Development of a pokeweed mosaic virus vector as a molecular tool to study host protein function in *Phytolacca americana*

Alexander Klenov

A dissertation submitted to the Faculty of Graduate Studies in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Graduate Program in Biology
York University
Toronto, Ontario

December 2020

Abstract

American pokeweed (*Phytolacca americana*) is a plant known for both the synthesis of pokeweed antiviral protein (PAP) and the hyperaccumulation of heavy metals. Progress in understanding the factors behind viral resistance and hyperaccumulation has been slow due to lack of genetic tools to study pokeweed proteins *in planta*. No transgenic pokeweed plants have been generated and the plant is known to produce phytolaccosides that inhibit *Agrobacterium* infection, one of the main tools for genetically engineering plants. Therefore, an infectious clone of Pokeweed mosaic virus (PkMV) was developed as the precursor of a viral vector that would allow the study of host protein function. PkMV is a potyvirus of the family *Potyviridae* and its 9.5 Kb single-stranded RNA genome shares the typical organization of potyviruses, in which a single polyprotein is cleaved by virally encoded proteases into 10 distinct proteins, with the 11th being derived from a frame-shift product. PkMV was the ideal chassis to build an infectious clone to study pokeweed as it infects the plant in the wild despite pokeweed producing a potent antiviral protein. Here, the assembly of the first infectious clone of PkMV is described and its ability to both infect pokeweed and deliver a foreign protein into the plant is demonstrated. Instability of the PkMV clone was encountered during its generation, due to cryptic bacterial promoters within its genome. A combination of sequence independent cloning techniques and assembly in *Agrobacterium* proved sufficient to overcome this challenge. Further manipulation in *Agrobacterium* was a limiting factor in modifying the PkMV clone, so a method to detect cryptic promoters throughout the entire viral genome was developed, which when silenced allowed recovery of an infectious clone that is readily manipulated in *E. coli*. This method relied on generation of a random library of PkMV fragments that were integrated into an eGFP or chloramphenicol reporter construct. Application of the low-throughput method detected

promoter activity by producing eGFP fluorescence. The high-throughput next generation sequencing variant sequenced colonies that survived high concentrations of chloramphenicol due to transcription of a chloramphenicol resistance gene. Both methods produced putative promoter regions (PPRs) containing a cryptic bacterial promoter. The +1 transcription start site (TSS) of each PPR was determined with template switching oligo (TSO)-5'RACE, which allowed accurate detection and silencing of the cryptic promoter. Silenced promoters in the PkMV genome restored stability of the viral clone in *E. coli*, permitting further manipulation. This work not only provides a straightforward method to detect and silence cryptic promoters in large viral genomes, but also begins to elucidate the mechanisms of plasmid toxicity in *E. coli*.

Acknowledgements

Thank you to my supervisor, Dr. Kathi Hudak, for giving young Alex a chance all those years ago, it changed the course of my life forever in the best way. I learned how to be a good scientist and teacher from you, as well as how to be resilient and independent. Thank you for your support and encouragement of all my creative endeavours, no matter the mess.

Thank you to my parents for bringing me to Canada so I could have a good life. You never gave up on me even during tumultuous times. Thank you, Dad, for your diplomatic skills and sense of humor. Thank you, Mom, for your philosophy, attention to detail and for every single warm meal. Thank you to Olga, for being the supportive big sister I always wanted.

Thank you to all the lab mates who have come and gone. Kass for his biochemistry expertise and infectious laugh I still carry to this day. Kira for emotional support, reminding me when things are due and always being there in the toughest times. Camille for your zest for life, enthusiasm for baking and for small rodents. Jen, for your steady attitude, your balance and debates which broadened my world view.

Thank you to the shoulders of giants on which we all stand upon.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3'RACE - 5’ Rapid Amplification of cDNA Ends

5'RACE - 3’ Rapid Amplification of cDNA Ends

6K1 - 6 KDa protein 1

6K2 - 6 KDa protein 2

ALSV - Apple latent spherical virus

AMLV-RT - Avian myeloblastosis virus reverse transcriptase

AR - Arkansas

ATCC - American Type Culture Collection

AtFT - Arabidopsis Flowering Locus T protein

BAM - Binary alignment map

BNYVV - Beet necrotic yellow vein virus

BYMV - Bean yellow mosaic virus

CI - Cylindrical inclusion

CMV- Cucumber mosaic virus

CP - Coat protein

CPMV - Cowpea mosaic virus

CPV- Canine parvovirus

CRISPR - Clustered regularly interspaced short palindromic repeats

CaMV - Cauliflower mosaic virus

ChIP-Seq - Chromatin immunoprecipitation sequencing

DNA - Deoxyribonucleic acid

GFP - Green fluorescent protein

GUS - B-glucoronidase

HC-Pro - Helper component protease

HIV-1 - Human immunodeficiency virus 1

IGV - Integrative Genomics Viewer

INPACT - In Plant Activation

KDa - Kilodalton

Kb - Kilobase

LB - Luria broth

MMLV-RT - Moloney murine leukemia virus reverse transcriptase

MWL - Molecular weight ladder

NCBI - National Center for Biological Information

NGS - Next generation sequencing

NI - No insert

NIa - Nuclear inclusion a

NIb - Nuclear inclusion b

NOS - Nopaline synthase

OD - Optical density

ORF – Open reading frame

P1 - Protease 1

P3 - Protease 3

PAP – Pokeweed antiviral protein

PCR - Polymerase chain reaction

PPR – Putative promoter region

PPV - Plum pox virus

PTGS - Post-transcriptional gene silencing

PVX - Potato virus X

PVY - Potato virus Y

PkMV – Pokeweed mosaic virus

Pos - Positive

RDRP - RNA dependent RNA polymerase

RE - Restriction enzyme

RHDV - Rabbit hemorrhagic disease virus

RIP - Ribosome inactivating protein

RNA - Ribonucleic acid

RT - Reverse transcriptase or transcription

RT-PCR - Reverse transcriptase PCR

SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SLIC - Sequence and ligation independent cloning

Scr - Scrambled

T-DNA - Transfer DNA

TEDA - T5 Exonuclease DNA assembly

TFL1 - Terminal Flower 1

TMV - Tobacco mosaic virus

TRBO - TMV RNA-Based overexpression

TRV - Tobacco rattle virus

TSO-5'RACE – Template switching oligo 5' rapid amplification of cDNA ends

TSS – Transcription start site

TVMV – Tobacco vein mottling virus

TYDV - Tobacco yellow dwarf virus

TuMV - Turnip mosaic virus

UTR - Untranslated region

VIGS - Virus induced gene silencing

VPg - Viral protein genome-linked

VRC – Viral replication complex

WT - Wild-type

YEP - Yeast extract peptone

ZYMV - Zucchini yellow mosaic virus

bp - base pair

cDNA - Complementary DNA

crRNA – CRISPR RNA

dsDNA - Double stranded DNA

dsRNA - Double stranded RNA

eGFP - Enhanced GFP

gDc - Glycoprotein D

mAb - Monoclonal antibody

nm - nanometer

sgRNA - Small guide RNA

siRNA - Small interfering RNA

ssDNA - Single stranded DNA

Chapter 1

Introduction

1.1 – *Phytolacca americana*, American pokeweed

Phytolacca americana, commonly referred to as pokeweed, is a perennial plant native to South America and eastern North America. Pokeweed is a member of the order Caryophyllales which contains agriculturally relevant species such as sugar beet and quinoa. The plant can reach heights of 3m in the wild, with leaves 30cm long and 15cm broad (Botta-Dukat et al., 2008). Its seeds are poisonous to mammals due to presence of phytolaccatoxins (Driver & Francis, 1979), however they are readily dispersed by birds (Ning et al., 2017).

Pokeweed is primarily known for one of its endogenous proteins, pokeweed antiviral protein (PAP), which exhibits antiviral activity against plant and animal viruses when expressed exogenously (Lodge et al., 1993; Rajamohan et al., 1999). PAP has been shown to remove purines from viral RNA, rendering the RNA ineffective for replication and translation (Rajamohan et al., 1999). PAP is a Type I Ribosome Inactivating Protein (RIP) that removes a purine from the large ribosomal RNA, causing inhibition of protein synthesis (Mansouri et al., 2006). Upon damage to the plant leaf, PAP is released from the apoplast where it is stored, into the cytoplasm to inactivate viral RNAs and ribosomes to limit the spread of infection. It would be advantageous to express PAP in crop plants to protect them from viruses.

Another notable trait of pokeweed is that it is a hyperaccumulator of heavy metals such as cadmium and zinc (McBride & Zhou, 2019). Pokeweed can grow in soil contaminated with heavy metals, such as mine tailings, and sequesters the metals within its tissues. The plants can then be harvested and disposed of in a safe manner. This trait combined with the significant amount of biomass pokeweed can accumulate means pokeweed is particularly suited for phytoremediation.

1.2 - Potyviruses

Potyviridae is a large family of plant viruses, whose members include potato virus Y and plum pox virus which cause significant agricultural losses yearly (Reddy et al., 2009). Potyviruses are spread by a wide variety of aphid, whitefly and mite species, which makes containment of viral spread difficult (Wylie et al., 2017). Potyvirus particles are filamentous, 800 nm long, and contain a 10 kb, positive-sense RNA genome that is covalently linked to viral genome-linked protein (VPg) at the 5' end and is also polyadenylated at its 3' end. About 2000 copies of capsid protein (CP) enclose the viral genome of each particle (Shukla & Ward, 1989). The genome of a potyvirus encodes a polyprotein that is subsequently cleaved into 10 individual proteins by virally encoded proteases.

From the most N-terminus, these proteins include P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP. Additionally, an 11th potyviral protein has been described, known as P3N-PIPO, which is translated through a +2 frame shift (Chung et al., 2008). P1 protease autocatalytically cleaves itself from the viral genome and is thought to function in genome replication and host silencing suppression (Pasin et al., 2014; Verchot & Carrington, 1995). NIa-Pro is the main viral protease and whose primary function is processing the viral polyprotein. HC-Pro is responsible for aphid transmission by bridging the aphid stylet and CP of the viral particle (Peng et al., 1998). As well, HC-Pro is a potent post transcriptional gene silencing (PTGS) suppressor (Kasschau et al., 2003). CI functions as a helicase in tandem with NIb, a viral RNA dependent RNA polymerase (RdRp) in the viral replication complexes (VRCs) formed by the virus (Merits et al., 1998). VPg is a multifunctional protein due to its intrinsically disordered regions and has been implicated in translation initiation and viral movement (Charon et al., 2016; Michon et al., 2006; Rajamäki & Valkonen, 2002). To add to the complexity, several cleavage

intermediates of the polyprotein such as VPg-Pro/6K2-VPg-Pro are observed *in planta* and are mediated by the strength of the proteolytic cleavage sites between them (Wan et al., 1998).

1.3 – Pokeweed mosaic virus

Pokeweed mosaic virus (PkMV) is a potyvirus that infects American pokeweed, *Phytolacca americana*. PkMV was an early plant virus to be described (Allard, 1918) and was shown through electron microscopy to form pinwheel inclusions in host cells, typical of potyviruses (Kim & Fulton, 1969). Recently, PkMV isolates from New Jersey, Arkansas, Maryland, Mississippi, and Pennsylvania have been sequenced. PkMV is 9512-9516 bp long (Aboughanem-Sabanadzovic et al., 2019), with a genome organization typical of potyviruses (Figure 1). Phylogenetic analysis showed that PkMV belongs to the potyvirus genus. It is curious that viruses such as PkMV or cucumber mosaic virus (CMV) can infect pokeweed (Davino et al., 2012), which produces PAP. PAP has been previously shown to be effective against both plant and animal viruses when expressed exogenously (Lodge et al., 1993).

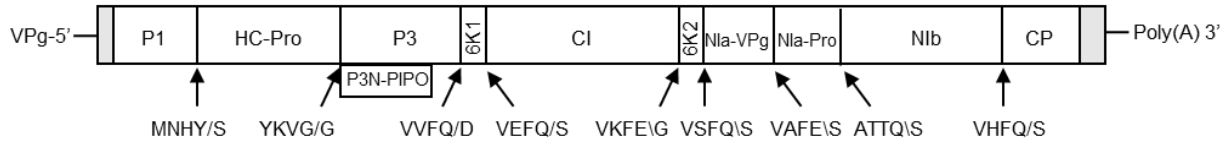

Figure 1 – PkMV genome organization. The PkMV-AR RNA genome spans 9516 nucleotides and shares the common potyvirus genome organization. The viral RNA is translated by host machinery into a single polyprotein, which is processed by virally encoded proteases.

1.4 – Infectious plant virus vectors

RNA virus vectors are DNA plasmids that contain the cDNA copy of the RNA virus genome with a promoter and terminator, allowing transcription of the viral genome when introduced into a host. Once the initial hurdle of generating a viral vector from a particular virus has been overcome, it is relatively simple to modify the viral genome with modern cloning techniques. Plants can be inoculated simply with mechanical inoculation or particle bombardment of plasmid DNA or RNA *in vitro* transcripts (Gal-On et al., 1995; Kim et al., 2003; Seo et al., 2009). Insertion of the viral genome into the T-DNA plasmid of *Agrobacterium* and subsequent infiltration has also been used to infect plants (Liu & Lomonossoff, 2002). After inoculation, the virus is naturally suited to systemically infect the entirety of the plant.

Since the very first tobacco mosaic virus (TMV) vector was introduced (Takamatsu et al., 1987), several viral vector "systems" have been developed to introduce a foreign sequence into a virus. The chosen system is based on what alterations a viral genome will tolerate and its intended purpose. As well, viral vectors can be categorized as either intact, containing the full viral genome, or deconstructed, where part of the genome is removed and must be supplied exogenously (Liu & Lomonossoff, 2002). Applications of plant viral vectors fall into three broad categories, either modifying host gene expression, tracking viral infectivity and biology or the overproduction of useful proteins (Abrahamian et al., 2020). The study of a host gene function can be accomplished through virus-induced gene silencing (VIGS), wherein a fragment of the target mRNA inserted into the viral genome mediates small interfering RNA (siRNA) knockdown of the target gene (Zhang et al., 2017). CRISPR-Cas9 based gene editing has been adapted to viral vectors, where the Cas9 nuclease and small guide RNA (sgRNA) are delivered simultaneously into the plant with a tobacco rattle virus based vector to knock down host genes

(Ali et al., 2015). Host gene expression can also be modulated by overexpression of a protein or transcription factor expressed from the viral genome (Yamagishi et al., 2014). The overproduction of proteins has also been used for the generation of antibodies and vaccine antigens (Legocki et al., 2005).

1.5 - Viral vector systems

Intuitively the simplest method of introducing a foreign sequence into a virus is to fuse it in-frame to a viral protein. As the viral genome is translated, the foreign protein is produced fused to the viral protein. This method assumes that having a fused protein does not significantly hamper the viral protein function. A viral vector of the potyvirus plum pox virus (PPV), was constructed and an epitope of canine parvovirus (CPV) was fused to CP (Fernández-Fernández et al., 1998). The virus was able to infect and accumulate to wild-type levels, and the viral particles purified from infected plants were able to induce an immunogenic response to CPV in mice. Cowpea mosaic virus (CPMV) has also been used for producing viral particles studded with epitopes of several animal viruses. The placement for optimum immunogenicity was determined to be within the βB-βC loop of the small subunit protein (Taylor et al., 2000). However, reports of viral instability and reduced yields spurred research determining that insert length and isoelectric point (pI) greatly affected the viability of the vector (Porta et al., 2003).

Some plant viruses have stringent size limits of their genomes, with foreign fragments being truncated through multiple viral passages (Choi et al., 2019). In these cases, the only option is to replace a non-essential protein with the foreign protein of interest. Cauliflower mosaic virus (CaMV) of the *Caulimovirus* genus is a double stranded DNA virus that exists in a circularized form in its host. It was found early on that a CaMV vector would only tolerate a 250 bp insertion and still be viable (Gronenborn et al., 1981). The only stable CaMV vectors

described are those that replaced the second open reading frame, responsible for aphid transmission (Futterer et al., 1990). Tobacco rattle virus (TRV), a *Tobravirus*, proved to be far more amenable to gene replacement vector construction. TRV has a bi-segmented RNA genome, with coat protein being the only essential gene on the second RNA segment, while genes responsible for nematode-facilitated transmission were disposable (Ratcliff et al., 2008). The vectors were used to demonstrate a popular application of viral vectors, gene silencing through siRNA-based interference. Green fluorescent protein (GFP) which had been stably transformed in tobacco was silenced with the TRV vector. The TMV vector by Takamatsu et al. was the first plant viral vector made (Takamatsu et al., 1987). While the clone was infectious, the replacement of the coat protein gene proved detrimental to the virus, which produced smaller lesions than wild-type virus. A side benefit of gene replacement vectors is that biocontainment can be significantly easier if the transmission genes are removed (Lindbo, 2007b).

The most common viral vectors have the foreign gene expressed alongside the full complement of viral genes. Potyviral vectors are particularly amenable to the addition of genes. The potyviral RNA genome codes for a single polyprotein which is cleaved at proteolytic sites. Artificial proteolytic sites were used effectively for turnip mosaic virus (TuMV) where GFP and β-glucuronidase (GUS) were inserted between various potyviral genes. Both reporter genes were initially expressed from the single vector, however 15 days post-inoculation, the GUS gene was replaced by a shorter sequence. To date, up to three foreign proteins can be inserted into a potyviral vector with a minimum of instability (Kelloniemi et al., 2008). However, the triple-insert vector multiplied at half the rate of the double-insert vector, suggesting that there are limits on genome size, perhaps due to instability of the viral RNA or protein capsid. Researchers found that the viral particle size increased with insertion of each consecutive coding region.

Filamentous viruses such as potyviruses may be less restricted in their maximum genome size when compared to their icosahedral brethren.

The aforementioned viral vector designs are known as full or intact vectors, as most or all of the original viral genes are present and functional. While being conceptually simple to design, full viral vectors have several disadvantages depending on the application (Abrahamian et al., 2020). If the vector is designed for mass protein production, there are viral genes that are either not necessary or could lower yield. As well, there are biocontainment issues inherent to delivering a fully functional virus. To address this issue, deconstructed vectors have been developed with a substantial part of the genome missing which must be supplemented in other ways. The TMV RNA-based overexpression (TRBO) system is based on a TMV vector with the CP gene removed, cloned into a T-DNA vector. Wild-type TMV is dependent on CP for systemic movement (Ryabov et al., 1999), therefore a CP deletion mutant of TMV inoculated by rubbing or particle bombardment would only produce localized infection. Researchers found that by deleting CP, transforming TRBO into *Agrobacterium* and vacuum inoculating plants led to very high levels of protein expression (Lindbo, 2007b). In this situation, the loss of systemic movement due to CP deletion was compensated by agro-infiltrating all parts of the plant. The removal of CP, which can make up to 10% of total soluble protein during infection, increased the yield from 1.5g/kg of protein to leaf tissue to 3-5g/kg. Previous work with the TMV vector showed that for a viable infection, a viral silencing suppressor was required to be co-expressed (Lindbo, 2007a). With the CP deletion mutant, the silencing suppressor was not required, suggesting that the CP sequence was preferentially targeted for RNA silencing. As well, no viral particles were produced with this vector.

CMV is a *Cucumovirus* with a tri-segmented RNA genome. Different strains of CMV have different host ranges, which is determined by RNA3 that encodes CP and movement protein 3a. Researchers constructed vectors based on the three genomic RNAs. By inoculating with identical RNA1 and RNA2 vectors while varying the RNA3 vector, it was possible to induce infection in previously immune hosts (Zhang et al., 2006). In this example, the tripartite nature of CMV was advantageous as it allowed swapping of limiting factors.

As an alternative viral delivery method, plants can be stably transformed with a viral genome wherein the plants are infected in every cell (Peyret & Lomonossoff, 2015). The advantages of this method include bypassing the inoculation step, which takes up to 2 weeks for a virus to systemically infect the plant. This strategy can improve the yield considering viral infection rarely infects all cells of a plant, with lower leaves sometimes not showing symptoms. In Plant Activation (INPACT) is one such system, in which the tobacco yellow dwarf virus (TYDV) is deconstructed into two cassettes which are stably transformed in tobacco (Dugdale et al., 2013). One cassette contains the foreign protein under a TYDV promoter, whereas the second cassette contains the TYDV replicase gene under an ethanol-inducible promoter. Upon spraying plants with ethanol, the replicase is induced, which transcribes the replicon. The transcribed replicon is also replicated by the viral replicase. Essentially, it is an inducible, hyper-expressing protein system. The main disadvantage is achieving the stable transformation itself, which is technically difficult and labor intensive. While yields approached 10% total soluble protein, the same yield has been achieved with full viral vectors, albeit with the extra step of infection.

1.6 - Usage of viral vectors

Viral vectors are the method of choice when a high yield of protein is required within a short time frame. This quality could prove extremely useful during epidemics when rapid development and delivery of vaccines are desired. Epitope presentation is a method where an immunogenic peptide is displayed on the surface of viral particles, which are then purified using density centrifugation (Peyret, 2015) or column chromatography (Chen et al., 1990). The caveat for epitope presentation is that some knowledge is required about the morphology of the viral particle in advance, and whether the introduction of extra peptides will destabilize capsid formation. Due to the availability of atomic crystal structures (Lin et al., 1999), CPMV has been investigated for epitope presentation within the βB-βC loop of the S subunit protein as mentioned above. The βE-αB loop (Brennan et al., 1999) of the large (L) subunit protein and the βC′-βC″ loop (Taylor et al., 2000) of the S protein are also viable for epitope presentation. CPMV-based vaccines successfully protected both minks (Dalsgaard et al., 1997) and dogs (Langeveld et al., 2001) against calicivirus and parvovirus infections, respectively. Protection from bacterial pathogens was also successful, with mice protected from *Pseudomonas aeruginosa* infection (Brennan et al., 1999). Unlike the icosahedral structure of CPMV, TMV is a rod-shaped virus, with roughly 2100 CP units arranged in helical symmetry compared to CPMV, which has 60 L and S subunits each (Liu et al., 2005). While theoretically a greater number of subunits would produce a stronger immunogenic response, the packing of CP on TMV is much tighter. Many initial attempts at TMV-based presentation systems failed, with no functional CP produced. Only when researchers inserted a read-through sequence to allow the production of both foreign and wild-type CP was TMV infection successful (Hamamoto et al.,

1993). After optimization of epitope location, the maximum length of introduced epitope was 23 amino acids (Bendahmane et al., 1999).

Producing immunogenic whole proteins *in planta* with viral vectors has also been investigated. This technique has an advantage over epitope presentation, where the maximum size of the foreign peptide is limited to ~110 amino acids unless wild-type CP is supplied (Cruz et al., 1996; Röder et al., 2018). The maximum size of a foreign protein that is not fused to the CP is much larger. Vectors based on potyviruses are well suited to this application, as their filamentous morphology is forgiving to accommodating genomes with larger inserts. The VP60 structural protein of Rabbit hemorrhagic disease virus (RHDV) was inserted between proteolytic sites within PPV (Fernández-Fernández et al., 2001). Purified viral extract was used to immunize rabbits which were challenged with otherwise lethal doses of RHDV. A TMV-based vector was also used to express Bovine herpesvirus type 1 (BHV-1) glycoprotein D (gDc) by insertion into the genome under a frameshift promoter (Pérez Filgueira et al., 2003). An oil-based vaccine derived from plant extract was able to protect cattle to the same level as a commercial vaccine. The same strategy was employed to insert a codon optimized Tat, an HIV-1 regulatory protein into TMV, which was used to infect spinach. When fed to mice, the spinach primed the mice for subsequent HIV-1 immunization (Karasev et al., 2005). Edible vaccines are of interest to the global community as they eliminate the need for transportation and cold storage of currently available vaccines. Several regulatory hurdles such as potential allergenicity in humans and contamination of the food chain must be overcome.

Mass production of expensive proteins and diagnostic enzymes has fewer regulatory issues and is likely to be adopted sooner than edible vaccines. Vitronectin is a cell attachment factor for stem cell production. It is normally produced in animals and the cost of a gram ranges

between 4-5 million dollars. Vitronectin coding region has been successfully inserted into an ethanol-inducible deconstructed TYDV based vector and expressed to levels approaching 10% of total soluble protein. The biotech company Farmacule has adopted this strategy and claims high yield of vitronectin in plants (Rybicki & Martin, 2011). The human tumor specific monoclonal antibody (mAB) a5 was produced in tobacco by co-inoculating plants with both TMV and potato virus X (PVX) (Giritch et al., 2006). Each virus carried the sequence of either the light chain or heavy chain of the antibody. These viruses were chosen because they do not compete and co-localize, unlike an infection of two TMV vectors, which segregate. Researchers found that the purified antibodies were both active and accumulated to 0.5g/kg of antibody per kilogram of plant.

Using plant viral vectors as a reverse genetics tool to study gene function is generally accomplished by triggering the plant defense mechanisms based on post-transcriptional gene silencing (PTGS). By inserting the sequence of a host mRNA into a viral vector and subsequently infecting the host, the endogenous gene will be silenced in a method that is now known as virus-induced gene silencing (VIGS). The first report of VIGS was the use of a hybrid TMV vector that contained the cDNA of phytoene desaturase, a key enzyme in the carotenoid synthesis pathway, under a sub-genomic promoter (Kumagai et al., 1995). Infection with the vector caused an over 50-fold increase in the amount of phytoene produced in the plant. Surprisingly, even plant viruses that produce potent viral silencing suppressors can be used in VIGS. Potyviruses encode for helper component-proteinase (HC-Pro) as part of the polyprotein, which has been shown to bind siRNAs and inhibit PTGS (Shiboleth et al., 2007). However, Potato virus A, a potyvirus, was able to successfully silence a GFP transgene in tobacco for several weeks. Interestingly, insert-less mutants of the infectious clone appeared at late stages of

infection. Apple latent spherical virus (ALSV) is an ideal example of a VIGS vector in that it has a broad host range and is symptomless in most of its hosts (Igarashi et al., 2009), allowing easier interpretation of phenotype without viral symptoms. ALSV was used to simultaneously overexpress *Arabidopsis* Flowering Locus T protein (AtFT) and silence Terminal Flower 1 (MdTFL1-1) in apple seedlings. While inducing no symptoms, the time to first flowering of the seedlings was reduced to less than a year (Yamagishi et al., 2014). An ALSV VIGS vector was also used to protect tobacco against bean yellow mosaic virus (BYMV), zucchini yellow mosaic virus (ZYMV) and CMV. This was accomplished by inserting up to 200 bp of CP sequence of the respective virus into the ALSV vector and infecting the plants (Satoh et al., 2014).

1.7 - Construction of viral vectors

Over 70% of plant viruses have RNA genomes that must first be reverse transcribed into cDNA before cloning into a suitable plasmid vector. Avian myeloblastosis virus reverse transcriptase (AMLV-RT) and moloney murine leukemia virus reverse transcriptase (MMLV-RT) are two reverse transcriptases that have been widely used for this purpose. Reverse transcription requires a reverse primer specific to the viral RNA to initiate synthesis. If the virus has not been sequenced previously, then the first step in construction of a viral vector is sequencing the virus. This is often initiated by a method known as 3' Rapid Amplification of cDNA Ends (3' RACE). The viral RNA is first reverse transcribed with a poly-dT primer that anneals to the polyadenylated viral RNA. After reverse transcription, an adapter oligo is ligated to the 5' end of the nascent cDNA which serves as a landing pad for an adapter primer used in subsequent PCR amplification steps. The amplified PCR product is subsequently cloned into a small bacterial vector and sequenced. The sequenced portion of the viral genome guides design of additional primers within the viral genome. The process is repeated until the entire genome is

sequenced. This sequencing by walking from the 3’ to 5’ end produces an *in-silico* genome that aids further construction (Pasin et al., 2019).

After the viral genome sequence is determined, the appropriate primers can be designed to amplify sections of the virus for assembly. During the early days of viral vectors, assembly was accomplished by the cloning of small portions of the viral genome into bacterial vectors using restriction enzymes (Domier et al., 1989), followed by assembling them sequentially into plasmids. Since a viral infection consists of both viable and non-viable genomes, sequential cloning is risky since there is a chance of amplifying sections of non-viable viral genomes. Population cloning was an improvement as it generated libraries of potentially viable viral sections which were combined and screened in large numbers (Yu & Wong, 1998). This technique was limited by the presence or absence of restriction enzyme cut sites within the viral genome. With the development of sequence-independent cloning techniques like sequence and ligation independent cloning (SLIC; Jeong et al., 2012), Gibson assembly (Gibson et al., 2009) and T5 Exonuclease DNA assembly (TEDA; Xia et al., 2019), viral genomes could be subdivided into arbitrary sized chunks for assembly (Bordat et al., 2015). These techniques rely on fragments with 15-60 bp terminal sequence complementarity, which are digested by exonucleases, producing sticky ends (Figure 2). These sticky ends anneal, the gaps are filled by a polymerase and the nick sealed by a ligase. This technique allows for assembly of 3+ fragments. The greater the number of fragments, the lower the transformation efficiency, therefore construction of a viral clone should be accomplished with as few fragments as possible.

