTAACAGTTGTGATGTTTTAAAACCATATGTTCCTCACATAGATG
ACATATTCAAGGACTTCACTTCTTACAAGGTTCGAGTTCCTGTAACTGGGGCCATCATCCTTGATGAAACTTTTGAAAGGGTCTTGGAAGAAA
CAGGATTTGATGTATCAAAACTTCTTC*

A-8. Sequences of new constructs used in Chapter III. The blue text indicates U6 promoter sequence, the green text indicates GFP coding sequence, and the red text indicates short flavaviral RNA sequence (sfRNA). Black text for the "U6" containing constructs, not VIGS construct, corresponds to the gGFP sequence. VIGS tool best prediction is *Nbdcp2* sequence cloned into the TRV-00 construct for silencing assay.