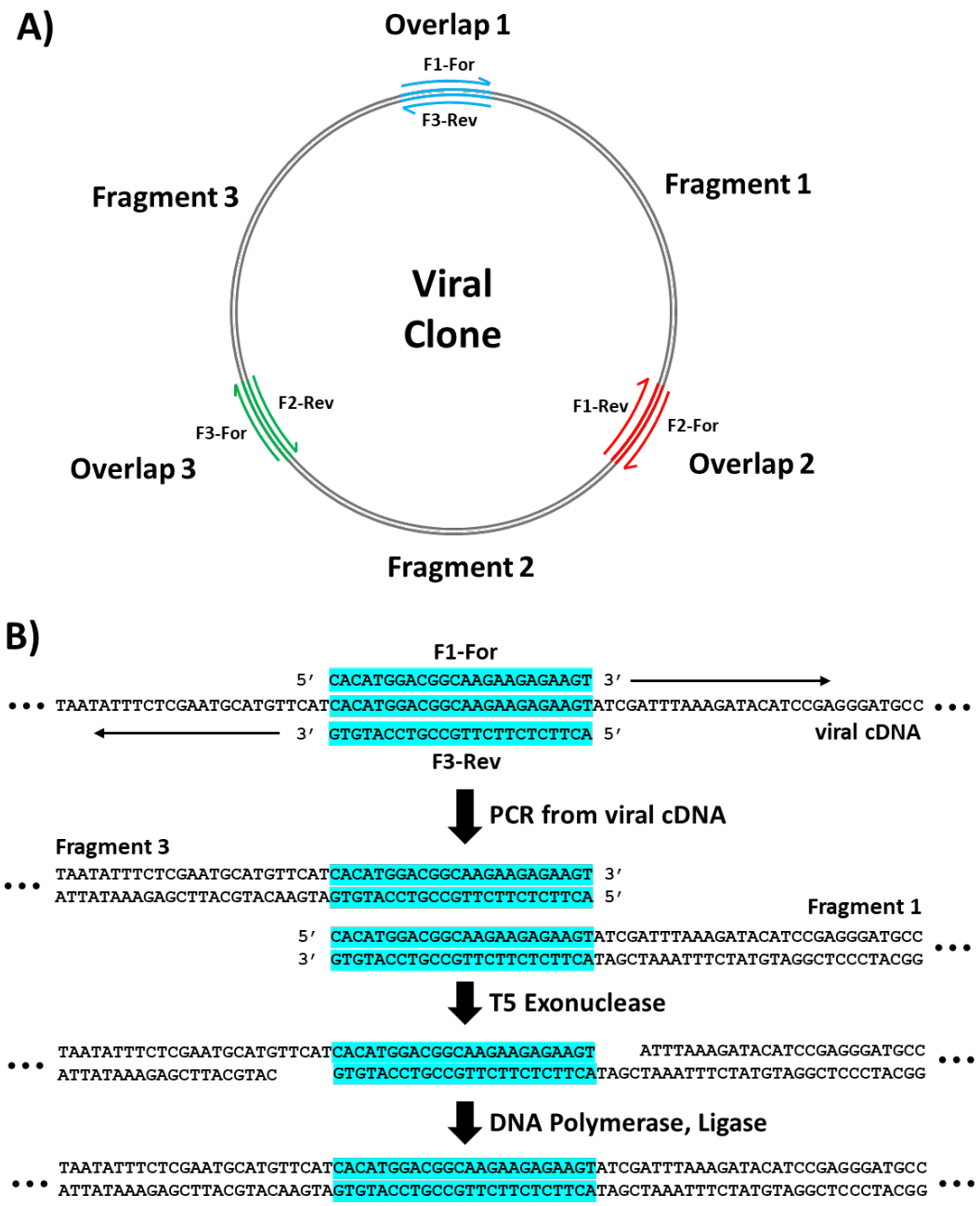

Figure 2 – Gibson assembly of an infectious viral clone. A) An *in silico* representation of the finished viral clone is designed to determine optimal location of overlaps and to select primers. B) Fragments of the viral genome are amplified from viral cDNA with primers containing 15-60 bp overlaps with the adjacent fragment. T5 exonuclease degrades nucleotides from the 5' to 3' direction, which exposes sticky ends where fragments can anneal. The gaps in the dsDNA are filled by a DNA polymerase and the nicks are sealed with a DNA ligase.

Golden Gate assembly is an alternative to overlap-based assembly techniques (Figure 3) (Engler et al., 2009). Golden Gate relies on the activity of Type IIS restriction enzymes (RE) like BsaI-HFv2, which cleave DNA distal to their recognition site. This allows PCR amplification of fragments, which upon cleavage with the RE, have sticky ends that allow ordered assembly into a vector. Due to the distal cleavage of the RE, the enzyme cleaves its own recognition site out of the assembly. Ligation between cleaved fragments cannot be cleaved again, and re-ligated cleavage products are recycled back into the assembly reaction (Engler et al., 2009). Therefore, the Golden Gate reaction will proceed until all cleaved products are correctly assembled, leading to a very high assembly efficiency. This assembly technique is superior to overlap based assemblies due to the much higher transformation efficiency with greater fragment numbers. The biggest disadvantage to this method is that the viral genome and the destination vector must not contain the Type IIS restriction site used in assembly. If the sequence of the virus is known, then several restriction sites can be silently mutated during assembly, alleviating this issue. As well, vectors which do not have BsaI-HFv2 recognition sites, and that allow agroinfiltration are available, such as pLX-B2 (Pasin et al., 2017).

The choice of destination vector for viral clone assembly is also important. If infiltration with *Agrobacterium* will be the primary method of inoculation, the vector must allow manipulation of the clone in *E. coli* while also possessing the required elements that permit *Agrobacterium* to express the plasmid in plants.

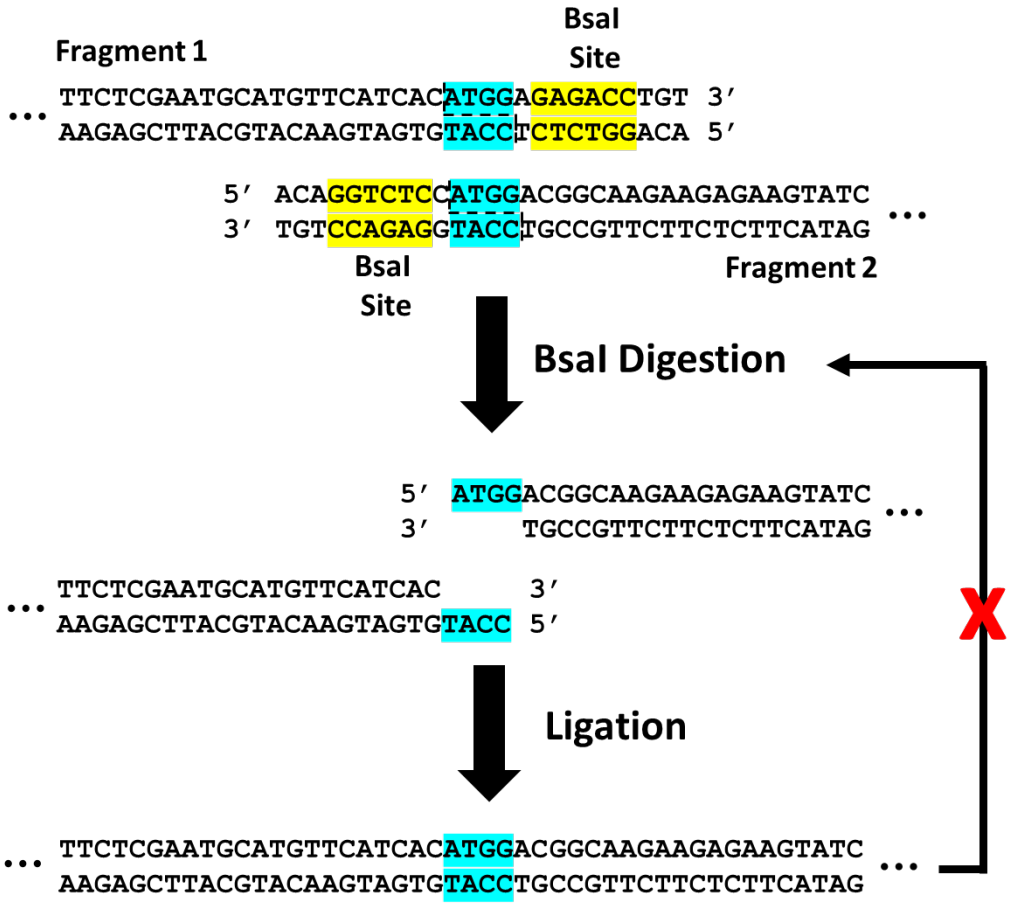

Figure 3 – Golden Gate assembly method. DNA fragments to be assembled are amplified with primers containing a BsaI restriction enzyme recognition site (yellow highlight). BsaI cleaves distal of its recognition site (blue highlight), leaving sticky ends that can then be ligated with T4 DNA ligase. Ligated products cannot be cleaved again as BsaI has removed its own recognition site from the end product.

The most popular of the early binary vectors were the pSoup/pGreen vectors (Hellens et al., 2000). In this configuration, the pGreen plasmid contained the T-DNA to be inserted, with elements that allowed manipulation of the plasmid in *E. coli*. pSoup contained the *vir* genes required for integration of the T-DNA into the plant genome during agroinfilitration, as well as replicase genes to maintain the pGreen plasmid in *Agrobacterium*. This dual vector system was eventually combined into the pPZP vector system that contained T-DNA, replicase genes and origin of replications compatible with *Agrobacterium* and *E. coli* in one package (Szakasits et al., 2007). pPZP was used as a basis for the pCambia series of vectors, which have been the vectors of choice until the recent release of the pLX-B2 vectors that provide the same functionality with a smaller vector size (Pasin et al., 2017).

Regardless of the vector system used, the promoter driving the transcription of the viral genome is most often the CaMV 35S promoter. Additionally, most infectious plant virus clones are assembled with a 20-60 nt poly-(A) tail into the destination vector. The length of poly-(A) tail necessary varies between virus types and method of inoculation. Using particle bombardment of plasmid DNA, a 25nt poly-(A) tail was sufficient for infection of a ZYMV clone (Kang et al., 2016). A poly-A tail of 12 A residues was insufficient for infectivity using *in vitro* transcribed tobacco vein mottling virus (TVMV) RNA, a 37nt poly-(A) produced an infection while a 96nt poly-(A) tail was even more infectious (Domier et al., 1989). Likely the length of the poly-(A) tail in the clone is required for robust replication in the initially infected cells, as *in vivo* the poly-(A) tail is reconstituted regardless of its initial length (Tacahashi & Uyeda, 1999). To terminate transcription of the viral genome, the nopaline synthase (NOS) terminator is most often used. The assembly of the 35S promoter, the genome of the virus, and the NOS terminator in a suitable vector should produce an infectious viral clone.

Verification of infectivity should be conducted after assembly to ensure a functional infectious vector. Initial visual symptoms appear between 7-14 days post infection, usually presenting as mottling and chlorotic veins in newly emerging leaves. Viral RNA and protein can be detected using RT-PCR and western blot, respectively, if an antibody is available to the virus. Viral particles can be isolated using density centrifugation techniques (Berger & Shiel, 1998) to prove their presence in infected plants. Infection resulting from rub inoculation and aphid transmission from infected to healthy plants are also additional pieces of evidence that the newly created clone is behaving like the wild-type virus.

1.8 - Toxicity of viral sequences

The early challenges of constructing viral clones were the lack of sequence independent cloning methods as well as the poor fidelity of polymerases used for PCR. Modern high-fidelity DNA polymerases such as Phusion and Q5 boast 20-100X fidelity when compared to wild-type Taq polymerase (Ricardo et al., 2020), meaning that the chances of encountering crippling mutations in a newly constructed infectious clone are very low. As well, a plethora of sequence independent techniques are now widely available. The last, and perhaps most difficult hurdle to overcome is that long, viral sequences are often toxic to the organism used during assembly.

Toxicity of viral sequences was observed early in the development of viral vectors. During the cloning of beet necrotic yellow vein virus (BNYVV) (Ouillet et al., 1989), researchers were unable to clone RNA2 of the quadripartite plant virus. They had attempted to clone RNA2 in two separate parts, neither one of which was toxic on its own. However, upon combining the cDNA into the full-length RNA2, toxicity was observed. They speculated that one portion of RNA2 contained the peptide toxic to bacteria, while the other had an adventitious prokaryotic promoter driving the expression of the toxic peptide. Animal virus clones have also

been known to induce toxicity when cloned in *E. coli*. Dengue fever virus, a 10 kb ssRNA flavivirus, was recognized early on to be difficult to clone due to toxicity (Lai et al., 1991) and has remained problematic.

To date, the working hypothesis is that toxicity encountered during cloning of large viral sequences is due to the presence of cryptic bacterial promoters (Li et al., 2011; Münster et al., 2018). *E. coli* is the most common organism for the construction of viral clones and is therefore of particular interest when investigating cryptic promoters. *E. coli* RNA polymerase complexes with several sigma factors to recognize various types of promoters under differing environmental conditions. Sigma 70 is the most common of the sigma factors and recognizes the consensus TTGACA-(17 bp spacer)-TATAAT (Djordjevic, 2011). The probability of a perfect Sigma 70 recognition site occurring in a 10 kb viral genome by chance is very low. However, the recognition site can vary from the consensus and still drive high levels of transcription. When examining ~300 experimentally validated *E. coli* promoters, the average number of matches to the perfect consensus sequence was 7.9/12, with matches over 10 being quite rare (Lisser & Margalit, 1993). As well, the length of the spacer between the -10 and -35 box varied between 15-19 bp. Based on a training set of 111 *E. coli* promoters, researchers found that a 250 bp region could contain between 2 and 35 promoter-like signals. Therefore, the construction of infectious viral clones will almost inevitably involve contending with the issue of spurious transcription in *E. coli*.

Numerous strategies to clone toxic viral sequences have been employed. Many *E. coli* strains exist commercially that have been developed for tasks such as protein expression (BL21), cloning of repetitive DNA (STBL) and copy number reduction for cloning toxic proteins (ABLE) (Aubry et al., 2015). Empirically testing the strain of *E. coli* used to generate the

infectious clone can be helpful in overcoming toxicity. For potyviral infectious clones DH10b has become a standard strain and is able to produce full-length clones where DH5a fails (Blawid & Nagata, 2015). Another strategy is to vary the copy number of the viral vector in *E. coli* by choosing low-copy backbones. This was employed in the successful generation of a stable yellow fever virus clone (Bredenbeek et al., 2003).

The most widely adopted method for overcoming toxicity is the insertion of introns into the viral genome during cloning. The intron is thought to interrupt the translation of toxic proteins in *E. coli*, while being spliced out within the plant host, to reconstitute an infectious virus. This method has been employed with animal viruses (Yamshchikov et al., 2001) and more commonly with plant viruses (Johansen, 1996; López-Moya & García, 2000). Intron insertion must be performed empirically to determine a position within the viral genome that stabilizes the clone (Johansen & Lund, 2008). In the case of lettuce mosaic virus (LMV), intron insertion was not necessary to retrieve a stable clone and actually slowed the rate of infection (Yang et al., 1998).

The most direct way to reduce or eliminate toxicity is to find and silently mutate the offending cryptic promoters. In one instance, an almost perfect Sigma 70 promoter was found in an isolate of potato virus Y (PVY) (Ali et al., 2011). When mutated, fully infectious and stable clones were recovered. Unfortunately, finding obvious promoters in a large viral genome is not trivial. One method employed involved breaking down the dengue virus genome into small 300 bp portions and cloning them upstream of a luciferase promoter (Pu et al., 2011). Using this technique, researchers found a promoter in one of the 300 bp fragments and silently mutated it, allowing the 3 kb portion of the dengue fever virus to be cloned. A drawback of this method is that it does not scale well for long RNA viruses like potyviruses or full-length dengue fever

virus, which are both approximately 10 kb long. Moreover, cryptic promoters can exist on either strand of the dsDNA viral clone, doubling the number of search windows.

1.9 – Goals of study

Pokeweed is a plant with significant potential application in medicine, agriculture and bioremediation. Unfortunately, no transgenic pokeweed plants have been generated with knockouts of PAP or other key proteins involved in stress response. Viral vectors are an attractive alternative to the generation of fully transgenic plants since they allow the knock down and overexpression of specific proteins. The goal of this study is to generate an infectious viral clone of PkMV to be used as a molecular tool to study pokeweed proteins *in planta*.

This project is divided into two parts:

1) Construct the first infectious clone of PkMV and demonstrate its infectivity.
2) Stabilize the PkMV clone by developing a high-throughput screening method to detect cryptic bacterial promoters in viral sequences.

Chapter 2:

Complete coding sequence and infectious clone of pokeweed mosaic virus Arkansas isolate

Published manuscript

Klenov, A., & Hudak, K. A. (2018). *European Journal of Plant Pathology*, 152(2), 541–547. https://doi.org/10.1007/s10658-018-1477-9

Contributions: AK and KH conceived the study. AK performed all wet lab work. AK and KH drafted the manuscript. AK and KH edited the manuscript.

All supplementary material referenced in this manuscript is available in the appendix.

2.1 - Abstract

The American pokeweed plant (*Phytolacca americana*) is resistant to several biotic and abiotic stressors. Investigation of the genes involved in defense would be aided by transient or stable over/under expression in pokeweed; however, there are no established protocols for generating transgenic pokeweed plants. Pokeweed is infected by pokeweed mosaic virus (PkMV), a potyvirus with recent genome sequence of some isolates. In this work, the complete coding sequence of the Arkansas isolate was determined and the first infectious clone of PkMV constructed. Agroinfiltration of the clone into pokeweed leaves produced the typical chlorotic mottling and systemic movement through the plant, indicative of infection by PkMV. Viral particles of purified preparations, isolated from plants inoculated with the infectious clone, were morphologically identical to PkMV, and presence of viral RNA and proteins were verified by RT-PCR and immunoblot assay. The clone-generated virus particles were successfully transmitted by aphids to healthy pokeweed plants. Agroinfiltration of the infectious clone encoding the open reading frame of enhanced green fluorescent protein (eGFP) resulted in systemic expression of eGFP and supports the introduction of foreign genes or siRNA in pokeweed. Construction of an infectious clone of PkMV facilitates creation of a viral vector to manipulate expression of pokeweed genes involved in defense, to better understand how pokeweed resists pathogens and abiotic stress.

2.2 – Body

The American pokeweed plant (*Phytolacca americana*) is a perennial herb native to eastern North America and has become naturalized, and often invasive, in parts of Europe and eastern Asia (Bentley et al., 2015). The plant is tolerant of heavy metal contaminated soils, and is a hyperaccumulator of cadmium and manganese (Dou et al., 2009; Zhao et al., 2011). A transcriptome-wide study of cadmium-treated plants showed the upregulation of gene products that chelate and transport heavy metals, characteristic of hyperaccumulators (Chen et al., 2017). Pokeweed is resistant to infection by some viruses and fungi which is due, in part, to synthesis of a ribosome-inactivating protein called pokeweed antiviral protein (PAP) (Endo et al., 1988; Stirpe, 2013). Expression of PAP in heterologous plants such as tobacco limits infection by potato virus X and the fungal pathogen *Rhizoctonia solani* (Wang et al., 1998; Zoubenko et al., 2000)

PAP expression is increased in pokeweed treated with jasmonic acid, a hormone involved in defense against herbivores and necrotrophic pathogens (Neller et al., 2016). Several other genes involved in defense are also upregulated, including pathogenesis related protein, chitinases, proteinases, peroxidases, and terpenoid biosynthesis enzymes. Examining the contribution of these genes to pokeweed hardiness would be assisted by either over- or under-expression; however, no established protocols for transformation of pokeweed exist. Therefore, we turned our attention to cloning a virus that infects this plant, from which a viral vector could be developed for transient transformation of pokeweed.

Pokeweed is infected by pokeweed mosaic virus (PkMV), a potyvirus of the family *Potyviridae*, genus *Potyvirus,* species *Pokeweed mosaic virus*. Recently, the genomes of isolates from Maryland, New Jersey and Pennsylvania were sequenced (Di, 2016; Xu et al., 2012). Here, complete coding sequence and construction of an infectious clone of the Arkansas isolate of PkMV is presented. Agroinfiltration of the clone into pokeweed leaves showed chlorotic mottling, typical of infection with PkMV. Electron micrographs illustrated filamentous particles with dimensional ratios characteristic of potyviruses, indicating that progeny virions were successfully formed in infected pokeweed leaves. In addition, these clone-generated particles were aphid-transmitted to healthy pokeweed plants, and introduction of eGFP cDNA to the clone resulted in systemic expression of the fluorescent protein. Together, these data confirm generation of the first PkMV infectious clone.

Description of sequencing the PkMV-AR isolate and constructing the clone follow, with all primers used in this study listed in Supplementary Table 1. The Arkansas isolate of PkMV was purchased from ATCC (PV-141) and rub-inoculated onto the third true leaves of intact pokeweed plants. After 14 days, uninoculated emerging leaves showing signs of infection with wrinkling along the mid-vein and chlorotic mottling throughout, were harvested for virus particle preparation. Viral particles were collected by sucrose gradient centrifugation, as described in the general potyvirus purification protocol (Berger & Shiel, 1998). To sequence the Arkansas isolate, 3’ RACE was first performed on viral RNA extracted from purified virus preparations, following the GeneRacer protocol (Thermo Fisher Cat#: L150201). The RNA was reverse transcribed with a poly-d(T) adapter primer (primer A1; Supplementary Table 1) and Superscript III reverse transcriptase (Thermo Fisher Cat#: 18080093). PCR was performed with Q5 DNA

polymerase (NEB Cat#: M0491S), a 5’ proximal primer specific to the Pennsylvania isolate (A2) and a primer specific to the adaptor (A3). The PCR product was sequenced and a reverse primer (A4), 3’ proximal from the initial forward primer and based on Arkansas isolate sequence, was designed and used for another PCR with a more 5’ proximal primer (A5) based on Pennsylvania isolate sequence and the original Arkansas isolate cDNA as template. A third round of walking was performed with a reverse primer (A6), 3’ proximal to the second forward primer (A5) and based on Arkansas sequence, and a 5’ primer based on consensus sequence of published PkMV 5’ termini (A7). Several attempts at 5’RACE of PkMV-AR were unsuccessful, likely due to the presence of VPg covalently linked to the 5’ end of the genomic RNA; therefore, the 5’ consensus sequence was used as primer in the third PCR. This primer walking strategy was used to produce three separate overlapping PCR products. Care was taken to avoid designing Arkansas isolate-specific reverse primers close to the Pennsylvania isolate-based forward primers so that *in silico* construction of the viral sequence from the three overlapping PCR products was based on sequence specific to the Arkansas isolate. Apart from 30 nts of the 5’ terminal consensus sequence (the A7 primer), all sequence of the genome is specific to the Arkansas isolate. The resulting sequence of PkMV-AR was used to design primers for subsequent cloning of the viral genome.

To clone the Arkansas isolate, PkMV-AR RNA was reverse transcribed with primer B1 and Superscript III (Thermo Fisher Cat#: 18080093). PCR was performed with two primer pairs (B1, B2; B3, B4) designed to amplify the PkMV-AR cDNA in two fragments with a 30 bp overlapping region. pCambia0305.2, chosen as the destination vector, was amplified by PCR with primers B5 and B6, designed to remove the GusPlus exon and create 30 bp overlapping

sequence complementary to the 5' and 3' fragments of the viral genome. After DpnI treatment and low-melt agarose purification, the vector, and the 5' and 3' fragments of the viral DNA were combined with HiFi DNA Assembly Master Mix (NEB Cat#: E2621S). Gibson assembly, a sequence-independent method, was chosen to clone PkMV-AR and the assembled clone was electroporated directly into *Agrobacterium tumefaciens*, strain Agl1, as described recently (Tuo et al., 2017). Plasmid DNA was extracted from transformed *Agrobacterium* and PCR performed to detect the 5'- and 3'- fragments of PkMV cDNA, with the same primers used for cloning (B1-B4). A flowchart of the cloning strategy is illustrated in Figure 4.

The entire PkMV-Ag01 plasmid DNA was sequenced by Illumina high throughput sequencing, and the 75 bp, paired-end reads were assembled with Geneious software. The assembled sequence aligned perfectly with the sequence initially obtained from primer walking. The nucleotide sequence of PkMV-AR clone was deposited to NCBI (accession# MG189944). PkMV-AR shares only 85% nucleotide identity and 95% protein identity with the Maryland, New Jersey, and Pennsylvania isolates, and is more divergent in sequence than the other three isolates, which share 97% protein identity. PkMV-AR has the genome organization of other potyviruses (Figure 5; Xu et al., 2012), and based on previously published phylogenetic trees of potyvirus polyprotein sequences, PkMV-AR is closely related to tobacco etch virus and potato virus A (Gibbs et al., 2015; Xu et al., 2012). To determine whether the viral clone was biologically active, *Agrobacterium* transformed with PkMV-Ag01 was infiltrated into the third true leaf of intact pokeweed plants.

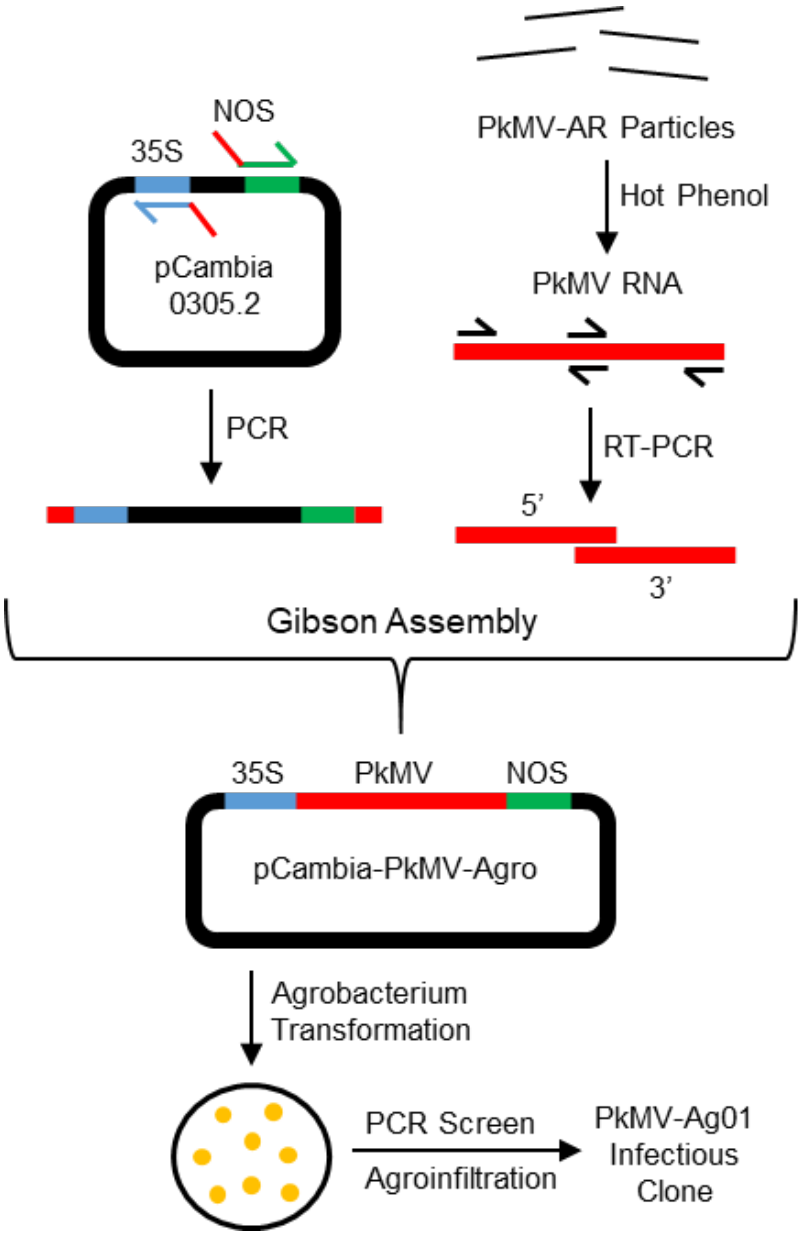

Figure 4 - Cloning strategy for PkMV-AR. Viral RNA was extracted from purified virus preparation of native PkMV-AR and amplified in two fragments, by RT-PCR using virus RNA-specific primers. 5' and 3' fragments were inserted into the pCambia0305.2 vector backbone between the 35S promoter and NOS terminator using Gibson assembly. The plasmid was electroporated into *Agrobacterium*, and colonies were screened for the PkMV insert. Colonies bearing positive clones were sequenced and infiltrated into leaves of pokeweed plants.

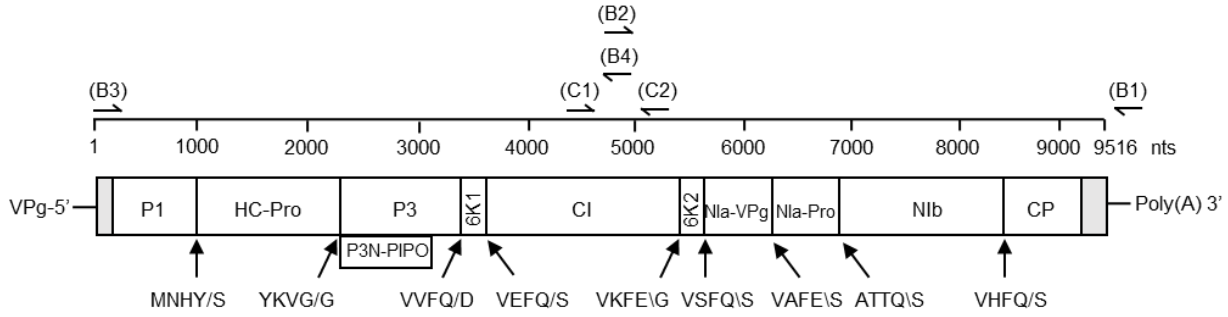

Figure 5 - PkMV-AR genome organization. The PkMV-AR RNA genome spans 9516 nucleotides and shares the common potyvirus genome organization. The viral RNA is translated by host machinery into a single polyprotein, which is cleaved by viral encoded proteases. Arrows indicate amino acid cleavage sites. Shaded boxes represent 5' and 3' UTR, respectively. B1-B4 and C1-C2 indicate location of primers used for cloning and validation of infection, respectively.

After 14 days, non-infiltrated emerging leaves showed wrinkling and chlorotic mottling (Figure 6A; PkMV-Ag01 infected). These symptoms were comparable with positive control plants that were rub-inoculated with purified PkMV-AR virus preparation (PkMV-AR infected). As negative controls, plants were uninoculated or agroinfiltrated with empty vector; these plants did not develop symptoms. As a test of infectivity, green peach aphids (*Myzus persicae*) were exposed to symptomatic pokeweed plants inoculated with the PkMV-Ag01 clone. Healthy, uninfected plants were then introduced and the aphids permitted to feed freely on the healthy plants. Fifteen days later, mosaic mottling of leaves was evident on the previously healthy plants (PkMV-Ag01 aphid transmitted). To avoid the possibility of contamination among plants of different treatments, aphid transmission experiments were conducted in a greenhouse, separate from plants infected with PkMV-AR or PkMV-Ag01, which were maintained in growth chambers. Transmission of progeny virus by the insect vector shows that the infectious clone of PkMV is fully functional and biologically active.

To confirm that mosaic symptoms were caused by potyvirus infection, viral particles were isolated from non-infiltrated emerging leaves of plants agroinfiltrated with PkMV-Ag01, again by sucrose gradient centrifugation (Berger & Shiel, 1998). Particles were concentrated by centrifugation and applied to copper grids, stained with uranyl acetate, and visualized with a FEI 3D Quanta FEG electron microscope. PkMV has filamentous particles, approximately 30 nm in width by 700 nm in length, bearing the classic morphology of potyviruses (Figure 6B). Viral RNA was extracted from purified virus preparation and separated on a 1% agarose gel. The presence of a band at approximately 9.5 Kb correlates well with the expected size of the PkMV genome at 9516 nucleotides (Figure 6C).

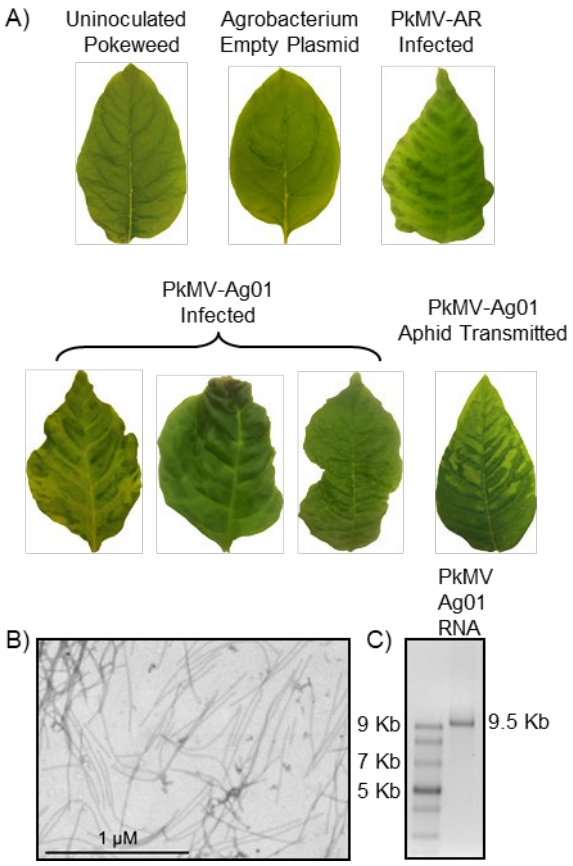

Figure 6 - Symptoms of infection with PkMV-Ag01 and aphid transmission. A) Leaves of intact pokeweed plants at the 3-4 leaf stage were infiltrated with transformed *Agrobacterium* bearing the viral clone. After 14 days, newly emerging leaves showed characteristic mosaic symptoms (PkMV-Ag01 infected). Leaves of pokeweed following aphid transmission of the clone-derived progeny virus also showed chlorotic mottling (PkMV-Ag01 aphid transmitted). Positive control leaves were rub-inoculated with purified virus preparation (PkMV-AR infected). Negative control leaves were either uninoculated or agroinfiltrated with an empty plasmid. **B)** Purified virus preparation isolated from plants agroinfiltrated with PkMV-Ag01. Following purification through a sucrose gradient and concentration by centrifugation, particles were applied to copper grids, stained with uranyl acetate, and viewed at 50,000X with an electron microscope. **C)** Viral RNA was extracted from purified virus preparation isolated from plants agroinfiltrated with PkMV-Ag01, and separated on 1% agarose in denaturing RNA loading dye with ssRNA ladder (NEB) to estimate size of RNA.

To further validate PkMV infection, total RNA was isolated from emerging symptomatic leaves and reverse transcribed with C2 primer, followed by PCR with C1 primer and the same reverse primer, to amplify a PkMV-specific region of RNA spanning nucleotides 4676-5175. Products of the expected size were evident from samples of plants agroinfiltrated with viral clones (Figure 7A; clones Ag01, 02 and 03) and from plants inoculated directly with purified virus preparation of native PkMV-AR. RNA was also isolated from clone-derived purified progeny virus preparation (PkMV-Ag01 particles) and tested for the PkMV-specific region by PCR. Negative control samples were either uninoculated or agroinfiltrated with empty vector. In addition, pokeweed cell lysates were examined by immunoblotting for the presence of PkMV coat protein using a monoclonal antibody directed against a conserved peptide found in coat proteins of many potyviruses (1:2000; Agdia cat# CAB 27200; Figure 7B). Non-infiltrated emerging leaves of plants agroinfiltrated with viral clones also expressed viral coat protein. An additional, higher molecular weight protein was observed in total cell lysates, suggesting that the antibody also detected the PkMV polyprotein intermediate or a host protein strongly associated with viral coat protein. Taken together, these results show that viral clones of PkMV are infectious and that progeny viral particles are synthesized that can move systemically through pokeweed.

To test the possibility of developing the PkMV clone into a viral vector to express foreign genes or siRNA in pokeweed, the open reading frame (ORF) of enhanced green fluorescent protein (eGFP) was inserted into the PkMV infectious clone. Addition of eGFP between P1 and HC-Pro was based on previously published work demonstrating this region as a viable insertion site (Figure 8A, B; Beauchemin et al. 2005).

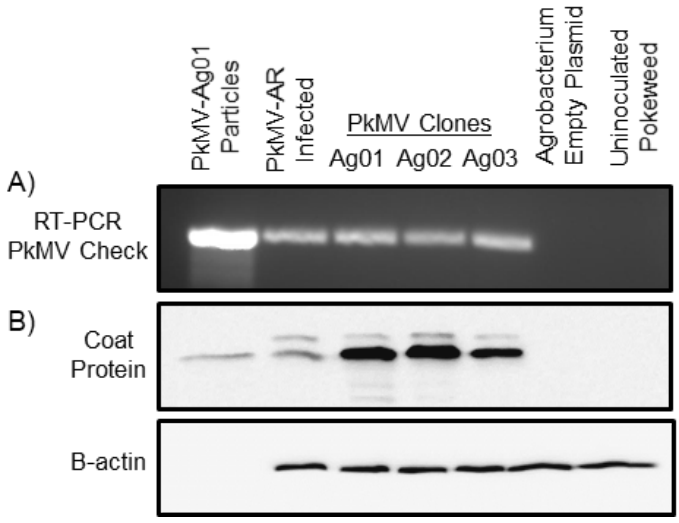

Figure 7 - Validation of PkMV-Ag01 infection. A) Total RNA was isolated from newly emerging leaves of pokeweed plants 14 days after agroinfiltration with PkMV-Ag01 and used as template for RT-PCR with PkMV RNA-specific primers (PkMV clones). As positive controls, RNA was isolated from clone-derived progeny virus preparation (PkMV-Ag01 particles) and plants infected with native particles (PkMV-AR infected). Negative control leaves were either uninoculated or agroinfiltrated with an empty plasmid. PCR products were separated through 1% agarose and visualized with ethidium bromide. **B)** Total cell lysate from pokeweed leaves (60 µg) was separated through 12% SDS-PAGE, transferred to nitrocellulose and probed with monoclonal antibodies specific to potyvirus coat protein (1:2000) and β-actin (1:2000; as loading control). The same experimental and control samples were used as in **A**.

With the PkMV-Ag01 clone as template, two PCR fragments were synthesized spanning the vector backbone and insertion site, using primers for fragment 1 (D1, D2) and fragment 2 (D3, D4). Using pCambia-eGFP as template, a third PCR product was synthesized containing the ORF of eGFP, with PkMV sequence at the 5' and 3' ends introduced with primers D5 and D6 and overlapping the insertion site. PCR products were purified with low melt agarose and combined with HiFi DNA Assembly Master Mix (NEB). As with PkMV-Ag01, Gibson assembly was chosen to clone PkMV-eGFP. The assembled clone was electroporated directly into *Agrobacterium tumefaciens*, strain Agl1 (Tuo et al., 2017). Colonies were selected and insertion of eGFP was confirmed with PCR using the same primers to amplify eGFP (D5, D6). *Agrobacterium* bearing PkMV-eGFP was agroinfiltrated into the third true leaf of intact pokeweed plants, as described for the PkMV-Ag01 clone.

After 14 days, non-infiltrated emerging leaves exposed to 450 nm light and viewed through a GFP bandpass filter (525AF30RED/EM930) showed fluorescence associated with eGFP expression (Figure 8C). This fluorescence pattern was specific to PkMV-eGFP and absent from leaves agroinfiltrated with PkMV-Ag01 or rub-inoculated with native PkMV-AR virus preparation. Leaf cell lysates probed with a polyclonal antibody specific to GFP (1:5000, Cell Signaling, cat# 2555) confirmed its presence in PkMV-eGFP agroinfiltrated plants (Figure 8D). The expression and systemic movement of eGFP supports the potential of PkMV-Ag01 as a viral vector, to transiently over-or under-express specific genes in pokeweed. Therefore, the infectious clone of PkMV will facilitate investigation of function of pokeweed genes that mediate stress response.

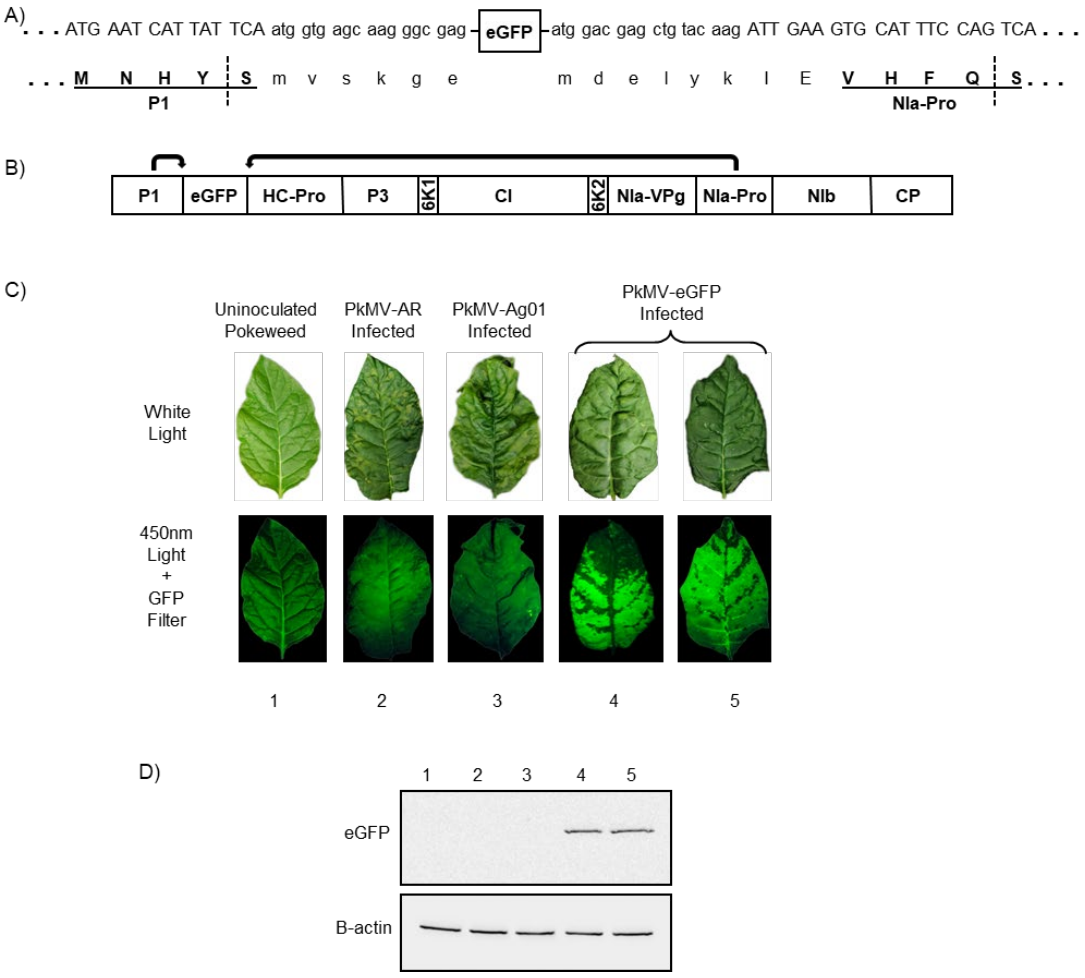

Figure 8 - PkMV-eGFP design and infection. A) Nucleotide and amino acid sequence of eGFP N- and C-terminal boundaries. Uppercase and lowercase letters denote PkMV and eGFP sequences, respectively; bold and underline indicate protease recognition sequences; dashed lines denote the protease cut sites. **B)** eGFP is excised from the PkMV polyprotein by P1 and NIa-Pro proteases. **C)** Leaves of intact pokeweed plants at the 3-4 leaf stage were either agroinfiltrated with PkMV-Ag01 or PkMV-eGFP, rub-inoculated with native particles (PkMV-AR infected), or uninoculated. After 14 days, emerging leaves were observed under white light or blue light (450 nm) with a GFP bandpass filter. **D)** Total cell lysate from pokeweed leaves (60 µg) was separated through 12% SDS-PAGE, transferred to nitrocellulose and probed with polyclonal antibody specific to GFP (1:5000) and monoclonal β-actin antibody (1:5000; as loading control). The same experimental and control samples were used as in **C**, indicated by numbers 1-5.

2.3 – Acknowledgements

This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to KAH and an Ontario Graduate Scholarship to AK.

Chapter 3:

Facile method of curing toxicity in large viral genomes by high-throughput identification and removal of cryptic promoters.

Published manuscript

Klenov, A., & Hudak, K. A. (2021). *Journal of Virological Methods*, 287, 113993. https://doi.org/10.1016/j.jviromet.2020.113993

Contributions: AK conceived the study. AK performed all wet lab work. AK drafted the manuscript. AK and KH edited the manuscript.

All supplementary material referenced in this manuscript are available in the appendix.

3.1 – Abstract

Infectious plant virus clones are challenging to construct and manipulate due to the presence of cryptic promoter sequences that induce toxicity in bacteria. Common methods to overcome toxicity include intron insertion to interrupt toxic open reading frames and the use of *Rhizobium* or yeast species that do not recognize the same cryptic promoters. Unfortunately, intron insertion must be attempted on a trial and error basis within full-length clones and may change the infection characteristics of the virus. We have developed a facile method that can detect multiple cryptic bacterial promoters within large virus genomes. These promoters can then be silenced to obtain infectious clones that can be manipulated in *E. coli*. Our strategy relies on the generation of a viral library which is cloned upstream of either an eGFP open reading frame for low-throughput analysis or chloramphenicol for next generation sequencing. Pokeweed mosaic virus (PkMV), a 9.5 Kb ssRNA potyvirus, was used as a proof of concept. We found 16 putative promoter regions within 150-250 bp library fragments throughout the PkMV genome. 5'RACE allowed identification of the promoter sequence within each fragment, and subsequent silencing produced infectious clones. Our results indicate that cryptic promoters are ubiquitous within large viral genomes and that promoter screening is a desirable first step when constructing a viral clone. Our method can be applied to large plant and animal viruses as well as any DNA sequence for which low level of background transcriptional activity is required.

3.2 - Introduction

The development of infectious plant virus clones has been pivotal in understanding plant-pathogen interactions as well as harnessing the biotechnological potential of plants. Infectious viral clones provide a consistent way of infecting plants with genetically altered forms of a virus to study aspects of virus biology such as viral movement (Cotton et al., 2009) and host response (Zilian & Maiss, 2011), in addition to applications such as protein overexpression (Touriño et al., 2008) and gene silencing (Gammelgard et al., 2007). Infectious clones of plant RNA viruses are primarily cDNA copies of viral genomes (Pasin et al., 2019) inserted into a plasmid with requisite promoters and terminators that enable expression and infection within the plant. The viral clone can be delivered by rub inoculation (Scholthof, 1999), particle bombardment (Tuttle et al., 2012) or agroinfiltration (Bordat et al., 2015). Comprehensive methods for infectious clone and viral vector delivery have been reviewed recently (Abrahamian et al., 2020).

The first plant virus clones were constructed with traditional restriction enzyme cloning, requiring sequential cloning of viral fragments. A significant advancement in generation of plant virus plasmids was the use of sequencing independent cloning techniques such as Gibson assembly (Blawid & Nagata, 2015). This method allows the assembly of long viral genomes in a single cloning reaction. Unfortunately, one significant bottleneck remains in the generation and manipulation of viral clones, specifically the toxicity often encountered with plasmid generation in *E. coli*. Toxicity in *E. coli* due to the presence of viral sequences has been observed early in infectious clone development for both plant (Ouillet et al., 1989) and animal (Chambers et al., 1990) viruses. It is hypothesized that this toxicity is due to the presence of cryptic bacterial promoters (Fakhfakh et al., 1996), driving transcription and subsequent translation of toxic

products. The resulting viral clones are recovered with mutations or deletions and are often non-infectious. However, the exact mechanism of toxicity in *E. coli* is not well understood.

Several strategies have been used to counter this toxicity, including the use of alternate *E. coli* strains (Aubry et al., 2015), low copy number plasmids (Gritsun & Gould, 1998), assembly in yeast (Tuo et al., 2015) or *Agrobacterium* (*Rhizobium radiobacter*) (Tuo et al., 2017), intron insertion (Johansen, 1996), and silent mutagenesis (Pu et al., 2011). Varying *E. coli* strains and plasmid backbones aim to reduce the level of toxicity to permit manipulation of viral clones. However, the choice of backbone and strain must be determined empirically for new viral clones. Assembly of clones in *Agrobacterium* or yeast is useful for the initial generation of an infectious clone, but is hampered by poor plasmid yield, which slows further manipulation and targeted mutagenesis of the clone (Sripriya et al., 2011). Intron insertion is the most widely adopted strategy, however the location of insertion is often determined on a trail and error basis (Johansen, 1996). Silent mutagenesis is the most direct approach of curing toxic cryptic promoters and is generally accomplished by subdividing the viral genome into smaller parts, detecting the presence of promoters with reporter genes and finally silently mutating them (Pu et al., 2011). This approach, while ultimately the most effective, is extremely labor intensive when applied to large RNA viruses such as potyviruses. Potyviruses are a large group of plant viruses whose members cause significant damage to agricultural crops (Nicaise, 2014).

We therefore developed a method that allows for the rapid detection and silencing of cryptic bacterial promoters within large plant virus genomes. This method may be performed in a low-throughput manner, by cloning a viral cDNA library upstream of a eGFP open reading frame, which allows screening of transformed *E. coli* by eye. A complementary, high-throughput approach utilizes the same cDNA library cloned upstream of a chloramphenicol resistance open

reading frame, allowing surviving colonies to be collected *en masse* for next generation sequencing. Both approaches effectively construct putative promoter region (PPR) maps of the viral genome. The location of promoters within PPRs is then determined by identification of transcription start sites immediately downstream of cryptic promoter sequences. Identified promoters are then silenced during the generation of the viral clone. This method allows the construction of infectious viral vectors that can be manipulated without rearrangement in *E. coli*, and is applicable to plant and animal viruses as well as any DNA sequence shown to be toxic during cloning.

3.3 - Material and Methods

Viral particle isolation

Viral particles were isolated from leaves of 12-14 leaf pokeweed plants infected with the pokeweed mosaic virus (PkMV) Arkansas (AR) isolate of the family *Potyviridae*, genus *Potyvirus* 30 dpi. Viral particles were collected by sucrose gradient centrifugation, as described in the general potyvirus purification protocol (Berger & Shiel, 1998). Following isolation, purified particles were either processed immediately for RNA isolation or stored in 50% glycerol at -20ºC.

RNA isolation

Viral RNA was extracted with the hot phenol method detailed in the general potyvirus purification protocol (Berger & Shiel 1998). Viral RNA was resuspended in RNA storage buffer (22.5 mM DTT, 1 mM Citrate pH 6.3), quantified with a Nanodrop spectrophotometer and stored at -40ºC. Integrity of RNA was checked by visual observation following separation of 0.5 µg of

RNA in 1X RNA loading dye (47.5% formamide, 0.5 mM EDTA, 0.01% SDS, 0.01% bromophenol blue. 0.005% xylene cyanol) on a 1.5% agarose gel.

Reverse transcription

Viral RNA (100 ng) was combined with 1 µL of 50 µM random hexamers (NEB Cat#: S1230S) and 1 µL of 10 mM dNTPs in a total volume of 13 µL. RNA was denatured at 65ºC for 10 minutes and placed immediately on ice. To the sample, 4 µL of 5X Superscript IV buffer, 1 µL 0.1 M DTT, 200 U of Superscript IV reverse transcriptase (Thermo Cat #: 18090010) and 40 U of murine RNase inhibitor (NEB Cat #: M0314S) was added. The sample was incubated at room temperature for 10 minutes and then 50ºC for 1 hour. After cDNA synthesis, the reaction was stopped by heating to 80ºC for 10 minutes. Afterwards, 5 U of RNase H (NEB Cat #: M0297S) was added and the sample was incubated for an additional 30 minutes at 37ºC. cDNA was stored at -20ºC until PCR amplification.

Viral library preparation

To prepare the viral library for promoter screening, the cDNA of the viral RNA was diluted 1:10 with dH_2O (v:v) and used as template for PCR amplification and addition of Golden Gate adapter sequences. The PCR reaction was performed on 2 µL of diluted cDNA template with 2 U of Q5 DNA polymerase (NEB Cat#: M0491S) according to manufacturer's protocols with 0.5 µM forward adapter primer (Lib-For; Table S1) and reverse adapter primer (Lib-Rev; Table S1). All primers used in this study are listed in Supplementary Tables 2-5. The following cycling parameters were used: first 5 cycles of an initial denaturation at 94ºC for 90 seconds followed by a ramp from 20ºC to 70ºC in 1ºC increments with an incubation of 3 seconds at each temperature. Afterwards, PCR proceeded with 40 cycles of denaturation at 94ºC for 20 seconds,

annealing at 58ºC for 20 seconds, and extension at 72ºC for 20 seconds. A final extension of 40 seconds at 72ºC finished the reaction. Multiple reactions were pooled and PCR products purified with the Biobasic PCR Products Purification Kit (Cat #: BS363), concentrated with vacuum centrifugation and separated on a 1% low-melt agarose gel. PCR products between 200-500 bp were excised and purified with the Biobasic Gel Extraction Kit (Cat#: BS353). Following quantification with a Nanodrop spectrophotometer, libraries were stored at -20ºC.

Positive and scrambled control DNA fragments were constructed by annealing top (sense) strand primers (Pos-Top, Scr-Top; Table S1) and bottom (antisense) strand primers (Pos-Bott, Scr-Bott; Table S1) at a final concentration of 10 µM of each primer in 1X annealing buffer (25 mM MOPS pH 7.9, 5 mM EDTA) to a volume of 20 µL. Annealing was performed in a thermocycler with the following protocol: 94ºC for 90 seconds, ramp from 94ºC to 74ºC in 1ºC increments and 4 seconds per degree, followed by ramping from 74ºC to 54ºC at 1ºC and 15 seconds per step and finishing with a ramp from 54ºC to 34ºC in 1ºC increments and 30 seconds per degree. Positive and scrambled control fragments were diluted 1:200 to a working concentration of 50 fmol/µL.

Two reporter vectors were constructed for insertion of library and control fragments, one encoding eGFP and the other encoding a chloramphenicol resistance gene. pGG-eGFP was constructed by inserting a Golden Gate cloning site upstream of the Shine-Delgarno sequence and eGFP reporter of the template plasmid pJ02B2Gm_AE (Iverson et al., 2015). pGG-Chl was generated using pGG-eGFP as template by replacing the open reading frame of eGFP for the chloramphenicol resistance open reading frame using Gibson assembly. Golden Gate cloning was used to ligate the library and control fragments into the reporter vectors. In 20 µL, 1X T4 DNA Ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP) was

combined with 200 U of T4 DNA Ligase (NEB Cat#: M0202S), 10 U of BsaI-HFv2 (NEB Cat #: R3733S), 50 fmol of vector (either eGFP or antibiotic reporter) and 150 fmol of insert library or control DNA fragment. The vector contained a Golden Gate site inserted upstream of the eGFP or chloramphenicol resistance ORF. The ligation was performed in a thermocycler with the following protocol: 37ºC for 3 minutes, 16ºC for 6 minutes, 25 cycles of 37ºC for 1.5 minutes and 16ºC for 3 minutes, 37ºC for 6 minutes followed by heat denaturation at 80ºC for 10 minutes. Each ligation (2 µL) was transformed into DH5α cells and plated on either LB agar with 50 µg/mL kanamycin for eGFP screening, or LB agar with 50 µg/mL kanamycin and 210 µg/mL chloramphenicol for antibiotic screening. eGFP library plates were incubated at 37ºC for 24 hours and left to mature at room temperature for 48 hours. Chloramphenicol library plates were incubated at 37ºC for 24 hours.

Low-throughput eGFP screening and sequenced reads alignment

Bacterial colonies transformed with the viral library or controls on LB kanamycin plates were screened visually with a 470 nm LED light observed through a 510BP20 filter or through orange-lensed glasses. Fluorescent colonies were picked and grown in small cultures of LB, 50 µg/mL kanamycin and plasmids isolated with the Biobasic Plasmid DNA Miniprep Kit (Cat #: BS413). Resulting plasmids were Sanger sequenced with primers Lib-Seq-For and Lib-Seq-Rev (Table S1) flanking the Golden Gate insertion site. Sanger sequencing results were aligned to the PkMV genome using the ClustalO algorithm (Sievers et al., 2011). Regions with an overrepresentation of aligned reads were chosen as putative promoter regions (PPRs) and used as input for promoter identification.

High-throughput antibiotic screening and sequenced reads alignment

Bacterial colonies transformed with the viral library or controls on LB kanamycin/chloramphenicol agar plates were scraped by resuspending all colonies on a plate in 4 mL LB medium. Plasmid DNA was isolated using the Biobasic Plasmid DNA Kit and 2 ng of isolated plasmid DNA was amplified with forward and reverse primers (NGS-For, NGS-Rev; Table S1) using the following parameters: 94ºC for 90 seconds, 12 cycles of 94ºC for 30 seconds, 58ºC for 30 seconds, 72ºC for 30 seconds, with a final extension of 72ºC for 1 minute. Several PCR reactions were pooled and Cutsmart buffer (NEB Cat#: B7204S) was added to a 1X concentration along with 20 U of DpnI (NEB Cat #: R0176S) and incubated overnight at 37ºC. The PCR products were isolated with the Biobasic PCR Cleanup Kit, eluted and quantified with the Nanodrop spectrophotometer. The samples were then sent for Illumina next-generation sequencing. Illumina paired-end sequencing was performed on the isolated PCR products (Bio Basic Inc) and bioinformatics analysis was performed on a Galaxy server (Giardine et al., 2005). Adapters were trimmed and reads with a quality score lower than 30 were discarded. Alignment to the PkMV genome was performed with Bowtie2 (Langmead & Salzberg, 2012) and the output was visualized with the Integrative Genomics Viewer (IGV, Thorvaldsdóttir et al., 2013). Overrepresented stacks of reads in the alignment were defined as PPRs.

Putative promoter identification and testing

TSO-5'RACE (template switching oligo 5' rapid amplification of cDNA ends) was used to experimentally identify the location of promoters within PPRs by obtaining +1 transcription start sites. PPRs were PCR amplified from the PkMV genome with PkMV specific primers (Table S2) using the following conditions: initial denaturation at 94ºC for 90 seconds, cycling 35 times at 94ºC for 45 seconds, 63ºC for 30 seconds, 72ºC for 60 seconds with a final extension of

72ºC for 120 seconds. PCR fragments were cloned into the eGFP reporter construct using Golden Gate assembly and plated on LB agar, 50 μg/mL kanamycin. Total RNA extraction was performed on high eGFP expressing colonies using the RNAsnap protocol (Stead et al., 2012). Reverse transcription was performed on 100 ng of bacterial total RNA using the Template Switching RT Enzyme Mix (NEB Cat#: M0466S), a TSO oligo (TSO; Table S2) and an eGFP specific reverse primer (Lib-5R-RT; Table S2). The cDNA was used as template for PCR using the manufacturers recommended protocol with adapter specific (TSO-For; Table S2) and eGFP specific primers (RevA, RevA1; Table S2). Resulting PCR products were cloned into a pUC-19 vector, sequenced and aligned to the PkMV genome. At least 3 colonies were sequenced per PPR to develop a consensus sequence. The first 50 bp upstream of the consensus +1 transcription start site was designated as the promoter, synthesized as top (sense) and bottom (antisense) strand oligos (Table S3) with Golden Gate adapter sites and ligated into the eGFP reporter construct to test for promoter activity. As positive control, a strong Sigma 70 promoter (TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGC) was introduced 5' of the eGFP open reading frame and a corresponding randomly scrambled promoter sequence (GATCGGATACGTACCTCGAGTTAATGCTACGCTAT) served as a negative control.

To quantify eGFP expression, promoters cloned into the eGFP construct were transformed into DH5α and plated on LB agar, 50 μg/mL kanamycin. Colonies were resuspended in 0.8% NaCl, diluted to an OD_{600} of 0.2 and 200 μL of cells were placed into black sided 96 well plates. The level of eGFP fluorescence was quantified using a Synergy H4 Hybrid Reader with 480 nm excitation and 520 nm emission filters. eGFP fluorescence was normalized to the OD_{600} of each sample and means ± S.E. were calculated for triplicate samples with four technical replicates each. Promoters with high levels of eGFP fluorescence were chosen as

targets for mutagenesis within the viral genome. Promoters with silent mutations were generated as described above and the location of mutations were based on the approximate location of the -10 and -35 box.

Generating cured PkMV clones

Cured PkMV constructs were generated by PCR using PkMV-eGFP as plasmid template. The first PCR section spanned from the most 5' end of the virus to the identified cryptic promoter (Primers: pLX-PkMV-F1-For and pLX-PkMV-F1-10F/11R-SM-Rev, Table S4) and the second PCR section spanned from the cryptic promoter to the most 3' end of the viral genome (Primers: pLX-PkMV-F2-10F/11R-SM-For and pLX-PkMV-F2-Rev, Table S4). The PCR program used was as follows: initial denaturation at 94ºC for 90 seconds, cycling 35 times at 94ºC for 45 seconds, 63ºC for 30 seconds, 72ºC for 3 minutes with a final extension of 72ºC for 6 minutes. The resulting PCR products were gel purified as described above and 300 fmol DNA was cloned into 50 fmol of PCR linearized pLX-B2-RFP vector (Primers: pLX-PkMV-Vec-For and pLX-PkMV-Vec-Rev, Table S4) (Pasin et al., 2017) with TEDA cloning (Xia et al., 2019). The assembled products were transformed into DH10β competent cells, plated on LB agar, 50 µg/mL kanamycin and incubated at 30ºC for 2 days. Colonies were selected and screened by restriction digestion with PvuI-HF. Plasmids of the correct size, approximately 13.5 Kb, were transformed into *Agrobacterium* (*Rhizobium radiobacter*) for subsequent agroinfection.

Testing cured PkMV clones for infectivity

Agroinfection

Agroinfection was performed as described (Tuo et al., 2017). Transformed *Agrobacterium* colonies were selected and grown to an OD_{600} of 1.0 in YEP medium, washed

once with agroinfiltration buffer (10 mM MES pH 5.6, 10 mM $MgCl_2$, and 100 μM acetosyringone), resuspended to an OD_{600} of 0.5 with the same buffer and incubated in the dark for 3 hours at room temperature. *Agrobacterium* cultures were syringe infiltrated into 4-leaf stage pokeweed plants; symptoms appeared within 4-7 days post infection.

eGFP fluorescence viewing and immunoblotting

eGFP expression was visualized in infected plants by exciting the fluorescent protein with a 470 nm LED and viewing through a 510/20 bandpass filter attached to a Canon Rebel SL2 camera. Expression of viral coat protein from the PkMV clones was detected by immunoblotting. Approximately 20 mg of leaf tissue from infected pokeweed was macerated with 200 μL Extraction buffer (25 mM Tris-HCl, pH 7.5, 1 mM EGTA, 1 mM DTT, 1 mM PMSF, 5% glycerol). Crude lysate was centrifuged at 16,000 x g for 5 minutes to pellet insoluble matter. Supernatant protein was quantified by Bradford Assay and 60 μg was separated through 12% SDS-PAGE and transferred to nitrocellulose. Blots were probed for PkMV coat protein with a monoclonal antibody directed against a conserved peptide of potyviral coat protein (1:2000; Agdia cat# CAB 27200). Blots were stripped using 8 M guanidine hydrochloride and re-probed with β-actin-specific monoclonal antibody (1:2000) as indicator of loading amount. Proteins were visualized by chemiluminescence using a MicroChemi imager (DNR Micro Imaging).

3.4 - Results

We reported previously the construction of an infectious clone of PkMV in *Agrobacterium* instead of *E. coli* because of toxicity we observed during cloning attempts in *E. coli* (Klenov & Hudak, 2018). Indication that the PkMV genome (Figure 9A) causes toxicity in

E. coli was visualized by plasmid DNA extraction from cells transformed with the wild-type PkMV vector and PCR amplification of three sections of the viral genome (1, 2 and 3). Comparison of PCR fragment separation on agarose gel with fragments resulting from RT-PCR performed with the same primers on viral RNA extracted from particles indicated that PCR fragment derived from part 2 of the plasmid was reduced in length by approximately 3.5 Kb (Figure 9B). This difference in size suggests that during cloning, a portion of plasmid DNA within part 2 of the viral genome was deleted in *E. coli.* Instability of the clone suggested the presence of cryptic promoters.

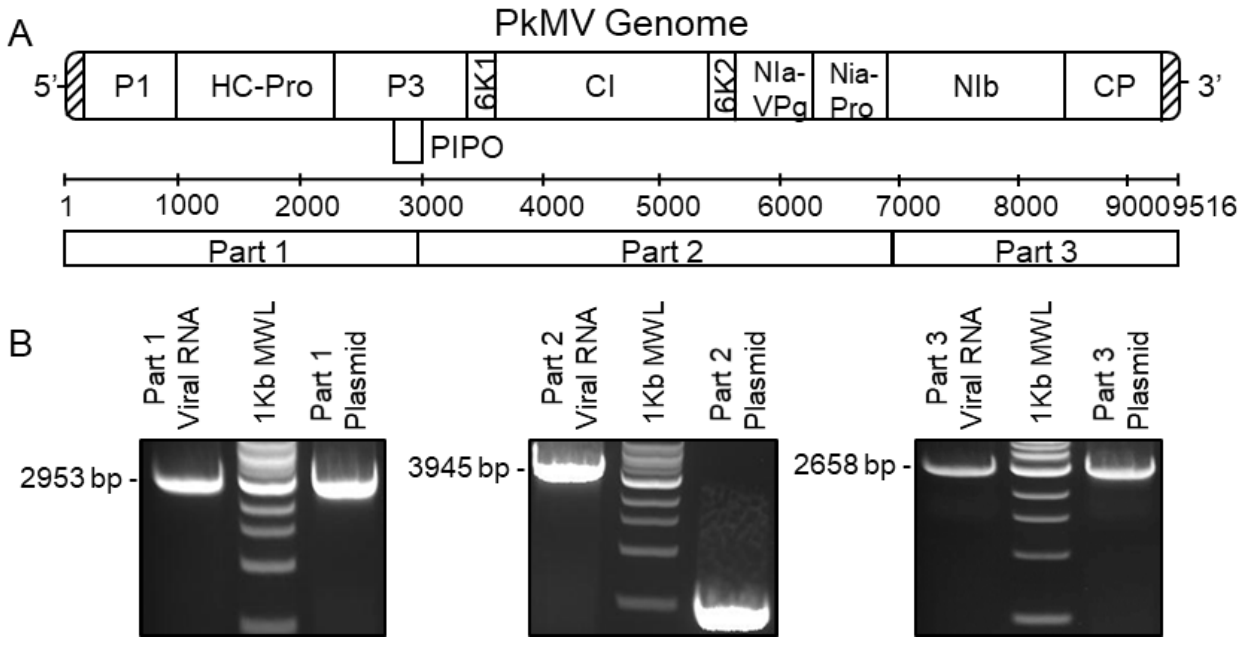

Figure 9 – Region of toxicity in PkMV. A) The PkMV genome is 9516 bp long and shares the genome organization typical of potyviruses. The ssRNA is translated as a single polyprotein that is cleaved into 10 viral proteins, with an 11th protein, PIPO, derived from a +2 frameshift. B) Separation of DNA fragments generated by PCR of PkMV plasmid isolated from *E. coli* or RT-PCR of viral RNA isolated from PkMV particles. PCR/RT-PCR was performed in three sections (1, 2 and 3) corresponding to the indicated regions of the PkMV genome. The brightest band on the 1 Kb molecular weight ladder (MWL) indicates 3 Kb.

Low-throughput promoter screening was performed as an initial step to identify these cryptic promoters within the PkMV genome. Viral RNA was reverse transcribed with random hexamers and the cDNA was amplified with PCR. PCR products between 200-500 bp were cloned into an eGFP reporter vector (pGG-eGFP). *E. coli* transformed with the assembled plasmids were viewed under blue light (Figure 10A, B). Compared to the negative control, a sequence-scrambled promoter, many of the PkMV transformants produced fluorescent colonies indicating promoter activity (Figure 11). The brightest colonies, as judged by eye, were selected for plasmid isolation and Sanger sequencing. Approximately 40 individual colonies were sequenced and the reads aligned to the PkMV genome (Figure 12A). Reads segregated into stacks, defining areas of promoter activity within 200-500 bp regions of the PkMV genome.

As an alternative to the low-throughput screening, the high-throughput method was performed in a similar manner, except that the viral library was size selected for 150-400 bp fragments and cloned upstream of an antibiotic resistance marker in vector pGG-Chl. Surviving colonies were collected *en masse* for plasmid isolation and the library was amplified for paired-end Illumina sequencing (Figure 10A, C). Approximately 8 million sequence reads were aligned to the PkMV genome and output files were visualized with the Integrative Genomics Viewer (Figure 12B). The high-throughput screening method provided a higher resolution promoter activity map of the PkMV genome, allowing the definition of 16 stacks of reads, corresponding to 16 putative promoter regions (PPRs). The high-throughput PPRs were comparable with the reads obtained from the low-throughput method particularly in regions encoding CI and CP of the viral genome, with differences observed in the P3 and NIb encoding regions. PPRs 6-12 were found to be within part 2 of the PkMV genome suspected of toxicity and were selected for further characterization.

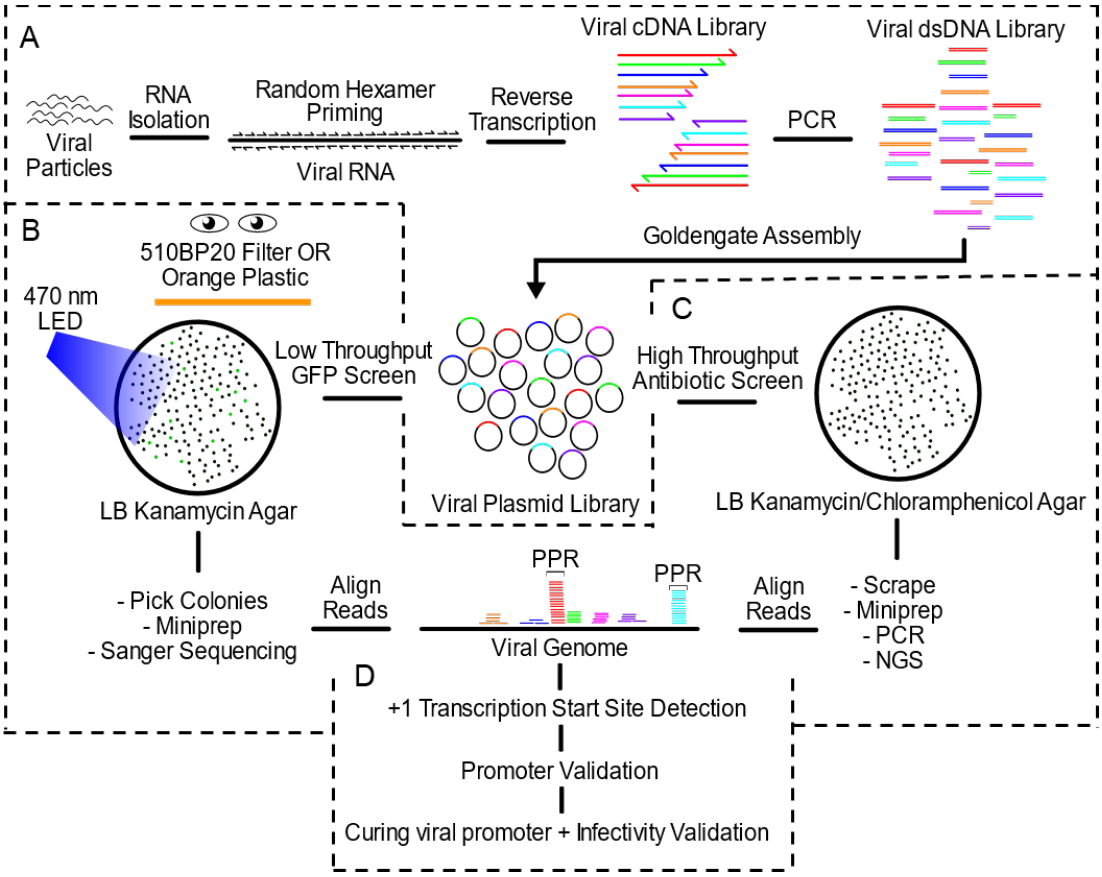

Figure 10 - Promoter screen of viral genomes. A) RNA was isolated from viral particles and reverse transcribed with random hexamers, amplified with adapter primers and gel extracted. The dsDNA viral library was then cloned using Golden Gate assembly into an eGFP or antibiotic reporter plasmid. B) For low-throughput promoter screening, the eGFP library was transformed into *E. coli* and colonies were observed under 470 nm light through a 510BP20 filter or orange plastic filter. Colonies with high level of eGFP fluorescence were chosen for plasmid DNA isolation and Sanger sequencing. C) For high-throughput promoter screening, antibiotic libraries were plated on LB agar with chloramphenicol. Surviving colonies were scraped and their plasmid DNA isolated. The plasmid pool was amplified at the Golden Gate insertion site and Illumina sequenced (NGS). D) Sequencing reads from high- or low-throughput methods were aligned to the viral genome and regions with an over-representation of reads were designated putative promoter regions (PPRs). PPRs were selected and silently mutated to stabilize subsequent viral clones.

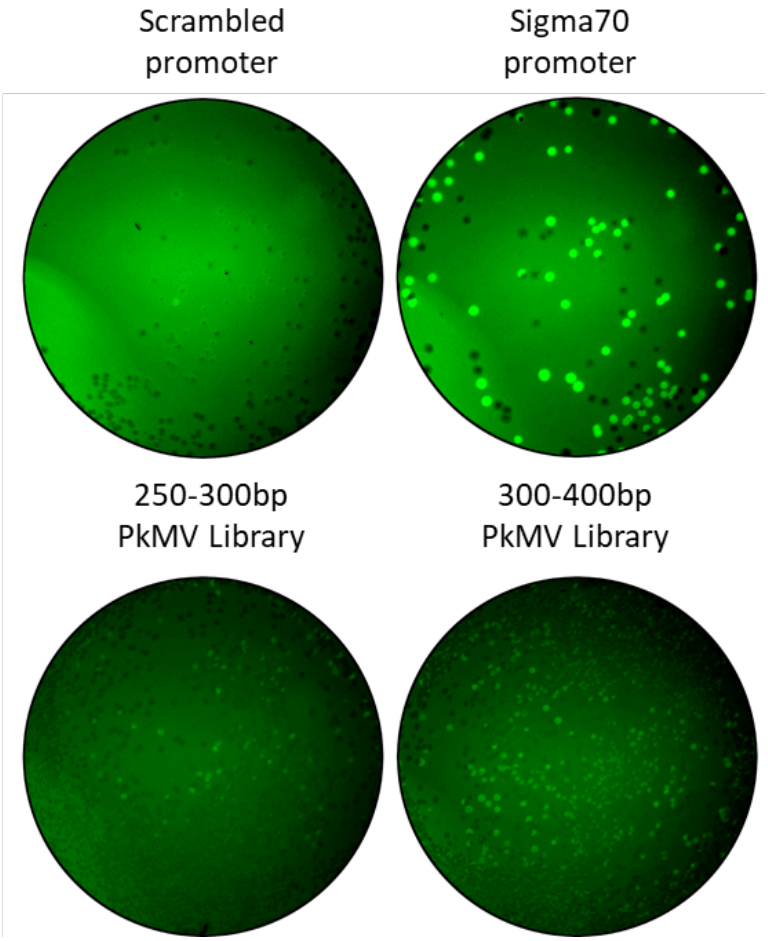

Figure 11 – Low throughput promoter screen of PkMV. The pool of amplified viral cDNAs was cloned into the eGFP reporter vector and transformed into *E. coli*. Resulting colonies were visualized under 470 nm light and colonies with high level eGFP fluorescence were chosen for insert sequencing and alignment to the PkMV genome. Two populations of viral insert sizes are shown (250-300 and 300-400 bp). Positive control was a strong Sigma 70 bacterial promotor and the negative control was the randomly scrambled sequence of the positive control.

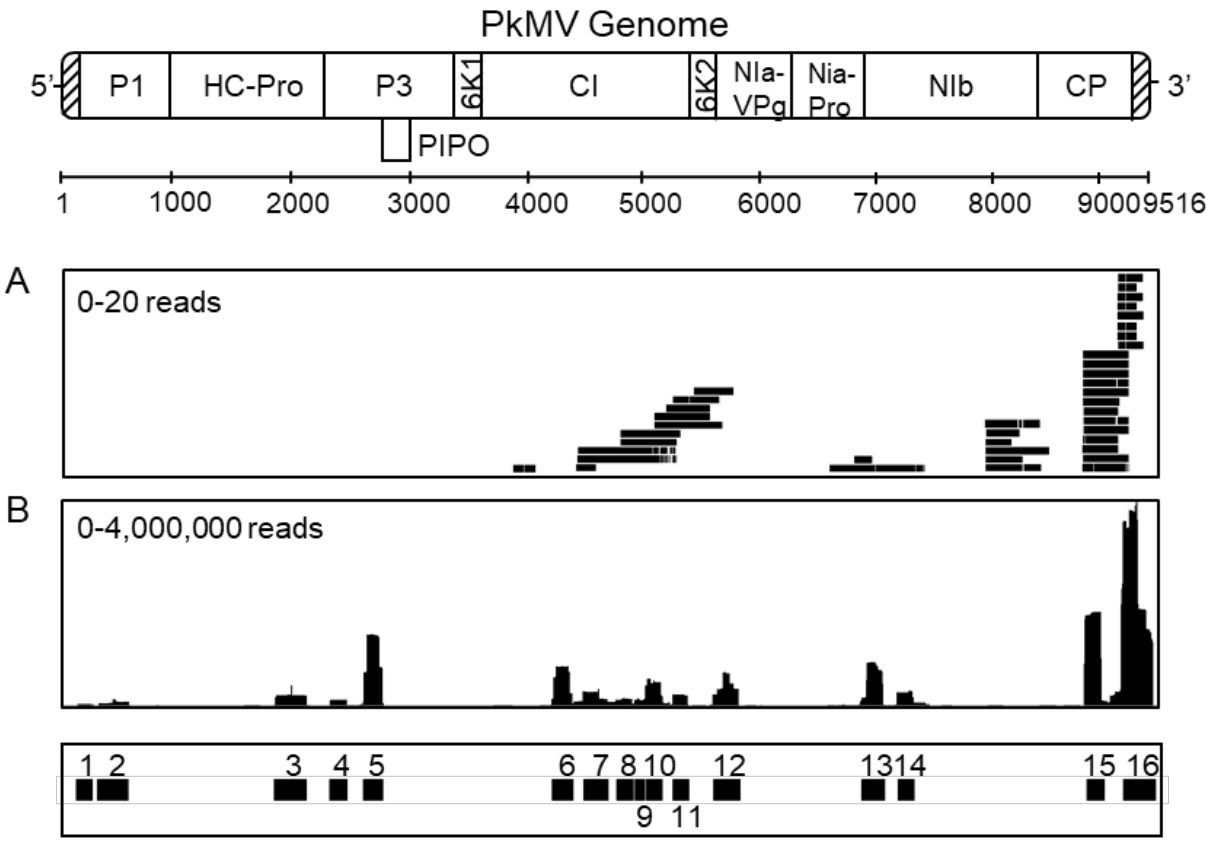

Figure 12 – Alignment of high- and low-throughput sequencing reads to the PkMV genome. A) Sanger sequencing reads from colonies expressing a high degree of eGFP fluorescence were aligned to the PkMV genome using the ClustalO algorithm. B) NGS reads were aligned to the PkMV genome with Bowtie2 and visualized with the Integrative Genomics Viewer. Sequences containing PPRs, numbered 1 to 16, are indicated based on peaks identified from high-throughput sequencing alignment.

To determine the transcriptional strength of the putative promoters and the DNA strand on which they occurred, PPRs 6-12 were amplified in both the forward and reverse orientation and cloned into the eGFP reporter vector. Resulting fluorescence measurements indicated that some PPRs showed minimal promoter activity, whereas levels of fluorescence of PPRs 10F and 11R were comparable to the positive control, a strong constitutive *E. coli* promoter (Figure 13). There was no sequence similarity among these PPRs apart from the presence of -10 and -35 boxes. As well, promoter activity occurred on either the sense or antisense DNA strand of the infectious viral clone, a factor not often considered when searching for cryptic promoters.

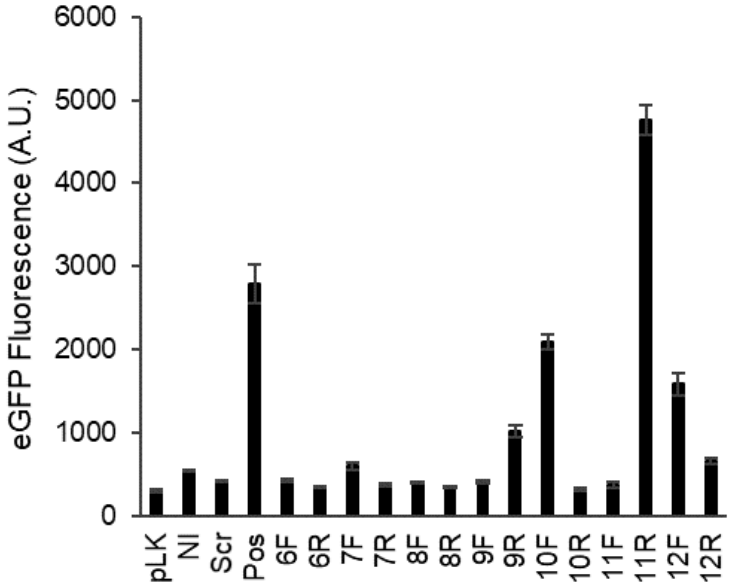

Figure 13 – Promoter activity of identified PPRs within broad toxic region in Part 2 of the PkMV genome. PPRs 6-12 were selected and cloned in forward (F) and reverse (R) orientation into the eGFP reporter vector and transformed into *E. coli*. Resulting colonies were resuspended in saline solution and fluorescence quantified with a 96 well plate reader. Fluorescence values, normalized to OD_{600} for each culture, are means +/- S.E. for three replicates per sample. Positive control (Pos) was a strong Sigma 70 promoter, scrambled control (Scr) was the positive control sequence randomly scrambled, no insert control (NI) was the eGFP reporter vector with no insert, and pLK-1 is a small cloning vector with no eGFP ORF.

The exact location of the cryptic promoters within the PPRs showing a high level of eGFP fluorescence was determined by identifying the transcription start site downstream of these promoters. TSO-5'RACE was performed on total RNA isolated from *E. coli* carrying PPR eGFP constructs and the first 50 bp upstream of the identified +1 transcription start site was defined as the promoter (Table 1). These promoter sequences were silently mutated within their -10 and -35 boxes and cloned into the eGFP reporter construct alongside their wild-type counterparts to measure promoter strength and determine if silencing was effective. In all cases, the levels of eGFP fluorescence were high for wild-type promoters and substantially reduced following the introduction of silent mutations.

Figure 14A illustrates the strategy to introduce mutations within the genome to silence cryptic promoters. A viral clone or reverse-transcribed viral RNA is amplified by PCR using primers designed to bear silent mutations in the cryptic promoter. Resulting DNA fragments are assembled with Gibson assembly, transformed into *E. coli* and screened for infectivity. The introduction of silent mutations into cryptic promoters found within the PkMV genome enabled cloning and manipulation of the genome in *E. coli* without toxicity. However, it was imperative that the viral clone remained infectious in plants. To determine whether silencing the identified promoters produced infectious clones, PkMV-eGFP, a wild-type clone with eGFP open-reading frame inserted between P1 and HC-Pro genes (Klenov & Hudak, 2018) was mutated to silence promoters 10F and 11R, as an example of the method. Both mutant clones bearing silent mutations within promoters 10F or 11R were generated entirely in *E. coli*, producing colonies with consistent growth and no detectable deletions of the viral genome. When agroinfiltrated into pokeweed leaves, these clones produced eGFP fluorescence indicative of successful infection in plants (Figure 14A). As well, cell lysates of leaves agroinfiltrated with mutant PkMV-eGFP

clones and probed with potyvirus-specific antibody indicated the presence of viral coat protein (Figure 14B). Taken together, these results illustrate that introduction of silent mutations into identified cryptic promoters allows for manipulation of the viral clone in *E. coli* while maintaining the infectious nature of the clone in plants.

Table 1 – Identified cryptic promoters and their silenced sequences. TSO-5'RACE was performed on PPRs to identify location of -10 and -35 promoter elements (underlined). Silent mutations (SM) are indicated in grey highlight. Promoter fragments were cloned into an eGFP reporter plasmid, transformed into *E. coli*, and fluorescence was quantified. Values indicate means ± S.E. for three biological replicates and four technical replicates each. Arrows indicate decrease in fluorescence relative to the wild-type (WT) cryptic promoter. Positive control (Pos) was a strong Sigma 70 promoter, scrambled control (Scr) was the positive control sequence randomly scrambled, no insert control (NI) was the eGFP reporter vector with no insert, pLK-1 is a small cloning vector with no eGFP ORF.

PPR	-35 Box -10 Box	eGFP Fluorescence (A.U.)
9R-WT	GGGATCTTGACATTTTCTGG**TATTGT**TGTACCAATCAAATTCC**TGTAAT**T	1065 ± 16
9R-SM	GGGATCTTGACATTTTCTGG**GATCGT**TGTACCAATCAAATTCC**GGTAGT**T	↓ 405 ± 15
10F-WT	TAGGCAGA**TTGACT**ATGGCTCAGGCCACCAAAG**TAGCAT**ACACTTTGCAA	3708 ± 373
10F-SM	TAGGCAGA**CTAACG**ATGGCTCAGGCCACCAAAG**TCGCTT**ACACTTTGCAA	↓ 945 ± 143
11R-WT	CTGAAGAATTTCAATA**TTTTCT**GCCGTGTGATTCCGTG**TATAAT**GACTTC	3448 ± 35
11R-SM1	CTGAAGAATTTCAATA**TTCTCC**GCCGTGTGATTCCGTG**TGTAAT**GACTTC	↓ 334 ± 5
12F-WT	AGTTAG**GCTATG**AGGTGCACGCTGATGATG**ATACAA**TAGAACATTTCTTT	3416 ± 225
12F-SM	AGTTAG**GCTACG**AGGTGCACGCTGATGATG**ACACCA**TAGAACATTTCTTT	↓ 540 ± 16
Pos	**TTGACA**GCTAGCTCAGTCCTAGG**TATAAT**GCTAGC	2792 ± 226
Scr	GATCGGATACGTACCTCGAGTTAATGCTACGCTAT	425 ± 9
NI	N/A; pGG-eGFP without insert	543 ± 10
pLK-1	N/A; cloning vector without eGFP ORF	300 ± 16

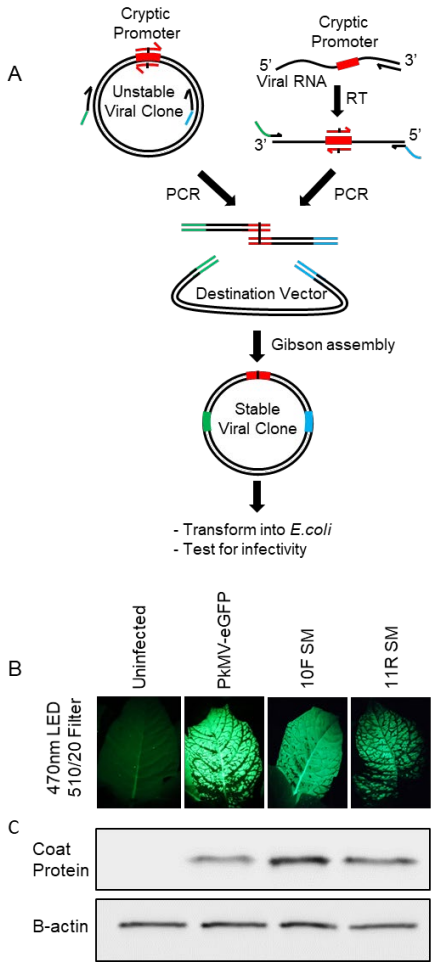

Figure 14 – Silencing of cryptic promoters stabilizes plant virus clones. Following detection of a cryptic promoter in a viral genome, silent mutagenesis using Gibson assembly can stabilize the clone in *E. coli*. A) A previously isolated viral clone or reverse-transcribed viral RNA can be used as template for PCR with primers bearing the silent mutations in the cryptic promoter (red with black vertical line). Fragments are assembled with Gibson assembly, transformed into *E. coli* and screened for infectivity. B) PkMV-eGFP clones with mutations silencing promoters 10F (10F-SM) or 11R (11R-SM) were agroinfiltrated into 4-leaf stage pokeweed plants. Negative control was an uninfected pokeweed, the positive control was unmodified PkMV-eGFP originally constructed in *Agrobacterium*. C) After 14 dpi, fluorescence was observed with a 470 nm LED and 510/20 nm filter attached to a Canon Rebel SL2 camera. B) Total cell lysate from pokeweed leaves (60 µg) was separated through 12% SDS-PAGE, transferred to nitrocellulose and probed with monoclonal antibodies specific to potyvirus coat protein (1:2000) and β-actin (1:2000; as loading control).

3.5 - Discussion

The process of cloning a large viral genome is challenging due to the presence of sequences that induce toxicity and instability in *E. coli*. During our initial cloning of PkMV (Klenov & Hudak, 2018), we observed typical signs of toxicity when assembling fragments of wild-type PkMV in *E. coli*, such as slow-growing colonies and deletion of a portion of the viral genome. The strategy that produced a complete clone relied on Gibson assembly in *Agrobacterium* rather than *E. coli*, and while effective, the process was hampered by longer generation time of *Agrobacterium* compared to *E. coli* and poor quality plasmid DNA.

We have developed a superior method for detecting and curing cryptic bacterial promoters within large viral genomes. Both the low-throughput and high-throughput options reduce the window for promoter search from 9.5 Kb to regions as small as 150 bp. By identifying transcription start sites within PPRs, locations of putative promoters are further narrowed to within 50 bp, from which promoter elements can be identified and silenced effectively. Moreover, when the silencing mutations are integrated into a viral genome, the resulting clone can be manipulated in *E. coli* and retain infectivity in plants.

Intron insertion is the most widely reported method of producing infectious potyviral clones. Though often effective, one drawback is that intron insertion has the potential to slow the onset of viral symptoms following inoculation of the host plant (Yang et al., 1998). As well, intron insertion increases the size of the viral genome, which limits the size of fragments that can be inserted in a virus vector for over-expression within a plant (Kelloniemi et al., 2008). PkMV has 58 AG/G splice acceptor/donor sequences as potential intron insertion sites in the sense DNA strand and 27 in the antisense DNA strand of the 3.5 Kb central toxic region alone. Since infectious viral clones exist as dsDNA, a cryptic promoter can occur on either the sense or

antisense strand of DNA. It is not known whether toxicity in *E. coli* occurs specifically from the transcription and translation of viral proteins, or an abundance of short peptides translated from any of the open reading frames in the viral genome. As well, instability of a plasmid may be due to the formation of double-stranded RNA intermediates with regulatory roles. Therefore, the presence of cryptic promoters should be considered in both strands of a viral clone. Only attempting intron insertion in the sense strand risks significant time and effort loss if no infectious clones are recovered due to persistent instability in *E. coli*.

Curing the toxic promoters directly by silent mutation is therefore the preferred method for obtaining stable infectious clones. However, due to the large genome size, direct silent mutation of a potyvirus genome has only been reported for potato virus Y, in which the authors silenced an almost perfect Sigma 70 promoter sequence within the virus genome (Ali et al., 2011). An improved strategy for finding cryptic promoters was developed by Pu et al., 2011 who cloned individual 200 bp portions of a 3 Kb region of dengue fever virus to detect cryptic promoters. This method requires a previously cloned region of the virus in question and would not be tenable for the full-length dengue fever genome, or a potyviral genome, which are approximately 10 Kb in length. Instead of a luciferase-based reporter, our method employs eGFP and needs no additional reagents besides a 470 nm LED for viewing, observable by eye or with a plate reader if available.

The low-throughput eGFP-based method resulted in a quarter of sequence reads aligning to the broad toxic region centered around the CI cistron, which reduced the promotor region causing toxicity to 1.2 Kb. Interestingly, the majority of reads aligned elsewhere in the PkMV genome, primarily in the CP coding region. This region has not been previously reported to be toxic to bacteria, suggesting that not all cryptic bacterial promoters induce toxicity. The

identification of promoter regions was accomplished by sequencing approximately 40 bacterial colonies. More colonies can be screened and sequenced to generate a higher resolution map of the viral genome. Our method is also scalable depending on the resources available. The chloramphenicol reporter allowed us to utilize Illumina sequencing of surviving colonies bearing fragments of the PkMV genome. Sequencing reads generated from the high-throughput and low-throughput methods generally overlapped within the CI and CP coding regions of the PkMV genome indicating that either screen will identify active promoters. The higher resolution afforded by millions of aligned sequence reads revealed new regions of promoter activity within the P1, HC-Pro and P3 coding regions of the PkMV genome, suggesting that cryptic promoters are quite common within DNA sequences. The Illumina sequencing discovered 16 putative promoter regions (PPRs) between 150-250 bp in the PkMV genome, significantly reducing the search window for promoters from the 9.5 Kb viral genome. Interestingly, PPRs 6-11 fall within the CI coding region, which is a popular choice for intron insertion to stabilize potyviral clones (Gao et al., 2012; Olsen & Johansen, 2001).

The eGFP reporter plasmid was used to easily determine if an active promoter remained in the increasingly small promoter region. During this refinement phase, the strand orientation of the promoter may be determined by inserting it in the forward and reverse orientation into the reporter. As expected, the identified PPRs were active either in the forward or reverse direction, which is typical of Sigma 70 promoters that are active in one direction only. Interestingly, PPR 6, which had a well-defined peak from the high-throughput screen, showed very low promoter activity. This could be caused by certain regions of the PkMV RNA being more accessible during reverse transcription than others, resulting in a biased number of reads.

PPRs obtained in this way were small enough to identify 50 bp of the promoter sequence upstream of the transcription start site. It was convenient then to synthesize the promoters as oligos and test them individually for activity, with and without silent mutations. Interestingly, PPR 12R, which promoted eGFP expression when cloned as a larger 3.9 Kb region, resulted in minimal fluorescence when the promoter was isolated from that region following TSO-5'RACE. Therefore, the level of promoter activity in a DNA sequence may be additive from several weaker promoters or regulatory elements which enhance the main promoter. The synthesis of putative promoters as oligos proved to be a rapid way of testing the strength of the promoter and the effectiveness of silencing prior to investing effort in generating the full-length infectious clone. It is interesting that either mutation of 10F or 11R promoter stabilized the viral clone and questions the hypothesis that toxicity in *E. coli* is due to the accumulation of toxic proteins, as the silencing of one promoter would not prevent the production of a toxic protein from another. Additionally, no ORFs greater than 10 amino acids were found near the cured promoters. Instead, it is possible that strong promoters like 10F and 11R produce large amounts of transcript RNA that can form double-stranded intermediates, triggering an RNAi-like response in the cell.

This method is not only applicable to potyviruses, but to large mammalian viruses such as flaviviruses, whose members include West Nile and zika virus. The genomes of flaviviruses are ssRNA and exceed 10 Kb, making them very difficult to clone. Other notable examples include members of the *Coronaviridae*, which can have genomes up to 30 Kb (Sah et al., 2020). The ability to rapidly screen a newly emerging virus for toxic, cryptic promoter activity will accelerate the generation of infectious clones and thus potential therapeutics. The screening method described here is also useful for identifying cryptic promoters in non-viral DNA sequences such as plasmids that require low levels of background transcription. For example,

cryptic promoters may artificially inflate expression levels measured by reporter constructs and hamper the study of weak promoters. In addition, expression plasmids encoding proteins highly toxic to *E. coli*, such as phage T4 endoribonuclease RegB (Saida et al., 2006), can be improved by silencing cryptic promoters in the plasmid backbone.

Altogether, we have shown a screening method for identifying cryptic promotors that is flexible depending on the available resources. Low-throughput screening can be achieved with simple molecular biology tools and low-cost blue light illumination. With access to Illumina sequencing and bioinformatics analysis, a high-resolution map of promoter activity can be established with minimal labor. Both screens can filter an entire potyviral genome to a small number of PPRs. These regions can then be subjected to TSO-5'RACE to determine the approximate transcription start site, and therefore sequence of the promotor contained upstream. Promoters can then be silently mutated to speed the development of a stable, infectious viral clone.

3.6 - Acknowledgements

This work was supported by a Discovery Grant to K.A.H. from the Natural Sciences and Engineering Research Council of Canada, and CGS-D Scholarship to A.K.

Chapter 4:

Discussion and future directions

4.1 - Generation of the first PkMV infectious clone

We have reported the construction of the first infectious clone of PkMV, which represents a new molecular tool to study the pokeweed plant and the virus itself. The impetus for the research was that no reverse genetics tools were available to study pokeweed, which is known for its synthesis of a potent antiviral protein as well as a heavy metal hyperaccumulator. To our knowledge, no transgenic pokeweed plants have been generated, with pokeweed callus and cell culture being the furthest that researchers have gone down this path (Kobayashi et al., 1995). As well, Phytolaccaceae species are known to be resistant to agroinfiltration (Kanzaki et al., 1999), cutting off a potential avenue for manipulation of the plant.

Cloning of PkMV was challenging from the beginning. The strain chosen to be our wild-type was the Arkansas (AR) strain, originally deposited by Dr. R.J. Sheppard to ATCC in 1970. The nucleotide sequence of this strain was unknown, so 3' RACE and sequencing by primer walking was performed until the 5' end of PkMV was reached. This proved effective in generating the initial sequence that was used for the first attempts at cloning PkMV (N. Yu et al., 2020).

The methods used to clone PkMV were a progressive escalation from classic protocols to modern cloning techniques. We had begun by cloning equally sized portions of PkMV into cloning vectors before assembly with restriction enzymes. This sequential assembly with restriction sites was one of the first methods developed for cloning viruses (Domier et al., 1989). Then, 4-fragment Gibson assembly was attempted along with the more obscure technique of homologous recombination in yeast (Sun et al., 2017). While Gibson assembly uses a mixture of exonuclease, ligase and polymerase to assemble overlapping DNA fragments (Gibson et al., 2009), assembly with yeast is performed by transforming overlapping fragments and relying on

homologous recombination to piece them together. All plasmids isolated from *E. coli* and yeast were non-infectious either due to point mutations, frame shifts, or large deletions within the viral polyprotein. Presence of mutations in the viral clones were assayed with either commercial sequencing by primer walking or through NGS sequencing of the viral plasmids. Clones with the insertion of the IV2 intron, derived from the potato ST-LS1 gene (Johansen, 1996) into the CI coding region were similarly not infectious. This was surprising as intron insertion is one of the best described ways of curing potyviruses (González et al., 2002; Johansen, 1996; López-Moya & García, 2000). Potentially, the lack of infectivity could have been an artifact of individually cloning portions of PkMV into plasmids before assembly, allowing additional chances for mutations to accumulate. At this point, it was clear that unlike the researchers who cloned LMV without needing to stabilize their clone (Yang et al., 1998), we were encountering the classic symptoms of plasmid toxicity in *E. coli*.

The major breakthrough came with a paper that described bypassing the manipulation of a viral clone in *E. coli* and instead performing the assembly in *Agrobacterium* directly (Tuo et al., 2017). This time, we also avoided the sub-cloning of PkMV genome fragments into cloning vectors. All fragments of PkMV were amplified with RT-PCR from viral RNA and used directly for assembly. This strategy allowed us to generate our first infectious clone, PkMV-Ag01, which was shown to be infectious through the production of viral particles, protein and was aphid transmissible (Klenov & Hudak, 2018).

While obtaining the first infectious clone was extremely encouraging, the goal of cloning PkMV was to produce a tool capable of delivering foreign proteins into pokeweed. Typically, potyviruses can be engineered to carry foreign proteins within their polyproteins, which are cleaved and released by virally-encoded proteases during infection. The polyprotein cleavage

sites were determined after comparing their protein sequences to previously published cleavage sites of potyviruses (Adams et al., 2005). It was determined that the cleavage site between P1 and HC-Pro of the polyprotein was optimal for insertion of a foreign protein sequence to generate PkMV-eGFP. This is a popular insertion site for foreign proteins in potyviruses (Beauchemin et al., 2005). Insertion between the NIb and CP coding regions did not prove to be infectious. Fluorescence was photographed throughout the plant using a custom-built setup. For exciting eGFP, a royal blue LED diode (450nm) was filtered through a 470 nm short-pass filter. The emitted light was passed through a 510/20 nm bandpass filter, which effectively removed the excitation wavelengths as well as the red autofluorescence typical of plant chlorophylls.

These experiments with PkMV-eGFP showed that our viral clone was infectious and capable of delivering a foreign protein into pokeweed. At this point, we began work to clone other proteins of interest into PkMV, as well as convert the clone into a VIGs vector to knock down pokeweed genes. Unfortunately, the difficulties of manipulating an increasing number of clones in *Agrobacterium* began to accumulate.

When compared with *E. coli*, *A. tumefaciens* has a generation time 3-fold longer (Wessel et al., 2006). Therefore, fewer transformations can be performed in a given period. To confirm that a PkMV clone is full length, colony PCR performed directly on *Agrobacterium* colonies is often unpredictable. For more reliable screening, a small culture must be grown and the plasmid DNA extracted from the culture. The plasmids that replicate within *Agrobacterium* are at a very low copy number, so the yield of plasmid is very low. As well, the plasmid is poor quality, as standard miniprep methods are optimized for *E. coli*. While *Agrobacterium* had served its purpose for generating an infectious clone, the issue of toxicity in *E. coli* had to be dealt with directly.

4.2 - Developing a method to detect and cure cryptic promoters in PkMV

Toxicity of viral sequences in *E. coli* has been observed from the very early days of infectious clone generation. While several indirect methods of stabilizing clones, such as alternate plasmids, bacterial strains and intron insertion were options (Pasin et al., 2019), we wanted to find and silence the cryptic promoters directly. The most successful example of this to date has been the curing of a cryptic promoter in dengue fever virus, in which 300 bp portions of a 3000 bp genome chunk were assayed for promoter activity (Pu et al., 2011). While effective for smaller portions of a virus, this method was inefficient for potyviruses which are 9.5-10 kb long. The method we have developed allows a researcher to start directly from viral RNA or from a previously isolated clone, determine the location of cryptic bacterial promoters and silence them.

Toxicity in *E. coli* during cloning is thought to be due to the transcription and translation of viral sequences (Guo et al., 2007; Li et al., 2011; Pluchino et al., 2015). In *E. coli*, transcription is initiated by binding of the initiation factor to the bacterial promoter, allowing RNA polymerase to bind. There are seven distinct sigma factors in *E. coli*, each of which recognizes a subset of promoters under specific conditions. The Sigma 70 factor is the most abundant initiation factor that binds promoters during the growth of the bacteria. Therefore, if there was a cryptic promoter somewhere in the PkMV genome, it would likely be recognized by the Sigma 70 factor, whose consensus is TTGACA(-35) and TATAAT(-10). Originally we thought it would be possible to use *in silico* prediction software, such as PromoterHunter (Klucar et al., 2009), to find Sigma 70 binding sites. Although these algorithms are adequate for predicting potential promoters in small 250 bp regions, the number of potential binding sites in PkMV was intractable. If we could filter down the PkMV genome to a few transcriptionally active regions, then perhaps the *in silico* prediction would be more helpful.

To reduce the search window for cryptic promoters, we constructed a reporter plasmid carrying a Golden Gate insertion site upstream of an eGFP ORF (Figure 15). This construct would allow me to quickly insert a DNA library of any size upstream of eGFP. If the inserted fragment had a promoter that was recognized by the bacteria, then some amount of eGFP would be produced. By selecting colonies that produced eGFP fluorescence, we could enrich for DNA containing promoters.

Before the assay was performed, the fragment library of PkMV had to be generated. A significant amount of optimization was required to determine the best primers to use for random cDNA library generation both at the RT step and subsequent PCR optimization. Methods for amplification of random libraries have been described previously (Froussard, 1992; Zou et al., 2003) that employ random primer at either the RT or PCR step. The goal of both methods was to amplify cDNA or DNA libraries from low starting amount of materials, and provided optimal PCR cycling parameters and primer concentrations for that purpose. In the end, the optimal library prep method involved reverse transcription of the viral RNA with a poly-d(T)23-VN primer, followed by amplification with random hexamers bearing Golden Gate adapter sequences. Quality of library generation was judged on an agarose gel, ideally a random sampling of the entire PkMV genome by PCR would produce a smear versus defined bands. The number of PCR cycles was also kept intentionally low, between 8-12, to reduce PCR bias (Aird et al., 2011) This bias manifests itself as depletion of PCR products with high GC content as the number of cycles increase.

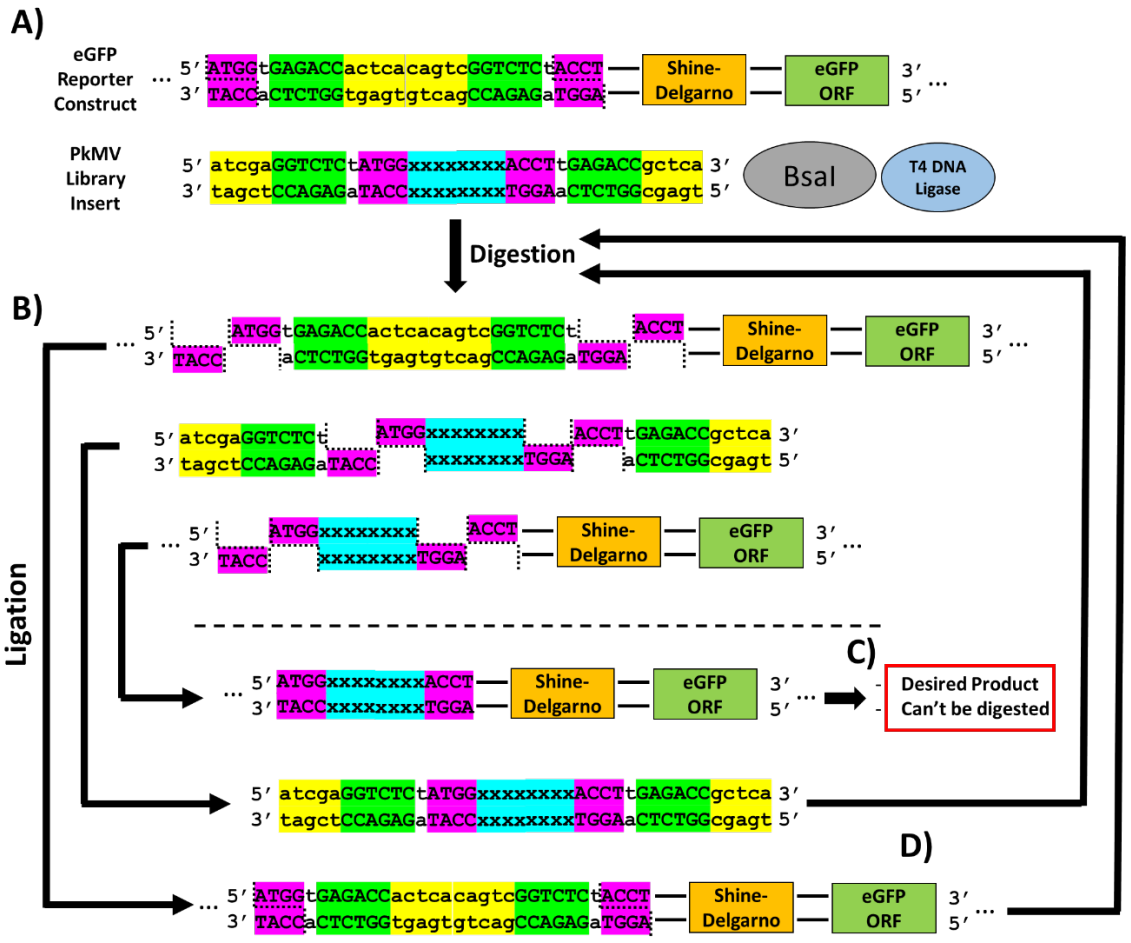

Figure 15 – Overview of Golden Gate assembly based eGFP reporter assay. Golden Gate technique enables a one pot assembly of DNA without prior linearization of the vector backbone. A) Circular eGFP library construct, PCR amplified insert containing PkMV library sequence (blue highlight), BsaI restriction enzyme and T4 DNA ligase are combined in a single tube and cycled between 16°C and 37°C. B) During the endonuclease cycle at 37°C, the reporter vector and insert are digested with BsaI, whose recognition site (green highlight) is distal of its cleavage site (pink highlight). Efficient cleavage requires a landing pad of at least 3 nt (yellow) 5' of the recognition site. After reducing the temperature to 16°C, three ligations will occur, two re-ligations of the cleaved vector and insert, and one between the vector and the desired insert. C) The desired ligation product between the vector and insert is a dead-end reaction as the restriction enzyme site is abolished. D) Vector or inserts that have re-ligated are recycled into the reaction and digested, as the BsaI restriction sites are retained.

Golden Gate assembly was chosen for cloning of the viral fragment library into the reporter constructs. For this application, Golden Gate assembly was ideal because of the scar-less insertion of the PkMV fragments into the reporter construct, as well as the lack of exonuclease in the enzyme mix. Gibson assembly type methods include an exonuclease to digest the ends of overlapping DNA to create sticky ends. For fragments <300 bp, the exonuclease can chew back the entire insert, creating a ssDNA fragment incapable of participating in the cloning reaction.

After cloning the PkMV fragment library into the eGFP reporter construct, the PkMV plasmid library was transformed into *E. coli*. While it only took a day for colonies to appear, observable amounts of eGFP accumulated over 2-3 days, which is typical for fluorescent reporter assays with bacterial colonies (Feilmeier et al., 2000). Several different library sizes were cloned into the reporter, from 150-250 bp fragments up to 650-750 bp. As fragment size increased, there was a concurrent increase in the frequency of fluorescent colonies, reaching a peak with insert sizes of 300-400 bp. Larger fragments than this actually produced less fluorescent colonies. Essentially, larger fragments are more likely to contain a promoter-like sequence, but fragments can get long enough to be too far from the eGFP ORF to be an effective promoter. This coincides with most *E. coli* promoters being within 250 bp of the ORF in the bacteria (Huerta & Collado-Vides, 2003). Picking individual colonies and sequencing the insert reveals PkMV sequences that induce transcription of eGFP, and thus contain a cryptic promoter. Upon aligning the sequence reads back to the PkMV genome, this low-throughput method can start defining putative promoter regions. Ideally, the inserts sequenced would be as short as possible to provide a small search window. However, this means that more bacteria need to be plated to get the same amount of colonies as on a large insert size plate. Sequencing 40 colonies yielded four putative promoter regions (PPRs) between 200-500 bp in size. This was a significantly reduced search

window from the 9.5 kb of the PkMV genome, and comparable to the search windows manually chosen when curing 3 kb of the dengue fever virus (Pu et al., 2011). These PPRs were used as input for the PromoterHunter software yet again, and 12 putative promoters were identified, only three of which were true promoters. This was still a fairly poor prediction rate that required the validation of many putative promoters.

To reduce the search window for putative promoters further and to reduce the manual labor of picking colonies and sequencing, a NGS approach was applied. First, a reporter construct with the chloramphenicol resistance ORF downstream of the PkMV library insertion site was constructed. Any cryptic promoter inserted will drive expression of the chloramphenicol ORF and allow survival of bacteria carrying this plasmid on chloramphenicol. A similar approach was applied when researchers were evolving bacterial promoters from random sequences, wherein they cloned random 19 bp sequences into the -35 position upstream of a tetracycline ORF to find active promoters (Horwitz & Loeb, 1986). This technique allowed the researchers to find active promoters that were significantly diverged in sequence and spacing from the consensus Sigma 70 binding site. They also found promoters diverged from the Sigma 70 consensus with stronger promoter activity, mirroring a few of the strong promoters we found in this work.

It took an unusually high level of chloramphenicol (>200 μg/mL) to begin to see a difference between negative control plates carrying a scrambled promoter insert and plates with colonies carrying fragments of the PkMV genome. As we later discovered, this was partially explained by a cryptic promoter in the reporter plasmid itself, driving a very small amount of chloramphenicol expression but enough to increase the background of surviving colonies. Several plates with high levels of chloramphenicol were scraped, representing several thousand

individual colonies, and plasmids were isolated *en masse*. The sequencing company required linear, dsDNA for sequencing, so the plasmid pool was amplified with low cycle PCR using primers flanking the fragment insertion site. dsDNA was sequenced using 150 bp paired-end reads and the sequences were aligned to the PkMV genome with the aligner Bowtie2 through the North American Galaxy server (Giardine et al., 2005; Langmead et al., 2009). The output binary alignment map (BAM) file was visualized with the Integrative Genomics Viewer (Thorvaldsdóttir et al., 2013). Viewing the alignment showed that PkMV had 16 PPRs distributed throughout its genome. To our knowledge, this was the first high resolution promoter map of a large RNA virus. As well, each individual PPR was only 150-300 bp, about half the size of the PPRs obtained with the low-throughput method. Not only did the high throughput method largely agree with the low throughput, there was an over-representation of reads in the middle of the PkMV genome, which we have shown to be preferentially mutated during cloning. As well, this region contains the CI coding sequence, which has been hypothesized to be the source of toxicity in potyviruses (Ali et al., 2011; Jakab et al., 1997). The mechanism behind this toxicity is not understood at this time. Hypothetically the same cylindrical inclusions that are tolerated in larger plant cells could potentially disrupt the *E. coli* plasma membrane.

While the promoter prediction software had successful predicted active promoters, a more accurate and direct approach was to perform 5'RACE on the PPRs. Traditional 5'RACE normally involves ligation of adapters either to RNA or cDNA and is quite laborious for large amounts of samples. Template switching oligo (TSO)-5'RACE allows the addition of an adapter during the reverse transcription step due to the template switching properties of the reverse transcriptase. This technique allowed me to extract total RNA from *E. coli* colonies carrying a PPR upstream of an eGFP ORF, perform TSO-5'RACE with an eGFP specific reverse primer

and obtain the +1 transcription start site (+1 TSS) of the promoter contained within. Even though prokaryotic mRNAs are uncapped, the reverse transcriptase used in the TSO-5'RACE protocol will still add several untemplated C's to the nascent cDNA strand, allowing the adapter oligo to anneal.

This technique permitted us to identify the +1 TSS of each PPR to within a few nucleotides. The ambiguity came from the observation that the reverse transcriptase did not always add three C's to the cDNA, the number ranged from two to four (Wulf et al., 2019). As well, the typical -35 and -10 boxes used to denote the binding site of the Sigma 70 factor were not always the same distance from the transcription start site. Therefore, the predictive power of this method can narrow down a 9.5 Kb RNA genome to promoter regions of just 50 bp. At this resolution, the -35 and -10 boxes can usually be detected by eye or with PromoterHunter (Klucar et al., 2009). Throughout this method, the eGFP reporter plasmid was extremely useful for testing the promoter activity of any DNA fragment. The reporter was used to test the large PPRs for activity, in both forward and reverse orientation, as well as the isolated 50 bp promoter regions and their silenced counterparts. Any 50 bp putative promoter region to be tested was simply synthesized as overlapping primers and inserted into the reporter construct in a one-pot reaction. The 50 bp promoters isolated with the TSO-5'RACE were all examined for the presence of a -35 and -10 box and mutated to abolish any resemblance to the Sigma 70 consensus binding site while retaining the original amino acid sequence of PkMV. It must be noted that additional caution should be observed for viruses that heavily utilize secondary structure, frame-shift proteins and sub-genomic RNAs for their replication. In those cases, a change in nucleotide sequence may have a significant impact on the biology of the virus. At the moment, PkMV is only known to produce the frame-shift protein P3N-PIPO, which arises from a

polymerase slippage site in the P3 coding region (Atsumi et al., 2016). Fortuitously, none of the detected PPRs were directly overlapping with the slippage site or coding region of P3N-PIPO in PkMV.

We chose to focus on PPRs 6-12 since they fell directly into the CI coding region, which is known to be a problematic region to clone for potyviruses. We discovered that promoters within regions 10F, 11R and 12F were highly active, producing levels of eGFP comparable and greater than the positive control, a consensus Sigma 70 binding site. After silencing the -35 and -10 boxes of these promoters within in the PkMV genome, the stability of the clone and the infectivity was tested. Surprisingly, infectious clones were recovered for all three silenced promoters. This result challenges the notion that toxicity in *E. coli* from large viral sequences is due to translation of toxic peptides. If that were the case, we would expect only a single cryptic promoter to be responsible for the observed toxicity. We analyzed all three reading frames downstream of the +1 TSS for promoters 10F, 11R and 12F (Figure 16). Promoters 10F and 12F did not have a peptide ORF in close proximity to the +1 TSS. Promoter 11R had two potential ORFs, one 6 AA long, the other 54 AA, 12-16bp downstream of the TSS with no Shine-Delgarno (SD) sequence. It is possible that a peptide can be translated in a SD independent manner, and ATG is not always the start codon during translation (Hecht et al., 2017; Skorski et al., 2006). However, the chances of having three unique promoters that all transcribe RNA capable of SD independent translation upstream of a peptide toxic to *E. coli* are low. Moreover, while promoters 10F and 12F appear on the top strand and could cause the transcription and translation of a truncated PkMV protein, 11R is found on the bottom strand. This means that any peptide produced from 11R would not share the amino acid sequence of any PkMV protein.

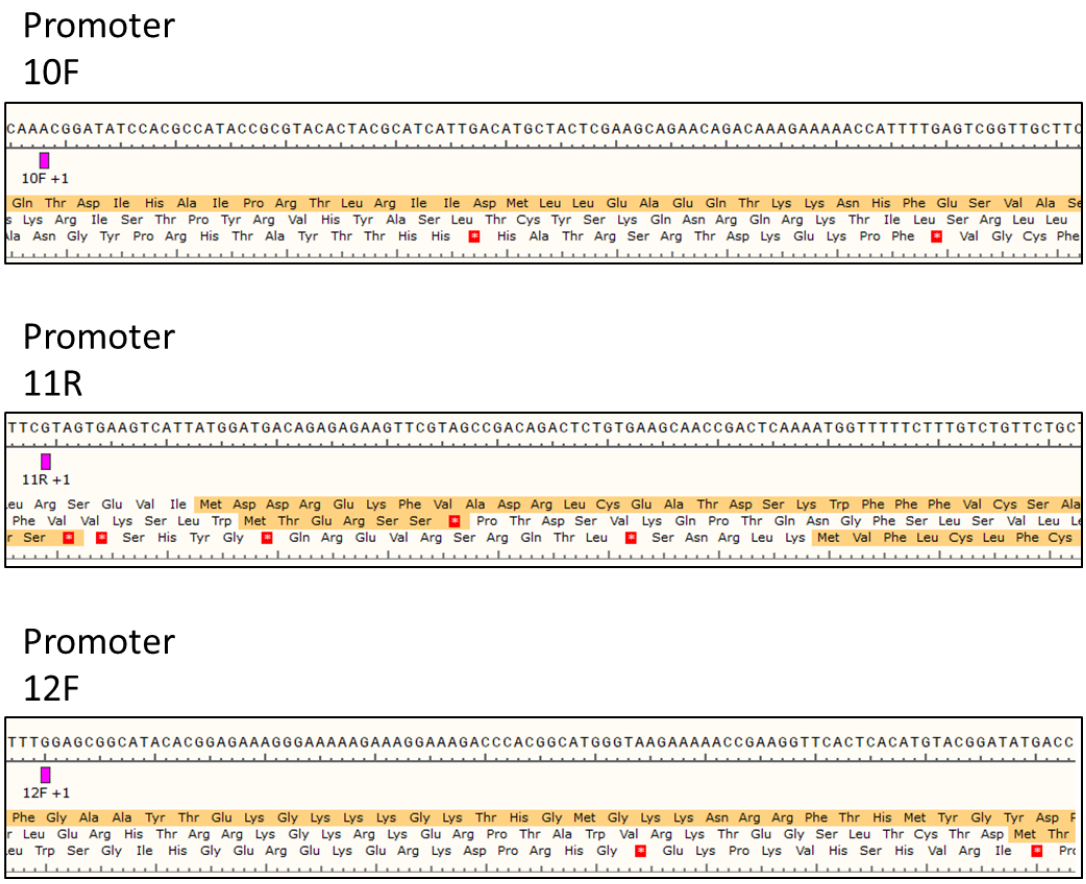

Figure 16 – ORF analysis of cryptic promoters. All three open reading frames downstream of the +1 TSS (Pink square) of promoters 10F, 11R, and 12F were analyzed for coding potential of toxic peptides. ORFs beginning with an ATG and ending with a stop codon (red blocks) are highlighted in beige.

Instead, we propose that the observed plasmid instability is due to accumulation of transcribed RNAs that form double-stranded intermediates that are processed and induce mutations in plasmid DNA through a mechanism mediated by clustered regularly interspaced short palindromic repeats (CRISPR) (Figure 16). Some components of small RNA processing are shared between eukaryotes and prokaryotes, including an enzyme analogous to human dicer, RNase III (Jin et al., 2019). This enzyme recognizes dsRNA and cleaves it into 11-22 bp fragments. Spurious transcription from cryptic promoters at multiple locations on both strands of viral DNA would provide a source of dsRNA, so the processing of plasmid-derived dsRNA is plausible. As well, the CRISPR prokaryotic immunity system was originally discovered in *E. coli* (Touchon et al., 2011). This system works in three stages, the adaptive stage during which protospacers are acquired from foreign dsDNA and integrated into the CRISPR transcriptional cassette, the transcription and processing of the pre-CRISPR-RNA from the protospacers, and finally targeting of the Cas protein to the dsDNA for cleavage (Koonin & Makarova, 2013). While much work has focused on the acquisition of protospacers from dsDNA breaks (Amitai & Sorek, 2016), it is possible that the plasmid-derived RNAs are being integrated as protospacers by a reverse transcriptase-Cas fusion protein (Silas et al., 2016). These spacers would be transcribed and processed into guide RNAs that would bring the Cas proteins to the viral plasmid, causing cleavage and the observed instability.

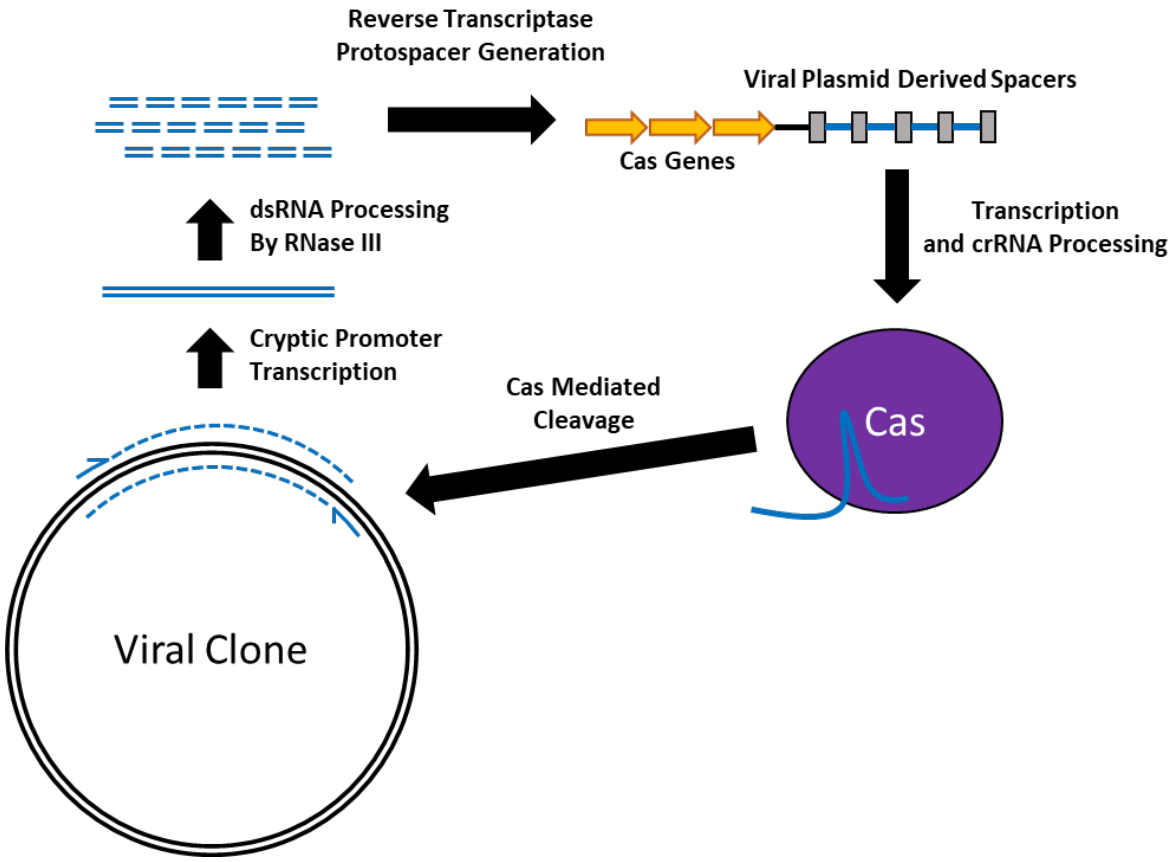

Figure 17 – Model for plasmid instability in *E. coli* mediated by CRISPR. Large viral clones contain numerous cryptic promoters on both strands of the dsDNA plasmid. RNA is transcribed from cryptic promoters on both DNA strands, forming dsRNA intermediates that are processed by RNase III, a prokaryotic dicer analog. The small fragments of RNA are reverse transcribed by a reverse transcriptase-Cas fusion protein and integrated into the spacer/repeat region of the Cas array. After transcription and processing of the spacer/repeat region, the guide RNA with Cas protein cleaves the viral clone, inducing instability.

While infectious clones were obtained from all three silenced promoters, the growth patterns of the *E. coli* colonies were different among them. From our observations, every transformation of a PkMV clone into *E. coli* produced agar plates with both large and small colonies. The larger colonies contained non-infectious plasmids, while the smaller colonies harboured the full-length infectious clones. Interestingly, PkMV with silenced promoters 10F and 11R had a greater number of small colonies than 12F. This suggests that while a single promoter may not dictate whether a plasmid is toxic or not, some promoters may contribute more to the toxicity than others. So, there is a threshold of transcriptional activity within a plasmid above which toxicity can be observed. Plasmids below that threshold can be obtained and manipulated without a high risk of plasmid mutation, while recovering plasmids with increasingly more promoter activity will be challenging.

In summary, we have developed a method to detect and cure cryptic bacterial promoters in large viral genomes. We have taken the PkMV genome in its entirety and discovered a handful of active promoters which when silenced allowed recovery of fully infectious clones. This work also proposes a new mechanism for plasmid toxicity in *E. coli*.

4.3 - Future directions

There are two exciting fronts on which this work can continue: the further development of the PkMV clone as a tool to study pokeweed and the elucidation of the mechanisms that cause plasmid induced toxicity in *E. coli*.

Now that the PkMV infectious clone is stably manipulated in *E. coli*, work has restarted to introduce VIGS functionality into the clone. VIGS silencing in plants is not typically done using potyviruses, with the concern being that the HC-Pro silencing suppressor contained within the viral genome may interfere with effective silencing in the plant. However, VIGS with a potyvirus has been shown once before; researchers used it to silence endogenous GFP in transgenic tobacco (Gammelgard et al., 2007). Phytoene desaturase (PDS) will be the first gene to be silenced, as it is the classic control for VIGS vectors due to the easily detected white, bleached phenotype in the leaves of infected plants. Knocking down levels of PAP will also be attempted again. Preliminary efforts were made to knock down PAP with PkMV VIGS before the curing of the toxic promoters. PAP levels were knocked down somewhat, however the difficulty in generating the VIGS clones prevented further optimization of the VIGS sequence.

Due to the effectiveness and ease of use of Golden Gate cloning during the promoter screening method, a Golden Gate site will be introduced into PkMV into the previously proven insertion site. This will allow rapid cloning of different VIGS targets without PCR linearization of the large vector.

Answering the question of the exact nature of toxicity of large viral sequences is another long-term goal. First, evidence is needed that there is RNA transcript accumulation aligning back to the plasmid, which will be obtained with RNA-seq experiments. As well, the small RNAs

present within *E. coli* will be sequenced, to see if processing into small RNA species is occurring. Creating mutant *E. coli* strains with knocked-out genes involved in small RNA biogenesis would also help tease apart the molecular mechanisms involved in plasmid toxicity. As well, ChIP-seq could be performed with antibodies specific to *E. coli* Cas proteins, to determine if the mutations observed in plasmids with a high level of promoter activity are due to Cas-induced cleavage. Another interesting experiment would be to sequence the spacer/repeat array of the *E. coli* Cas proteins to determine if parts of the PkMV genome are being integrated as protospacers.

We would also like to perform protein mass-spectrometry experiments to confirm or reject the hypothesis that plasmid toxicity in *E. coli* is due to overproduction of viral peptides. Essentially, we would align mass-spec results back to the PkMV polyprotein to determine whether any such peptides exist. There may be a time sensitive factor to these experiments, as the cleavage and mutation of a toxic plasmid may occur shortly after bacterial transformation, so a time course for any of these assays would be important.

References

Aboughanem-Sabanadzovic, N., Allen, T. W., Henn, A., Lawrence, A., & Sabanadzovic, S. (2019). Partial characterization of a divergent isolate of pokeweed mosaic virus from Mississippi. *Journal of Plant Pathology*, *101*(4), 1127–1131. https://doi.org/10.1007/s42161-019-00277-8

Abrahamian, P., Hammond, R. W., & Hammond, J. (2020). Plant Virus–Derived Vectors: Applications in Agricultural and Medical Biotechnology. *Annual Review of Virology*, *7*(1). https://doi.org/10.1146/annurev-virology-010720-054958

Adams, M. J., Antoniw, J. F., & Beaudoin, F. (2005). Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Molecular Plant Pathology*, *6*(4), 471–487. https://doi.org/10.1111/j.1364-3703.2005.00296.x

Aird, D., Ross, M. G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. B., Nusbaum, C., & Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*(2), R18. https://doi.org/10.1186/gb-2011-12-2-r18

Albrechtsen, M., & Heide, M. (1990). Purification of plant viruses and virus coat proteins by high performance liquid chromatography. *Journal of Virological Methods*, *28*(3), 245–256. https://doi.org/10.1016/0166-0934(90)90118-y

Ali, Z., Abul-Faraj, A., Piatek, M., & Mahfouz, M. M. (2015). Activity and specificity of TRV-mediated gene editing in plants. *Plant Signaling and Behavior*, *10*(10), 1–4. https://doi.org/10.1080/15592324.2015.1044191

Allard, H. A. (1918). The mosaic disease of Phytolacca decandra. *Phytopathology*, *8*, 51–54.

Amitai, G., & Sorek, R. (2016). CRISPR-Cas adaptation: Insights into the mechanism of action. *Nature Reviews Microbiology*, *14*(2), 67–76. https://doi.org/10.1038/nrmicro.2015.14

Atsumi, G., Suzuki, H., Miyashita, Y., Choi, S. H., Hisa, Y., Rihei, S., Shimada, R., Jeon, E. J., Abe, J., Nakahara, K. S., & Uyeda, I. (2016). P3N-PIPO, a Frameshift Product from the P3 Gene, Pleiotropically Determines the Virulence of Clover Yellow Vein Virus in both Resistant and Susceptible Peas. *Journal of Virology*, *90*(16), 7388–7404. https://doi.org/10.1128/JVI.00190-16

Aubry, F., Nougairède, A., Gould, E. A., & De Lamballerie, X. (2015). Flavivirus reverse genetic systems, construction techniques and applications: A historical perspective. *Antiviral Research*, *114*, 67–85. https://doi.org/10.1016/j.antiviral.2014.12.007

Beauchemin, C., Bougie, V., & Laliberté, J. F. (2005). Simultaneous production of two foreign proteins from a potyvirus-based vector. *Virus Research*, *112*(1–2), 1–8. https://doi.org/10.1016/j.virusres.2005.03.001

Bendahmane, M., Koo, M., Karrer, E., & Beachy, R. N. (1999). Display of epitopes on the surface of tobacco mosaic virus: impact of charge and isoelectric point of the epitope on virus-host interactions. *Journal of Molecular Biology*, *290*(1), 9–20. https://doi.org/10.1006/jmbi.1999.2860

Bentley, K. E., Berryman, K. R., Hopper, M., Hoffberg, S. L., Myhre, K. E., Iwao, K., Lee, J. B., Glenn, T. C., & Mauricio, R. (2015). Eleven Microsatellites in an Emerging Invader, Phytolacca americana (Phytolaccaceae), from Its Native and Introduced Ranges. *Applications in Plant Sciences*, *3*(3), 1500002. https://doi.org/10.3732/apps.1500002

Berger, P. H., & Shiel, P. J. (1998). Potyvirus Isolation and RNA Extraction. In G. D. Foster & S. C. Taylor (Eds.), *Plant Virology Protocols* (pp. 151–160). Humana Press. https://doi.org/10.1385/0-89603-385-6:151

Blawid, R., & Nagata, T. (2015). Construction of an infectious clone of a plant RNA virus in a binary vector using one-step Gibson Assembly. *Journal of Virological Methods*, *222*, 11–15. https://doi.org/10.1016/j.jviromet.2015.05.003

Bordat, A., Houvenaghel, M. C., & German-Retana, S. (2015). Gibson assembly: An easy way to clone potyviral full-length infectious cDNA clones expressing an ectopic VPg Plant viruses. *Virology Journal*, *12*(1), 1–8. https://doi.org/10.1186/s12985-015-0315-3

Botta-Dukat, Z., Balogh, L., & Feher, A. (2008). American and Chinese pokeweed (Phytolacca americana, Phytolacca esculenta). In Z. Botta-Dukat & L. Balogh (Eds.), *The Most Important Invasive Plants in Hungary* (pp. 35–46). HAS Institute of Ecology and Botany.

Bredenbeek, P. J., Kooi, E. A., Lindenbach, B., Huijkman, N., Rice, C. M., & Spaan, W. J. M. (2003). A stable full-length yellow fever virus cDNA clone and the role of conserved RNA elements in flavivirus replication. *Journal of General Virology*, *84*(5), 1261–1268. https://doi.org/10.1099/vir.0.18860-0

Brennan, F. R., Jones, T. D., Gilleland, L. B., Bellaby, T., Xu, F., North, P. C., Thompson, A., Staczek, J., Lin, T., Johnson, J. E., Hamilton, W. D. O., & Gilleland, H. E. (1999). Pseudomonas aeruginosa outer-membrane protein F epitopes are highly immunogenic in mice when expressed on a plant virus. *Microbiology*, *145*(1), 211–220. https://doi.org/10.1099/13500872-145-1-211

Brennan, Frank R, Gilleland, L. B., Staczek, J., Bendig, M. M., Hamilton, W. D. O., & Gilleland, H. E. (1999). A chimaeric plant virus vaccine protects mice against a bacterial infection. *Microbiology*, *145*(8), 2061–2067. https://doi.org/10.1099/13500872-145-8-2061

Chambers, T. J., McCourt, D. W., & Rice, C. M. (1990). Production of yellow fever virus proteins in infected cells: Identification of discrete polyprotein species and analysis of cleavage kinetics using region-specific polyclonal antisera. *Virology*, *177*(1), 159–174. https://doi.org/10.1016/0042-6822(90)90470-C

Charon, J., Theil, S., Nicaise, V., & Michon, T. (2016). Protein intrinsic disorder within the Potyvirus genus: from proteome-wide analysis to functional annotation. *Mol. BioSyst.* https://doi.org/10.1039/C5MB00677E

Chen, Y., Zhi, J., Zhang, H., Li, J., Zhao, Q., & Xu, J. (2017). Transcriptome analysis of Phytolacca americana L. in response to cadmium stress. *PLOS ONE*, *12*(9), e0184681. https://doi.org/10.1371/journal.pone.0184681

Chikh Ali, M., Said Omar, A., & Natsuaki, T. (2011). An infectious full-length cDNA clone of potato virus YNTN-NW, a recently reported strain of PVY that causes potato tuber necrotic ringspot disease. *Archives of Virology*, *156*(11), 2039–2043. https://doi.org/10.1007/s00705-011-1062-4

Choi, B., Kwon, S. J., Kim, M. H., Choe, S., Kwak, H. R., Kim, M. K., Jung, C., & Seo, J. K. (2019). A plant virus-based vector system for gene function studies in pepper. *Plant Physiology*, *181*(3), 867–880. https://doi.org/10.1104/pp.19.00836

Chung, B. Y.-W., Miller, W. A., Atkins, J. F., & Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(15), 5897–5902. https://doi.org/10.1073/pnas.0800468105

Cotton, S., Grangeon, R., Thivierge, K., Mathieu, I., Ide, C., Wei, T., Wang, A., & Laliberte, J.-F. (2009). Turnip Mosaic Virus RNA Replication Complex Vesicles Are Mobile, Align with Microfilaments, and Are Each Derived from a Single Viral Genome. *Journal of Virology*, *83*(20), 10460–10471. https://doi.org/10.1128/JVI.00819-09

Cruz, S. S., Chapman, S., Roberts, A. G., Roberts, I. M., Prior, D. A. M., & Oparka, K. J. (1996). Assembly and movement of a plant virus carrying a green fluorescent protein overcoat. *Proceedings of the National Academy of Sciences*, *93*(13), 6286–6290. https://doi.org/10.1073/pnas.93.13.6286

Dalsgaard, K., Uttenthal, A., Jones, T. D., Xu, F., Merryweather, A., Hamilton, W. D., Langeveld, J. P., Boshuizen, R. S., Kamstrup, S., Lomonossoff, G. P., Porta, C., Vela, C., Casal, J. I., Meloen, R. H., & Rodgers, P. B. (1997). Plant-derived vaccine protects target animals against a viral disease. *Nature Biotechnology*, *15*(3), 248–252. https://doi.org/10.1038/nbt0397-248

Davino, S., Panno, S., Rangel, E. A., Davino, M., Bellardi, M. G., & Rubio, L. (2012). Population genetics of cucumber mosaic virus infecting medicinal, aromatic and ornamental plants from northern Italy. *Archives of Virology*, *157*(4), 739–745. https://doi.org/10.1007/s00705-011-1216-4

Di, R. (2016). Complete Genome Sequence of the Pokeweed Mosaic Virus (PkMV)-New Jersey Isolate and Its Comparison to PkMV-MD and PkMV-PA. *Genome Announcements*, *4*(5), e00929-16. https://doi.org/10.1128/genomeA.00929-16

Djordjevic, M. (2011). Redefining Escherichia coli σ 70 Promoter Elements: -15 Motif as a Complement of the -10 Motif. *Journal of Bacteriology*, *193*(22), 6305–6314. https://doi.org/10.1128/JB.05947-11

Domier, L. L., Franklin, K. M., Hunt, A. G., Rhoads, R. E., & Shaw, J. G. (1989). Infectious in vitro transcripts from cloned cDNA of a potyvirus, tobacco vein mottling virus. *Proceedings of the National Academy of Sciences*, *86*(10), 3509–3513. https://doi.org/10.1073/pnas.86.10.3509

Dou, C. M., Fu, X. P., Chen, X. C., Shi, J. Y., & Chen, Y. X. (2009). Accumulation and detoxification of manganese in hyperaccumulator Phytolacca americana. *Plant Biology (Stuttgart, Germany)*, *11*(5), 664–670. https://doi.org/10.1111/j.1438-8677.2008.00163.x

Driver, M. G., & Francis, F. J. (1979). Purification of Phytolaccanin (Betanin) by removal of phytolaccatoxin from Phytolacca Americana. *Journal of Food Science*, *44*(2), 521–523. https://doi.org/10.1111/j.1365-2621.1979.tb03826.x

Dugdale, B., Mortimer, C. L., Kato, M., James, T. A., Harding, R. M., & Dale, J. L. (2013). In plant activation: an inducible, hyperexpression platform for recombinant protein production in plants. *The Plant Cell*, *25*(7), 2429–2443. https://doi.org/10.1105/tpc.113.113944

Endo, Y., Tsurugi, K., & Lambert, J. M. (1988). The site of action of six different ribosome-inactivating proteins from plants on eukaryotic ribosomes: The RNA N-glycosidase activity of the proteins. *Biochemical and Biophysical Research Communications*, *150*(3), 1032–1036. https://doi.org/10.1016/0006-291X(88)90733-4

Engler, C., Gruetzner, R., Kandzia, R., & Marillonnet, S. (2009). Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE*, *4*(5), e5553. https://doi.org/10.1371/journal.pone.0005553

Fakhfakh, H., Vilaine, F., Makni, M., & Robaglia, C. (1996). Cell-free cloning and biolistic inoculation of an infectious cDNA of potato virus Y. *Journal of General Virology*, *77*(3), 519–523. https://doi.org/10.1099/0022-1317-77-3-519

Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H., & Phillips, G. J. (2000). Green fluorescent protein functions as a reporter for protein localization in Escherichia coli. *Journal of Bacteriology*, *182*(14), 4068–4076. https://doi.org/10.1128/JB.182.14.4068-4076.2000

Fernández-Fernández, M.Rosario, Martínez-Torrecuadrada, J. L., Casal, J. I., & García, J. A. (1998). Development of an antigen presentation system based on plum pox potyvirus. *FEBS Letters*, *427*(2), 229–235. https://doi.org/10.1016/S0014-5793(98)00429-3

Fernández-Fernández, M R, Mouriño, M., Rivera, J., Rodríguez, F., Plana-Durán, J., & García, J. A. (2001). Protection of rabbits against rabbit hemorrhagic disease virus by immunization with the VP60 protein expressed in plants with a potyvirus-based vector. *Virology*, *280*(2), 283–291. https://doi.org/10.1006/viro.2000.0762

Froussard, P. (1992). A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nucleic Acids Research*, *20*(11), 2900. https://doi.org/10.1093/nar/20.11.2900

Futterer, J., Bonneville, J. M., & Hohn, T. (1990). Cauliflower mosaic virus as a gene expression vector for plants. *Physiologia Plantarum*, *79*(1), 154–157. https://doi.org/10.1111/j.1399-3054.1990.tb05878.x

Gal-On, A., Meiri, E., Huet, H., Hua, W. J., Raccah, B., & Gaba, V. (1995). Particle bombardment drastically increases the infectivity of cloned DNA of zucchini yellow mosaic potyvirus. *Journal of General Virology*, *76*(12), 3223–3227. https://doi.org/10.1099/0022-1317-76-12-3223

Gammelgard, E., Mohan, M., & Valkonen, J. P. T. (2007). Potyvirus-induced gene silencing: the dynamic process of systemic silencing and silencing suppression. *Journal of General Virology*, *88*(8), 2337–2346. https://doi.org/10.1099/vir.0.82928-0

Gao, R., Tian, Y. P., Wang, J., Yin, X., Li, X. D., & Valkonen, J. P. T. (2012). Construction of an infectious cDNA clone and gene expression vector of Tobacco vein banding mosaic virus (genus Potyvirus). *Virus Research*, *169*(1), 276–281. https://doi.org/10.1016/j.virusres.2012.07.010

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W. J., & Nekrutenko, A. (2005). Galaxy: A platform for interactive large-scale genome analysis. *Genome Research*, *15*, 1451–1455. https://doi.org/10.1101/gr.4086505

Gibbs, A. J., Nguyen, H. D., & Ohshima, K. (2015). The 'emergence' of turnip mosaic virus was probably a 'gene-for-quasi-gene' event. *Current Opinion in Virology*, *10*, 20–26. https://doi.org/10.1016/j.coviro.2014.12.004

Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Giritch, A., Marillonnet, S., Engler, C., van Eldik, G., Botterman, J., Klimyuk, V., & Gleba, Y. (2006). Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(40), 14701–14706. https://doi.org/10.1073/pnas.0606631103

González, J. M., Pénzes, Z., Almazán, F., Calvo, E., & Enjuanes, L. (2002). Stabilization of a Full-Length Infectious cDNA Clone of Transmissible Gastroenteritis Coronavirus by Insertion of an Intron. *Journal of Virology*, *76*(9), 4655–4661. https://doi.org/10.1128/jvi.76.9.4655-4661.2002

Gritsun, T. S., & Gould, E. A. (1998). Development and analysis of a tick-borne encephalitis virus infectious clone using a novel and rapid strategy. *Journal of Virological Methods*, *76*(1–2), 109–120. https://doi.org/10.1016/S0166-0934(98)00130-X

Gronenborn, B., Gardner, R. C., Schaefer, S., & Shepherd, R. J. (1981). Propagation of foreign DNA in plants using cauliflower mosaic virus as vector. *Nature*, *294*(5843), 773–776. https://doi.org/10.1038/294773a0

Guo, Y., German, T. L., & Schultz, R. D. (2007). A cryptic promoter in potato virus X vector interrupted plasmid construction. *BMC Molecular Biology*, *8*, 1–6. https://doi.org/10.1186/1471-2199-8-17

Hamamoto, H., Sugiyama, Y., Nakagawa, N., Hashida, E., Matsunaga, Y., Takemoto, S., Watanabe, Y., & Okada, Y. (1993). A New Tobacco Mosaic Virus Vector and its Use for the Systemic Production of Angiotensin-I-Converting Enzyme Inhibitor in Transgenic Tobacco and Tomato. *Bio/Technology*, *11*(8), 930–932. https://doi.org/10.1038/nbt0893-930

Hecht, A., Glasgow, J., Jaschke, P. R., Bawazer, L. A., Munson, M. S., Cochran, J. R., Endy, D., & Salit, M. (2017). Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Research*, *45*(7), 3615–3626. https://doi.org/10.1093/nar/gkx070

Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S., & Mullineaux, P. M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Molecular Biology*, *42*(6), 819–832. https://doi.org/10.1023/a:1006496308160

Horwitz, M. S. Z., & Loeb, L. A. (1986). Promoters selected from random DNA sequences. *Proceedings of the National Academy of Sciences*, *83*(19), 7405–7409. https://doi.org/10.1073/pnas.83.19.7405

Huerta, A. M., & Collado-Vides, J. (2003). Sigma70 promoters in Escherichia coli: Specific transcription in dense regions of overlapping promoter-like signals. *Journal of Molecular Biology*, *333*(2), 261–278. https://doi.org/10.1016/j.jmb.2003.07.017

Igarashi, A., Yamagata, K., Sugai, T., Takahashi, Y., Sugawara, E., Tamura, A., Yaegashi, H., Yamagishi, N., Takahashi, T., Isogai, M., Takahashi, H., & Yoshikawa, N. (2009). Apple latent spherical virus vectors for reliable and effective virus-induced gene silencing among a broad range of plants including tobacco, tomato, Arabidopsis thaliana, cucurbits, and legumes. *Virology*, *386*(2), 407–416. https://doi.org/10.1016/j.virol.2009.01.039

Jakab, G., Droz, E., Brigneti, G., Baulcombe, D., & Malnoë, P. (1997). Infectious in vivo and in vitro transcripts from a full-length cDNA clone of PVY-N605, a Swiss necrotic isolate of potato virus Y. *The Journal of General Virology*, *78 (Pt 12*(12), 3141–3145. https://doi.org/10.1099/0022-1317-78-12-3141

Jeong, J.-Y., Yim, H.-S., Ryu, J.-Y., Lee, H. S., Lee, J.-H., Seen, D.-S., & Kang, S. G. (2012). One-Step Sequence- and Ligation-Independent Cloning as a Rapid and Versatile Cloning Method for Functional Genomics Studies. *Applied and Environmental Microbiology*, *78*(15), 5440–5443. https://doi.org/10.1128/AEM.00844-12

Jin, L., Song, H., Tropea, J. E., Needle, D., Waugh, D. S., Gu, S., & Xinhua, X. (2019). The molecular mechanism of dsRNA processing by a bacterial Dicer. *Nucleic Acids Research*, *47*(9), 4707–4720. https://doi.org/10.1093/nar/gkz208

Johansen, E. (1996). Intron insertion facilitates amplification of cloned virus cDNA in Escherichia coli while biological activity is reestablished after transcription in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(22), 12400–12405. https://doi.org/10.1073/pnas.93.22.12400

Johansen, I. E., & Lund, O. S. (2008). Insertion of introns: A strategy to facilitate assembly of infectious full length clones. *Methods in Molecular Biology*, *451*, 535–544. https://doi.org/10.1007/978-1-59745-102-4-36

Kang, M., Seo, J., Choi, H., Choi, H., & Kim, K. (2016). *Establishment of a Simple and Rapid Gene Delivery System for Cucurbits by Using Engineered of Zucchini yellow mosaic virus*. *32*(1), 70–76.

Kanzaki, H., Kagemori, T., Yamachika, Y., Nitoda, T., & Kawazu, K. (1999). Inhibition of Plant Transformation by Phytolaccoside B from Phytolacca americana Callus. *Bioscience, Biotechnology, and Biochemistry*, *63*(9), 1657–1659. https://doi.org/10.1271/bbb.63.1657

Karasev, A. V., Foulke, S., Wellens, C., Rich, A., Shon, K. J., Zwierzynski, I., Hone, D., Koprowski, H., & Reitz, M. (2005). Plant based HIV-1 vaccine candidate: Tat protein produced in spinach. *Vaccine*, *23*(15 SPEC. ISS.), 1875–1880. https://doi.org/10.1016/j.vaccine.2004.11.021

Kasschau, K. D., Xie, Z., Allen, E., Llave, C., Chapman, E. J., Krizan, K. A., & Carrington, J. C. (2003). P1/HC-Pro, a Viral Suppressor of RNA Silencing, Interferes with Arabidopsis Development and miRNA Function. *Developmental Cell*, *4*(2), 205–217. https://doi.org/10.1016/S1534-5807(03)00025-X

Kelloniemi, J., Mäkinen, K., & Valkonen, J. P. T. (2008). Three heterologous proteins simultaneously expressed from a chimeric potyvirus: Infectivity, stability and the correlation of genome and virion lengths. *Virus Research*, *135*(2), 282–291. https://doi.org/10.1016/j.virusres.2008.04.006

Kim, K. S., & Fulton, J. P. (1969). Electron microscopy of pokeweed leaf cells infected with pokeweed mosaic virus. *Virology*, *37*(3), 297–308. https://doi.org/10.1016/0042-6822(69)90213-X

Kim, K. S., Oh, H. Y., Suranto, S., Nurhayati, E., Gough, K. H., Shukla, D. D., & Pallaghy, C. K. (2003). Infectivity of in vitro transcripts of Johnsongrass mosaic potyvirus full-length cDNA clones in maize and sorghum. *Archives of Virology*, *148*(3), 563–574. https://doi.org/10.1007/s00705-002-0942-z

Klenov, A., & Hudak, K. A. (2018). Complete coding sequence and infectious clone of pokeweed mosaic virus Arkansas isolate. *European Journal of Plant Pathology*, *152*(2), 541–547. https://doi.org/10.1007/s10658-018-1477-9

Klucar, L., Stano, M., & Hajduk, M. (2009). PhiSITE: Database of gene regulation in bacteriophages. *Nucleic Acids Research*, *38*(SUPPL.1), 366–370. https://doi.org/10.1093/nar/gkp911

Kobayashi, A., Hagihara, K., Kajiyama, S., Kanzaki, H., & Kawazu, K. (1995). Antifungal Compounds Induced in the Dual Culture with Phytolacca americana Callus and Botrytis fabae. *Zeitschrift Für Naturforschung C*, *50*(5–6), 398–402. https://doi.org/10.1515/znc-1995-5-610

Koonin, E. V., & Makarova, K. S. (2013). CRISPR-Cas: Evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biology*, *10*(5), 679–686. https://doi.org/10.4161/rna.24022

Kumagai, M. H., Donson, J., Della-Cioppa, G., Harvey, D., Hanley, K., & Grill, L. K. (1995). Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proceedings of the National Academy of Sciences*, *92*(5), 1679–1683. https://doi.org/10.1073/pnas.92.5.1679

Lai, C. J., Zhao, B., Hori, H., & Bray, M. (1991). Infectious RNA transcribed from stably cloned full-length cDNA of dengue type 4 virus. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(12), 5139–5143. https://doi.org/10.1073/pnas.88.12.5139

Langeveld, J. P. ., Brennan, F. R., Martínez-Torrecuadrada, J. L., Jones, T. D., Boshuizen, R. S., Vela, C., Casal, J. I., Kamstrup, S., Dalsgaard, K., Meloen, R. H., Bendig, M. M., & Hamilton, W. D. . (2001). Inactivated recombinant plant virus protects dogs from a lethal challenge with canine parvovirus. *Vaccine*, *19*(27), 3661–3670. https://doi.org/10.1016/S0264-410X(01)00083-4

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Legocki, A. B., Miedzinska, K., Czaplińska, M., Płucieniczak, A., & Wedrychowicz, H. (2005). Immunoprotective properties of transgenic plants expressing E2 glycoprotein from CSFV and cysteine protease from Fasciola hepatica. *Vaccine*, *23*(15), 1844–1846. https://doi.org/10.1016/j.vaccine.2004.11.015

Li, D., Aaskov, J., & Lott, W. B. (2011). Identification of a Cryptic Prokaryotic Promoter within the cDNA Encoding the 5′ End of Dengue Virus RNA Genome. *PLoS ONE*, *6*(3), e18197. https://doi.org/10.1371/journal.pone.0018197

Lin, T., Chen, Z., Usha, R., Stauffacher, C. V, Dai, J. B., Schmidt, T., & Johnson, J. E. (1999). The refined crystal structure of cowpea mosaic virus at 2.8 A resolution. *Virology*, *265*(1), 20–34. https://doi.org/10.1006/viro.1999.0038

Lindbo, J. A. (2007a). High-efficiency protein expression in plants from agroinfection-compatible Tobacco mosaic virus expression vectors. *BMC Biotechnology*, *7*(1), 52. https://doi.org/10.1186/1472-6750-7-52

Lindbo, J. A. (2007b). TRBO: A High-Efficiency Tobacco Mosaic Virus RNA-Based Overexpression Vector. *Plant Physiology*, *145*(4), 1232–1240. https://doi.org/10.1104/pp.107.106377

Lisser, S., & Margalit, H. (1993). Compilation of E.coli mRNA promoter sequences. *Nucleic Acids Research*, *21*(7), 1507–1516. https://doi.org/10.1093/nar/21.7.1507

Liu, L., Cañizares, M. C., Monger, W., Perrin, Y., Tsakiris, E., Porta, C., Shariat, N., Nicholson, L., & Lomonossoff, G. P. (2005). Cowpea mosaic virus-based systems for the production of antigens and antibodies in plants. *Vaccine*, *23*(15), 1788–1792. https://doi.org/10.1016/j.vaccine.2004.11.006

Liu, L., & Lomonossoff, G. P. (2002). Agroinfection as a rapid method for propagating Cowpea mosaic virus-based constructs. *Journal of Virological Methods*, *105*(2), 343–348. https://doi.org/10.1016/S0166-0934(02)00121-0

Lodge, J. K., Kaniewski, W. K., & Tumer, N. E. (1993). Broad-spectrum virus resistance in transgenic plants expressing pokeweed antiviral protein. *Proceedings of the National Academy of Sciences*, *90*(15), 7089–7093. https://doi.org/10.1073/pnas.90.15.7089

López-Moya, J. J., & García, J. A. (2000). Construction of a stable and highly infectious intron-containing cDNA clone of plum pox potyvirus and its use to infect plants by particle bombardment. *Virus Research*, *68*, 99–107. https://doi.org/10.1016/S0168-1702(00)00161-1

Mansouri, S., Nourollahzadeh, E., & Hudak, K. A. (2006). Pokeweed antiviral protein depurinates the sarcin/ricin loop of the rRNA prior to binding of aminoacyl-tRNA to the ribosomal A-site. *RNA (New York, N.Y.)*, *12*(9), 1683–1692. https://doi.org/10.1261/rna.70306

McBride, M. B., & Zhou, Y. (2019). Cadmium and zinc bioaccumulation by Phytolacca americana from hydroponic media and contaminated soils. *International Journal of Phytoremediation*, *21*(12), 1215–1224. https://doi.org/10.1080/15226514.2019.1612849

Merits, A., Guo, D., & Saarma, M. (1998). VPg, coat protein and five non-structural proteins of potato A potyvirus bind RNA in a sequence-unspecific manner. *Journal of General Virology*, *79*(12), 3123–3127. https://doi.org/10.1099/0022-1317-79-12-3123

Michon, T., Estevez, Y., Walter, J., German-Retana, S., & Gall, O. (2006). The potyviral virus genome-linked protein VPg forms a ternary complex with the eukaryotic initiation factors eIF4E and eIF4G and reduces eIF4E affinity for a mRNA cap analogue. *FEBS Journal*, *273*(6), 1312–1322. https://doi.org/10.1111/j.1742-4658.2006.05156.x

Münster, M., Płaszczyca, A., Cortese, M., Neufeldt, C. J., Goellner, S., Long, G., & Bartenschlager, R. (2018). A reverse genetics system for Zika virus based on a simple molecular cloning strategy. *Viruses*, *10*(7), 1–17. https://doi.org/10.3390/v10070368

Neller, K. C. M., Klenov, A., & Hudak, K. A. (2016). The Pokeweed Leaf mRNA Transcriptome and Its Regulation by Jasmonic Acid. *Frontiers in Plant Science*, *7*(March), 1–13. https://doi.org/10.3389/fpls.2016.00283

Nicaise, V. (2014). Crop immunity against viruses: Outcomes and future challenges. *Frontiers in Plant Science*, *5*(NOV), 1–18. https://doi.org/10.3389/fpls.2014.00660

Ning, L. I., Yang, W., Fang, S., Xinhai, L. I., Liu, Z., Leng, X., & Shuqing, A. N. (2017). Dispersal of invasive Phytolacca americana seeds by birds in an urban garden in China. *Integrative Zoology*, *12*(1), 26–31. https://doi.org/10.1111/1749-4877.12214

Olsen, B. S., & Johansen, I. E. (2001). Nucleotide sequence and infectious cDNA clone of the L1 isolate of Pea seed-borne mosaic potyvirus. *Archives of Virology*, *146*(1), 15–25. https://doi.org/10.1007/s007050170187

Ouillet, L., Guilley, H., Jonard, G., & Richards, K. (1989). In vitro synthesis of biologically active beet necrotic yellow vein virus RNA. *Virology*, *172*(1), 293–301. https://doi.org/10.1016/0042-6822(89)90131-1

Pasin, F., Bedoya, L. C., Bernabé-Orts, J. M., Gallo, A., Simón-Mateo, C., Orzaez, D., & García, J. A. (2017). Multiple T-DNA Delivery to Plants Using Novel Mini Binary Vectors with Compatible Replication Origins. *ACS Synthetic Biology*, *6*(10), 1962–1968. https://doi.org/10.1021/acssynbio.6b00354

Pasin, F., Menzel, W., & Daròs, J. A. (2019). Harnessed viruses in the age of metagenomics and synthetic biology: an update on infectious clone assembly and biotechnologies of plant viruses. *Plant Biotechnology Journal*, *17*(6), 1010–1026. https://doi.org/10.1111/pbi.13084

Pasin, F., Simón-Mateo, C., & García, J. A. (2014). The Hypervariable Amino-Terminus of P1 Protease Modulates Potyviral Replication and Host Defense Responses. *PLoS Pathogens*, *10*(3), e1003985. https://doi.org/10.1371/journal.ppat.1003985

Peng, Y. H., Kadoury, D., Gal-On, A., Huet, H., Wang, Y., & Raccah, B. (1998). Mutations in the HC-Pro gene of zucchini yellow mosaic potyvirus: Effects on aphid transmission and binding to purified virions. *Journal of General Virology*, *79*(4), 897–904. https://doi.org/10.1099/0022-1317-79-4-897

Pérez Filgueira, D. M., Zamorano, P. I., Domínguez, M. G., Taboga, O., Del Médico Zajac, M. P., Puntel, M., Romera, S. A., Morris, T. J., Borca, M. V, & Sadir, A. M. (2003). Bovine herpes virus gD protein produced in plants using a recombinant tobacco mosaic virus (TMV) vector possesses authentic antigenicity. *Vaccine*, *21*(27–30), 4201–4209. https://doi.org/10.1016/S0264-410X(03)00495-X

Peyret, H. (2015). A protocol for the gentle purification of virus-like particles produced in plants. *Journal of Virological Methods*, *225*, 59–63. https://doi.org/10.1016/j.jviromet.2015.09.005

Peyret, H., & Lomonossoff, G. P. (2015). When plant virology met Agrobacterium : the rise of the deconstructed clones. *Plant Biotechnology Journal*, *13*(8), 1121–1135. https://doi.org/10.1111/pbi.12412

Pluchino, K. M., Esposito, D., Moen, J. K., Hall, M. D., Madigan, J. P., Shukla, S., Procter, L. V., Wall, V. E., Schneider, T. D., Pringle, I., Ambudkar, S. V., Gill, D. R., Hyde, S. C., & Gottesman, M. M. (2015). Identification of a cryptic bacterial promoter in mouse (mdr1a) p-glycoprotein cDNA. *PLoS ONE*, *10*(8), 1–17. https://doi.org/10.1371/journal.pone.0136396

Porta, C., Spall, V. E., Findlay, K. C., Gergerich, R. C., Farrance, C. E., & Lomonossoff, G. P. (2003). Cowpea mosaic virus-based chimaeras: Effects of inserted peptides on the phenotype, host range, and transmissibility of the modified viruses. *Virology*, *310*(1), 50–63. https://doi.org/10.1016/S0042-6822(03)00140-5

Pu, S.-Y., Wu, R.-H., Yang, C.-C., Jao, T.-M., Tsai, M.-H., Wang, J.-C., Lin, H.-M., Chao, Y.-S., & Yueh, A. (2011). Successful Propagation of Flavivirus Infectious cDNAs by a Novel Method To Reduce the Cryptic Bacterial Promoter Activity of Virus Genomes. *Journal of Virology*, *85*(6), 2927–2941. https://doi.org/10.1128/JVI.01986-10

Rajamäki, M. L., & Valkonen, J. P. T. (2002). Viral genome-linked protein (VPg) controls accumulation and phloem loading of a Potyvirus in inoculated potato leaves. *Molecular Plant-Microbe Interactions*, *15*(2), 138–149. https://doi.org/10.1094/MPMI.2002.15.2.138

Rajamohan, F., Venkatachalam, T. K., Irvin, J. D., & Uckun, F. M. (1999). Pokeweed Antiviral Protein Isoforms PAP-I, PAP-II, and PAP-III Depurinate RNA of Human Immunodeficiency Virus (HIV)-1. *Biochemical and Biophysical Research Communications*, *260*(2), 453–458. https://doi.org/10.1006/bbrc.1999.0922

Ratcliff, F., Martin-Hernandez, A. M., & Baulcombe, D. C. (2008). Technical Advance: Tobacco rattle virus as a vector for analysis of gene function by silencing. *The Plant Journal*, *25*(2), 237–245. https://doi.org/10.1046/j.0960-7412.2000.00942.x

Reddy, D. V. R., Sudarshana, M. R., Fuchs, M., Rao, N. C., & Thottappilly, G. (2009). Genetically Engineered Virus-Resistant Plants in Developing Countries: Current Status and Future Prospects. In G. Loebenstein & J. P. B. T.-A. in V. R. Carr (Eds.), *Natural and Engineered Resistance to Plant Viruses, Part I* (Vol. 75, pp. 185–220). Academic Press. https://doi.org/10.1016/S0065-3527(09)07506-X

Ricardo, P. C., Françoso, E., & Arias, M. C. (2020). Fidelity of DNA polymerases in the detection of intraindividual variation of mitochondrial DNA. *Mitochondrial DNA Part B: Resources*, *5*(1), 108–112. https://doi.org/10.1080/23802359.2019.1697188

Röder, J., Dickmeis, C., Fischer, R., & Commandeur, U. (2018). Systemic Infection of Nicotiana benthamiana with Potato virus X Nanoparticles Presenting a Fluorescent iLOV Polypeptide Fused Directly to the Coat Protein. *BioMed Research International*, *2018*, 1–12. https://doi.org/10.1155/2018/9328671

Ryabov, E. V, Robinson, D. J., & Taliansky, M. E. (1999). A plant virus-encoded protein facilitates long-distance movement of heterologous viral RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(4), 1212–1217. https://doi.org/10.1073/pnas.96.4.1212

Rybicki, E. P., & Martin, D. P. (2011). Virus-Derived ssDNA Vectors for the Expression of Foreign Proteins in Plants. In K. Palmer & Y. Gleba (Eds.), *Plant Viral Vectors* (pp. 19–45). Springer Berlin Heidelberg. https://doi.org/10.1007/82_2011_185

Sah, R., Rodriguez-Morales, A. J., Jha, R., Chu, D. K. W., Gu, H., Peiris, M., Bastola, A., Lal, B. K., Ojha, H. C., Rabaan, A. A., Zambrano, L. I., Costello, A., Morita, K., Pandey, B. D., & Poon, L. L. M. (2020). Complete Genome Sequence of a 2019 Novel Coronavirus (SARS-CoV-2) Strain Isolated in Nepal. *Microbiology Resource Announcements*, *9*(11), 19–21. https://doi.org/10.1128/MRA.00169-20

Saida, F., Uzan, M., Odaert, B., & Bontems, F. (2006). Expression of Highly Toxic Genes in E. coli: Special Strategies and Genetic Tools. *Current Protein and Peptide Science*, *7*(1), 47–56. https://doi.org/10.2174/138920306775474095

Satoh, N., Kon, T., Yamagishi, N., Takahashi, T., Natsuaki, T., & Yoshikawa, N. (2014). Apple latent spherical virus vector as vaccine for the prevention and treatment of mosaic diseases in pea, broad bean, and eustoma plants by bean yellow mosaic virus. *Viruses*, *6*(11), 4242–4257. https://doi.org/10.3390/v6114242

Scholthof, H. B. (1999). Rapid Delivery of Foreign Genes into Plants by Direct Rub-Inoculation with Intact Plasmid DNA of a Tomato Bushy Stunt Virus Gene Vector. *Journal of Virology*, *73*(9), 7823–7829. https://doi.org/10.1128/jvi.73.9.7823-7829.1999

Seo, J. K., Lee, H. G., & Kim, K. H. (2009). Systemic gene delivery into soybean by simple rub-inoculation with plasmid DNA of a Soybean mosaic virus-based vector. *Archives of Virology*, *154*(1), 87–99. https://doi.org/10.1007/s00705-008-0286-4

Shiboleth, Y. M., Haronsky, E., Leibman, D., Arazi, T., Wassenegger, M., Whitham, S. A., Gaba, V., & Gal-On, A. (2007). The Conserved FRNK Box in HC-Pro, a Plant Viral Suppressor of Gene Silencing, Is Required for Small RNA Binding and Mediates Symptom Development. *Journal of Virology*, *81*(23), 13135–13148. https://doi.org/10.1128/JVI.01031-07

Shukla, D. D., & Ward, C. W. (1989). Structure of Potyvirus Coat Proteins and its Application in the Taxonomy of the Potyvirus Group. In K. Maramorosch, F. A. Murphy, & A. J. B. T.-A. in V. R. Shatkin (Eds.), *Advances in Virus Research* (Vol. 36, pp. 273–314). Academic Press. https://doi.org/10.1016/S0065-3527(08)60588-6

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(539). https://doi.org/10.1038/msb.2011.75

Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A. M., & Fire, A. Z. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science*, *351*(6276), aad4234–aad4234. https://doi.org/10.1126/science.aad4234

Skorski, P., Leroy, P., Fayet, O., Dreyfus, M., & Hermann-Le Denmat, S. (2006). The Highly Efficient Translation Initiation Region from the Escherichia coli rpsA Gene Lacks a Shine-Dalgarno Element. *Journal of Bacteriology*, *188*(17), 6277–6285. https://doi.org/10.1128/JB.00591-06

Sripriya, R., Sangeetha, M., Parameswari, C., Veluthambi, B., & Veluthambi, K. (2011). Improved Agrobacterium-mediated co-transformation and selectable marker elimination in transgenic rice by using a high copy number pBin19-derived binary vector. *Plant Science*, *180*(6), 766–774. https://doi.org/10.1016/j.plantsci.2011.02.010

Stead, M. B., Agrawal, A., Bowden, K. E., Nasir, R., Mohanty, B. K., Meagher, R. B., & Kushner, S. R. (2012). RNAsnap^{TM}: A rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Research*, *40*(20), 1–9. https://doi.org/10.1093/nar/gks680

Stirpe, F. (2013). Ribosome-inactivating proteins: From toxins to useful proteins. *Toxicon*, *67*, 12–16. https://doi.org/10.1016/j.toxicon.2013.02.005

Sun, K., Zhao, D., Liu, Y., Huang, C., Zhang, W., & Li, Z. (2017). Rapid construction of complex plant RNA virus infectious cDNA clones for agroinfection using a yeast-E. Coli-agrobacterium shuttle vector. *Viruses*, *9*(11). https://doi.org/10.3390/v9110332

Szakasits, D., Siddique, S., & Bohlmann, H. (2007). An Improved pPZP Vector for Agrobacterium-mediated Plant Transformation. *Plant Molecular Biology Reporter*, *25*(3–4), 115–120. https://doi.org/10.1007/s11105-007-0013-4

Tacahashi, Y., & Uyeda, I. (1999). Restoration of the 3’ end of potyvirus RNA derived from Poly(A)-deficient infectious cDNA clones. *Virology*, *265*(1), 147–152. https://doi.org/10.1006/viro.1999.0027

Takamatsu, N., Ishikawa, M., Meshi, T., & Okada, Y. (1987). Expression of bacterial chloramphenicol acetyltransferase gene in tobacco plants mediated by TMV-RNA. *The EMBO Journal*, *6*(2), 307–311. https://doi.org/10.1002/j.1460-2075.1987.tb04755.x

Taylor, K. M., Lin, T., Porta, C., Mosser, A. G., Giesing, H. A., Lomonossoff, G. P., & Johnson, J. E. (2000). Influence of three-dimensional structure on the immunogenicity of a peptide expressed on the surface of a plant virus. *Journal of Molecular Recognition*, *13*(2), 71–82. https://doi.org/10.1002/(SICI)1099-1352(200003/04)13:2<71::AID-JMR489>3.0.CO;2-V

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. https://doi.org/10.1093/bib/bbs017

Touchon, M., Charpentier, S., Clermont, O., Rocha, E. P. C., Denamur, E., & Branger, C. (2011). CRISPR distribution within the Escherichia coli species is not suggestive of immunity-associated diversifying selection. *Journal of Bacteriology*, *193*(10), 2460–2467. https://doi.org/10.1128/JB.01307-10

Touriño, A., Sánchez, F., Fereres, A., & Ponz, F. (2008). High expression of foreign proteins from a biosafe viral vector derived from Turnip mosaic virus. *Spanish Journal of Agricultural Research*, *6*(S1), 48. https://doi.org/10.5424/sjar/200806S1-373

Tuo, D., Fu, L., Shen, W., Li, X., Zhou, P., & Yan, P. (2017). Generation of stable infectious clones of plant viruses by using Rhizobium radiobacter for both cloning and inoculation. *Virology*, *510*(May), 99–103. https://doi.org/10.1016/j.virol.2017.07.012

Tuo, D., Shen, W., Yan, P., Li, X., & Zhou, P. (2015). Rapid Construction of Stable Infectious Full-Length cDNA Clone of Papaya Leaf Distortion Mosaic Virus Using In-Fusion Cloning. *Viruses*, *7*(12), 6241–6250. https://doi.org/10.3390/v7122935

Tuttle, J. R., Haigler, C. H., & Robertson, D. (2012). Method: low-cost delivery of the cotton leaf crumple virus-induced gene silencing system. *Plant Methods*, *8*(1), 1. https://doi.org/10.1186/1746-4811-8-27

Verchot, J., & Carrington, J. C. (1995). Evidence that the potyvirus P1 proteinase functions in trans as an accessory factor for genome amplification. *Journal of Virology*, *69*(6), 3668–3674.

Wan, J., Fellers, J., Hunt, A. G., Collins, G. B., & Hong, Y. (1998). In vitro interactions between a potyvirus-encoded, genome-linked protein and RNA-dependent RNA polymerase. *Journal of General Virology*, *79*(8), 2043–2049. https://doi.org/10.1099/0022-1317-79-8-2043

Wang, P., Zoubenko, O., & Tumer, N. E. (1998). Reduced toxicity and broad spectrum resistance to viral and fungal infection in transgenic plants expressing pokeweed antiviral protein II. *Plant Molecular Biology*, *38*(6), 957–964. https://doi.org/10.1023/a:1006084925016

Wessel, M., Klüsener, S., Gödeke, J., Fritz, C., Hacker, S., & Narberhaus, F. (2006). Virulence of Agrobacterium tumefaciens requires phosphatidylcholine in the bacterial membrane. *Molecular Microbiology*, *62*(3), 906–915. https://doi.org/10.1111/j.1365-2958.2006.05425.x

Wulf, M. G., Maguire, S., Humbert, P., Dai, N., Bei, Y., Nichols, N. M., Corrêa, I. R., & Guan, S. (2019). Non-templated addition and template switching by Moloney murine leukemia virus (MMLV)-based reverse transcriptases co-occur and compete with each other. *Journal of Biological Chemistry*, *294*(48), 18220–18231. https://doi.org/10.1074/jbc.RA119.010676

Wylie, S. J., Adams, M., Chalam, C., Kreuze, J., López-Moya, J. J., Ohshima, K., Praveen, S., Rabenstein, F., Stenger, D., Wang, A., & Zerbini, F. M. (2017). ICTV virus taxonomy profile: Potyviridae. *Journal of General Virology*, *98*(3), 352–354. https://doi.org/10.1099/jgv.0.000740

Xia, Y., Li, K., Li, J., Wang, T., Gu, L., & Xun, L. (2019). T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. *Nucleic Acids Research*, *47*(3), 1–11. https://doi.org/10.1093/nar/gky1169

Xu, D., Wang, M., Kinard, G., & Li, R. (2012). Complete genome sequence of two isolates of pokeweed mosaic virus and its relationship to other members of the genus Potyvirus. *Archives of Virology*, *157*(10), 2023–2026. https://doi.org/10.1007/s00705-012-1329-4

Yamagishi, N., Kishigami, R., & Yoshikawa, N. (2014). Reduced generation time of apple seedlings to within a year by means of a plant virus vector: A new plant-breeding technique with no transmission of genetic modification to the next generation. *Plant Biotechnology Journal*, *12*(1), 60–68. https://doi.org/10.1111/pbi.12116

Yamshchikov, V., Mishin, V., & Cominelli, F. (2001). A new strategy in design of +RNA virus infectious clones enabling their stable propagation in E. coli. *Virology*, *281*, 272–280. https://doi.org/10.1006/viro.2000.0793

Yang, S. J., Revers, F., Souche, S., Lot, H., Le Gall, O., Candresse, T., & Dunez, J. (1998). Construction of full-length cDNA clones of lettuce mosaic virus (LMV) and the effects of intron-insertion on their viability in Escherichia coli and on their infectivity to plants. *Archives of Virology*, *143*, 2443–2451. https://doi.org/10.1007/s007050050474

Yu, H.-H., & Wong, S.-M. (1998). Synthesis of biologically active cDNA clones of cymbidium mosaic potexvirus using a population cloning strategy. *Archives of Virology*, *143*(8), 1617–1620. https://doi.org/10.1007/s007050050402

Yu, N., Cai, Z., Xiong, Z., Yang, Y., & Liu, Z.-X. (2020). Complete genomic sequence of Noni mosaic virus (NoMV) associated with a mosaic disease in Morinda citrifolia L. *European Journal of Plant Pathology*, *156*(4), 1005–1014. https://doi.org/10.1007/s10658-020-01948-4

Zhang, J., Yu, D., Zhang, Y., Liu, K., Xu, K., Zhang, F., Wang, J., Tan, G., Nie, X., Ji, Q., Zhao, L., & Li, C. (2017). Vacuum and co-cultivation agroinfiltration of (germinated) seeds results in tobacco rattle virus (TRV) mediated whole-plant virus-induced gene silencing (VIGS) in wheat and maize. *Frontiers in Plant Science*, *8*(March), 1–12. https://doi.org/10.3389/fpls.2017.00393

Zhang, X., Yuan, Y.-R., Pei, Y., Lin, S.-S., Tuschl, T., Patel, D. J., & Chua, N.-H. (2006). Cucumber mosaic virus-encoded 2b suppressor inhibits Arabidopsis Argonaute1 cleavage activity to counter plant defense. *Genes & Development*, *20*(23), 3255–3268. https://doi.org/10.1101/gad.1495506

Zhao, L., Sun, Y.-L., Cui, S.-X., Chen, M., Yang, H.-M., Liu, H.-M., Chai, T.-Y., & Huang, F. (2011). Cd-induced changes in leaf proteome of the hyperaccumulator plant Phytolacca americana. *Chemosphere*, *85*(1), 56–66. https://doi.org/10.1016/j.chemosphere.2011.06.029

Zilian, E., & Maiss, E. (2011). Detection of plum pox potyviral protein-protein interactions in planta using an optimized mRFP-based bimolecular fluorescence complementation system. *Journal of General Virology*, *92*(12), 2711–2723. https://doi.org/10.1099/vir.0.033811-0

Zou, N., Ditty, S., Li, B., & Lo, S. C. (2003). Random priming PCR strategy to amplify and clone trace amounts of DNA. *BioTechniques*, *35*(4), 758–765. https://doi.org/10.2144/03354st06

Zoubenko, O., Hudak, K., & Tumer, N. E. (2000). A non-toxic pokeweed antiviral protein mutant inhibits pathogen infection via a novel salicylic acid-independent pathway. *Plant Molecular Biology*, *44*(2), 219–229. https://doi.org/10.1023/a:1006443626864

Chapter 5:

Appendix

5.1 – A small RNA targets pokeweed antiviral protein transcript

Published manuscript

Klenov, A., Neller, K. C. M., Burns, L. A., Krivdova, G., & Hudak, K. A. (2016). *Physiologia Plantarum*, 156(3), 241–251. https://doi.org/10.1111/ppl.12393

Contributions: LB and KH conceived the study. AK and GK performed the wet-lab experiments. KN performed bioinformatics analysis and drafted the manuscript. KH edited the manuscript.

Physiologia Plantarum
An International Journal for Plant Biology

Physiologia Plantarum 156: 241–251. 2016

A small RNA targets pokeweed antiviral protein transcript

Alexander Klenov, Kira C. M. Neller, Lydia A. Burns, Gabriela Krivdova and Katalin A. Hudak*

Department of Biology, York University, Toronto, Ontario, Canada

Correspondence
*Corresponding author,
e-mail: hudak@yorku.ca

Received 29 July 2015;
revised 9 September 2015

doi:10.1111/ppl.12393

Ribosome-inactivating proteins (RIPs) are a class of plant defense proteins with N-glycosidase activity (EC 3.2.2.22). Pokeweed antiviral protein (PAP) is a Type I RIP isolated from the pokeweed plant, *Phytolacca americana*, thought to confer broad-spectrum virus resistance in this plant. Through a combination of standard molecular techniques and RNA sequencing analysis, we report here that a small RNA binds and cleaves the open reading frame of PAP mRNA. Additionally, sRNA targeting of PAP is dependent on jasmonic acid (JA), a plant hormone important for defense against pathogen infection and herbivory. Levels of small RNA increased with JA treatment, as did levels of PAP mRNA and protein, suggesting that the small RNA functions to moderate the expression of PAP in response to this hormone. The association between JA and PAP expression, mediated by sRNA299, situates PAP within a signaling pathway initiated by biotic stress. The consensus sequence of sRNA299 was obtained through bioinformatic analysis of pokeweed small RNA sequencing. To our knowledge, this is the first account of a sRNA targeting a RIP gene.

Introduction

Plants have evolved numerous mechanisms to inhibit pathogen infection. Ribosome-inactivating proteins (RIPs) are a class of plant defense proteins with *N*-glycosidase activity. Specifically, RIPs catalyze the removal of purines from nucleic acid targets. Their first identified role was depurination of the conserved sarcin/ricin loop of large ribosomal RNA and a resultant inhibition of protein synthesis in vitro (Gessner and Irvin 1980, Endo et al. 1988). More recently, depurination of RNA-based viruses has been shown (Rajamohan et al. 1999, Karran and Hudak 2008, Kaur et al. 2011). Pokeweed antiviral protein (PAP), a Type I RIP isolated from the pokeweed plant *Phytolacca americana* (Irvin 1975) inhibits plant viruses including potato virus X and Y, tobacco mosaic virus and brome mosaic virus (Lodge et al. 1993, Karran and Hudak 2008). Transgenic plants expressing PAP exhibit virus resistance; however, these plants also display phenotypes indicative of toxicity, correlated with the level of PAP synthesis (Lodge et al. 1993). Unregulated toxicity of PAP is not observed in the native plant, pokeweed, suggesting the absence of important regulatory mechanisms in heterologous systems.

Little is known about the regulation of PAP expression in pokeweed, with the exception of preliminary temporal and spatial characterization (Honjo et al. 2002, Irvin et al. 1980, Kataoka et al. 1992). RIP induction by pathogen stress-related compounds, mechanical wounding, plant viruses and fungal pathogens has been reported (Dunaeva et al. 1999, Song et al. 2000, Qin et al. 2005, Iglesias et al. 2008). Temporal regulation of some RIPs has also been shown (Parente et al. 2008, Kawade and Masuda 2009, Loss-Morais et al. 2013). Although there has been some insight into transcriptional regulation of RIPs, their post-transcriptional regulation is not well characterized. Generally, positive correlation exists between transcript and protein levels (Kawade et al. 2008); however, discrepancies have been observed in some species

Abbreviations – EMSA, electrophoretic mobility shift assay; JA, jasmonic acid; LMW, low molecular weight; miRNA, micro RNA; nt, nucleotide; PAP, pokeweed antiviral protein; RACE, rapid extension of cDNA ends; RIP, ribosome-inactivating protein; siRNA, small interfering RNA; sRNA, small RNA.

(Iglesias et al. 2008,Vepachedu et al. 2003). Therefore, post-transcriptional regulation may be important for maintaining normal RIP expression patterns.

Small RNA (sRNA)-mediated regulation is a form of post-transcriptional control that has not been described for RIPs. sRNAs in plants are 20–24 nucleotide (nt), non-coding RNAs involved in fine-tuning gene expression. Depending on their biogenesis, sRNAs are either classified as microRNAs (miRNAs) or small interfering RNAs (siRNAs). Both types of sRNA participate in gene targeting mediated by the ARGONAUTE-containing RNA-Induced Silencing Complex (reviewed in Bologna and Voinnet 2014). In plants, sRNAs are induced in response to various biotic stresses, including pathogen infection, herbivory and related plant stress hormones (reviewed in Khraiwesh et al. 2012).

Based on the abundance of sRNAs involved in stress responses, and the fact that PAP is an antiviral RIP, we hypothesized that control of PAP expression involved sRNA-mediated gene regulation. Here, we identify a sRNA that causes cleavage of PAP mRNA, within its open reading frame. We also show that jasmonic acid (JA), a plant hormone important during herbivory and pathogen stress, differentially affects the expression of PAP mRNA and the sRNA. To our knowledge, this is the first report of a sRNA targeting a RIP message and provides novel insight into the post-transcriptional control of these plant defense enzymes.

Materials and methods

Pokeweed growth conditions and treatment with JA

Pokeweed seeds were immersed in 37% sulfuric acid for 5 min, rinsed in water and imbibed in water for 4 days at room temperature. Seeds were germinated in soil (Promix BX) and raised in a growth chamber (AC60, Biochambers, MB, CA) under fluorescent and incandescent lights at 180 $\mu mol\, m^{-2}\, s^{-1}$. Plants were fertilized once every 2 weeks with N:P:K 20:20:20. Plants at the 4-leaf stage were sprayed with 5 ml of 5 m*M* JA dissolved in 0.5% ethanol (to improve the solubility of JA). Negative control plants were sprayed with 0.5% ethanol alone. Plants were returned to the chamber and leaf tissue was harvested 24 h after treatment. All analyses for this study were conducted on pokeweed plants at the 4-leaf stage of growth.

Isolation of total and low molecular weight RNA

Total RNA was extracted from pokeweed to probe for the sRNA, and cabbage served as the negative control. Leaf tissue (7 g) was ground to a powder with liquid nitrogen. TRIzol reagent (Molecular Research Center, Inc., Cincinnati, OH, USA) and chloroform were added, samples were vortexed and incubated at room temperature for 10 min. Particulate matter was pelleted by centrifugation at 16 000 *g* for 15 min at 4°C. Isopropanol was added to supernatants and the RNA was precipitated at −20°C. Samples were centrifuged at 16 000 g for 20 min at 4°C, RNA pellets were resuspended in water and extracted in acidic phenol:chloroform:isoamyl alcohol (25:24:1). RNA was precipitated in ethanol and finally resuspended in water and stored at −80°C.

LMW RNA was extracted from pokeweed and cabbage to serve as an enriched pool of sRNA. Following isopropanol precipitation of total RNA, the RNA was diluted to 1 $\mu g\, \mu l^{-1}$ in water and incubated in 5% PEG8000 and 0.5 *M* NaCl for 30 min at −20°C. Samples were centrifuged at 16 000 *g* for 25 min at 4°C to pellet the high molecular weight RNA. Supernatants were incubated in isopropanol at −20°C and LMW RNA was pelleted by centrifugation at 16 000 *g* for 25 min at 4°C, washed in 75% ethanol and resuspended in water.

Northern blotting

Total pokeweed (PW) and cabbage (CB) RNA (38 µg) was separated through 1.5% agarose gel in 7% formamide, transferred to nylon membrane by capillary action and cross-linked to the membrane with UV light (120 $mJ\, cm^{-2}$ for 12 s). The RNA was probed by incubation with an α^{33}P-CTP labeled riboprobe (1×10^6 cpm) specific for a 245 nt portion of the PAP mRNA (Accession # AR009535.1). The internally labeled minus-strand RNA probe was synthesized by in vitro transcription with T7 polymerase from a polymerase chain reaction (PCR) template generated from a portion of the PAP mRNA spanning nt 680–925. The PAP-specific primers were PAP680for: 5′ gggagtaaaatcaagaagtcagg 3′ and PAP T7 925rev: 5′ taatacgactcactataggaaatcttaccccatgtctcttg 3′. An in vitro transcript of full-length PAP mRNA was used as a positive control.

Primer extension

A PAP-specific reverse primer (5′ gaagatcattcggaaaagtgg 3′) complementary to nt 340–361 of PAP mRNA was 5′ end-labeled with γ^{33}P-ATP. Radiolabeled primer (5×10^5 cpm) was denatured with total PW or CB RNA (30 µg) and annealed at room temperature for 15 min. The primer was extended by reverse transcription with 100 units of Superscript III (Life Technologies, Burlington, ON, CA) and incubation at 48°C for 90 min. Reactions were terminated by the addition of formamide buffer and denatured cDNA products were separated

by 7 *M* urea/6% acrylamide gel. To identify the cDNA 3′ ends, a dideoxynucleotide sequencing ladder of the PAP cDNA was generated with the same primer used for primer extension of total RNA.

5′ rapid extension of cDNA ends

Total RNA from pokeweed (36 µg) was used as template to generate PAP cDNA using a PAP-specific reverse primer. Briefly, total RNA was denatured with PAP reverse primer complementary to nt 607–632 (5′ ctaacacgagaattggcatttgggc 3′) and allowed to anneal at room temperature for 15 min. Reverse transcription was carried out at 48°C for 90 min using 100 units of Superscript III (Life Technologies, Burlington, ON, CA). The RNA was subsequently digested with 1.25 U of RNase H (New England Biolabs, Whitby, ON, CA) for 30 min at 37°C, and the cDNA was collected through Biobasic EZ-10 columns. cDNA samples (8.5 µl) were ligated with 10 units of T4 RNA Ligase 1 (New England Biolabs, Whitby, ON, CA) to 2.5 nmol of adapter primer (5′ PO_4- ccatggcaataccggtaaggtcctcactc 3′) in 1X RACE Buffer (20% PEG8000, 50 m*M* Tris–HCl pH 8.0, 10 m*M* $MgCl_2$, 10 ng µl^{-1} BSA, 2 m*M* hexamine cobalt chloride, 20 µ*M* ATP) at 22°C for 4 h. After purification through an EZ-10 column, 5 µl of ligation product was amplified with two rounds of PCR with adapter-specific (AP1: 5′ gagtgaggaccttaccgg 3′, AP2: 5′ ccttaccggtattgccatgg 3′) and PAP-specific primers (PAP532-512R: 5′ caaagggatcagaataaccc 3′, PAP432-411BglIIR: 5′ agatcttggatttgtatttgtattggg 3′). The final PCR product was isolated from low-melt agarose and sequenced.

Electrophoretic mobility shift assay

LMW RNA was probed with an α^{33}P-CTP radiolabeled sense RNA transcript of the PAP sequence from nt 286–309. Complementary cDNA primers (PAP T7 286–302 for: 5′ taatacgactcactatagggctgtgaatacaatc 3′; PAP 309–293 rev: 5′ tgtagatgattgtattc 3′) were annealed, with the forward primer also containing a T7 polymerase-binding site. Subsequent PCR produced sufficient DNA template for in vitro transcription of the riboprobe. LMW RNA (10 µg) of pokeweed and 1.2×10^5 cpm of riboprobe were denatured at 95°C for 5 min in RNA-binding buffer (5 m*M* HEPES-KOH pH 7.8, 100 m*M* KCl, 6 m*M* $MgCl_2$, 3.8% glycerol). LMW RNA from cabbage (10 µg) was also incubated with riboprobe as a negative control. Samples were incubated at room temperature for 30 min and separated through a 15% non-denaturing acrylamide gel. Gels were dried under vacuum and bands were visualized with a phosphorimager. As a positive control for band shift, the riboprobe was incubated with a complementary anti-sense RNA, generated by in vitro transcription of a PCR product with T7 polymerase-binding site at the 3′ end of the template (PAP 286–302 for: 5′ ggctgtgaatacaatc 3′; PAP T7 309–293 rev: 5′ taatacgactcactatagggtgtagatgattg tattc 3′).

Target construct for small RNA

The PAP cDNA with 2X FLAG sequence at the 5′ end was cloned downstream of the CaMV 35S promoter placed in the pBluescript vector. Primers were designed to introduce scrambled sequence at the putative sRNA299 target site within the PAP mRNA at nt 288–308. PCR using the PAP construct as template with primers NcoI2XFlagPAP216for (5′ gatgatccatggatggac tacaaagaccatgacggtgattataaagatcatgacatcgaagggaagatga agtcgatgcttgtggtg 3′) and Scram288-305PAP263-287rev (5′ tcgtcggtgcggacaataccaagttgaagttggtgcaagaatg 3′) generated the 5′ fragment of PAP, while PCR with primers BglIIstopPAPrev (5′ gatgatagatcttcaagttgtctgacag ctcccac 3′) and Scram291-308PAP309-331for (5′ tgtccg caccgacgagcaaatgttggaagtaccaccattag 3′) generated the 3′ fragment of PAP. Subsequent overlapping PCR using these two products as template, and primers NcoI2XFlagPAP216for and BglIIstopPAPrev, generated the full-length PAP construct with a scrambled sRNA299 target site. The scrambled sequence altered the amino acid sequence within this 21 nt section but did not introduce rare codons or stop codons that would cause ribosomes to stall and trigger mRNA decay. The wild-type PAP cDNA was amplified with the same PAP-specific 5′ and 3′ primers as used for the overlapping PCR.

Isolation and transfection of pokeweed protoplasts

Protoplasts were isolated essentially as described (Koch et al. 1996) with minor modifications. Pokeweed leaves (3 g) were sliced into thin (2 mm) strips and incubated with enzyme solution (0.4 *M* mannitol, 20 m*M* MES pH 5.7, 1% cellulose RS-10, 0.15% macerozyme, 0.2% BSA) for 3 h. The mixture was filtered through cheesecloth and centrifuged in an IEC clinical centrifuge at speed 2 for 5 min. Pellets were resuspended in 0.55 *M* mannitol and overlayed onto 20% sucrose cushions. Following centrifugation in an IEC clinical centrifuge at speed 2 for 5 min, protoplasts were collected from the interface. Protoplasts were washed in W5 solution (154 m*M* NaCl, 125 m*M* $CaCl_2$, 5 m*M* KCl, 2 m*M* MES pH 5.7), incubated on ice for 30 min and centrifuged again at the same speed. Protoplast pellets were resuspended in MMg solution (400 m*M* mannitol,

15 m*M* $MgCl_2$, 4 m*M* MES pH 5.7) and counted with a hemocytometer.

Protoplast transfection was performed at room temperature. Wild-type and scrambled FLAG-tagged PAP constructs (30 µg) were mixed with 100 µl of protoplasts (5×10^4 cells ml^{-1}) and 110 µl PEG/Ca solution (200 m*M* mannitol, 100 m*M* $CaCl_2$, 40% PEG4000, 50 m*M* PIPES pH 7.0) and incubated for 20 min. Samples were diluted with 0.44 ml of W5 solution followed by centrifugation in an IEC clinical centrifuge at speed 3 for 1 min. Pellets were resuspended in 100 µl of W5 solution and 1 ml of incubation medium (0.275 *M* mannitol, 1X Aoki salts, 10% sucrose, 1 µg ml^{-1} gentamycin). Protoplasts were incubated under low light (25 µmol m^{-2} s^{-1}) at room temperature for 20 h. After the incubation, samples were centrifuged at 1000 *g* for 10 min to recover the protoplasts for total RNA isolation.

Reverse transcription-quantitative PCR

To quantify levels of PAP mRNA following JA treatment of plants, total PW RNA (8 µg) was reverse transcribed with a PAP-specific primer (PAP925Rev; 5′ gaaatcttaccccatgtctcttgc 3′) and 100 units of Superscript III (Life Technologies, Burlington, ON, CA) at 48°C for 90 min. The cDNA product (14.5 µl) was used as template for quantitative PCR using SYBR Green PCR Master Mix (Clontech, Mountain View, CA, USA) with PAP-specific forward (PAP680For; 5′ gggagtaaaatcagaagtcagg 3′) and reverse (PAP925Rev; 5′ gaaatcttaccccatgtctcttgc 3′) primers, in 20 µl total sample volume. To quantify levels of FLAG-tagged wild-type PAP or FLAG-tagged PAP scrambled mRNAs, total RNA from pokeweed protoplasts (20 µg) was reverse transcribed as above with a PAP-specific primer (PAP370-338rev; 5′ gcttcattac gaagatcattcagaaaagtggcg 3′). Quantitative PCR was performed as above with cDNA (14.5 µl), a FLAG-specific forward primer (2X FLAG; 5′ gactacaaagaccatgacggtgat tataaagatcatgac 3′) and the same primer used for cDNA synthesis. In both cases, the 28S rRNA was used as an internal control, with forward (5′ aacgtgagctgggtttagac cgtcg 3′) and reverse (5′ tcagtagggtaaaactaacc 3′) primers specific to 28S rRNA. To quantify levels of sRNA299 following JA treatment, we followed the method of stem-loop PCR essentially as described (Varkonyi-Gasic et al. 2007). LMW RNA (5 µg) was reverse transcribed as above with a sRNA299-specific hairpin primer (s299HP: 5′ gttggctctggtgcagtgtgagaggtatgcgcaccagagc caacacgtttgggc 3′). Quantitative PCR was performed as above with cDNA (14.5 µl), a sRNA299-specific forward primer (sRNA299For: 5′ gcggcggtgtagat-gattgtatt 3′) and a hairpin-specific universal primer (UniversalRev: 5′ acctctcacactgcac 3′). miR156 was used as an internal control as this miRNA has been validated as a suitable reference for biotic stress studies in other plants (Kulcheski et al. 2010). LMW RNA (5 µg) was reverse transcribed with a miR156-specific hairpin primer (miR156HP: 5′ gttggctctggtgcagtgtgagaggtatgcgcaccagagccaacacgt-gtgctc 3′) followed by quantitative PCR (qPCR) using miR156-specific forward primer (miR156For: 5′ gcggcg gtgacagaagagag 3′) and the hairpin-specific universal primer used above.

Cellular lysate preparation and immunoblot

Three leaf discs per sample (1 cm diameter each) were taken from pokeweed plants and homogenized in Buffer A (25 m*M* Tris–HCl, pH 7.5, 1 m*M* EGTA, 1 m*M* DTT, 1 m*M* PMSF, 5% glycerol). Samples were clarified by centrifugation at 16 000 *g* for 5 min at 4°C. Supernatant protein was quantified using the Bradford Assay, and 5 µg protein per sample was separated through 12% SDS-PAGE. Proteins were transferred to nitrocellulose and probed with a polyclonal antibody specific for PAP (1:5000). Blots were also probed with a polyclonal antibody for L3, a ribosomal protein, as a loading control (1:5000). PAP and L3 were visualized by chemiluminescence.

Small RNA sequencing

Total RNA from leaves of pokeweed plants treated with or without JA was sent to the Genomics Core Facility (Sunnybrook Hospital, Toronto, CA) for processing. In total, three pokeweed sRNA libraries were constructed from one (–JA) or two (+JA) plants using the strand-specific SOLiD Total RNA-Seq Kit (Life Technologies, Burlington, ON, CA). Briefly, preparation of the sRNA libraries was as follows: acrylamide gel purification of the 15–30 nt size fraction of RNA; simultaneous ligation of directional 5p and 3p primers; cDNA synthesis and acrylamide gel purification; PCR amplification (15 cycles) and column purification. Sequencing of the sRNA libraries was performed on a SOLiD 5500 XL machine. For each sRNA library, two technical replicates were sequenced.

Bioinformatics analysis of the small RNA libraries

All low quality reads (containing a fastqcssanger value below 10) and those outside of the 18–28 nt size range were removed. Adapter sequences were trimmed with Clip v.1.0.1 from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). sRNA sequences derived from rRNA, tRNA, snRNA and snoRNA were identified with Bowtie v.1.1.0 (Langmead et al.

2009) and removed, using the *Beta vulgaris* genome (RefBeet-1.1) as a reference. *B. vulgaris* was chosen because it is the closest related species to pokeweed with a sequenced genome available (Dohm et al. 2014). Remaining reads in each library were grouped if they had the same sequence and termed 'unique sequences'. To identify potential regulatory sRNAs, libraries were aligned to the reverse complement strand of PAP mRNA (Accession # AR009535.1) with Bowtie v.1.1.0. Up to three mismatches were allowed in alignments. The Integrated Genomics Viewer v.2.3 (http://www.broadinstitute.org/igv/) was used to visualize and count aligned reads. A multiple sequence alignment was generated with Clustal Omega v.1.2.1 (http://www.ebi.ac.uk/Tools/msa/clustalo/) for sRNA299 reads in the –JA and +JA libraries.

Statistical analysis

Significance testing was performed for all quantitative reverse-transcription PCR (RT-PCR) results. A one-tailed, unpaired Student's *t*-test was conducted using GraphPad Prism v.5.01.

Results

PAP mRNA is cleaved 298 nucleotides 3′ of the transcription start site

To investigate whether PAP mRNA may be targeted for cleavage by a sRNA, total RNA from pokeweed plants (PW) was probed for PAP mRNA by northern blot. Total cabbage (CB) RNA was used as a negative control, based on a literature search indicating no known RIP in this plant. We detected full-length PAP mRNA, based on comparison with an in vitro transcript of PAP, plus a lower molecular mass band, suggesting cleavage of PAP mRNA (Fig. 1A). To identify the site of cleavage, a gene-specific primer for PAP was extended by reverse transcriptase on total RNA from the same samples. A band indicating termination of extension was observed in pokeweed RNA but not in cabbage RNA (Fig. 1B). Sequencing of a PAP cDNA with the same reverse primer indicated that the cDNA terminated at nt A299, relative to the 5′ transcription start site of PAP mRNA. 5′ rapid extension of cDNA ends analysis was also conducted on total PW RNA. A PCR product indicative of full-length PAP mRNA was observed, along with a smaller PCR product, supporting truncation of the PAP mRNA (Fig. 1C). Sequencing of the smaller product confirmed that PAP mRNA was cleaved between nt C298 and A299. This cleavage occurred within the open reading frame of the PAP mRNA (Fig. 1D) and indicates the mRNA may be targeted by a small RNA.

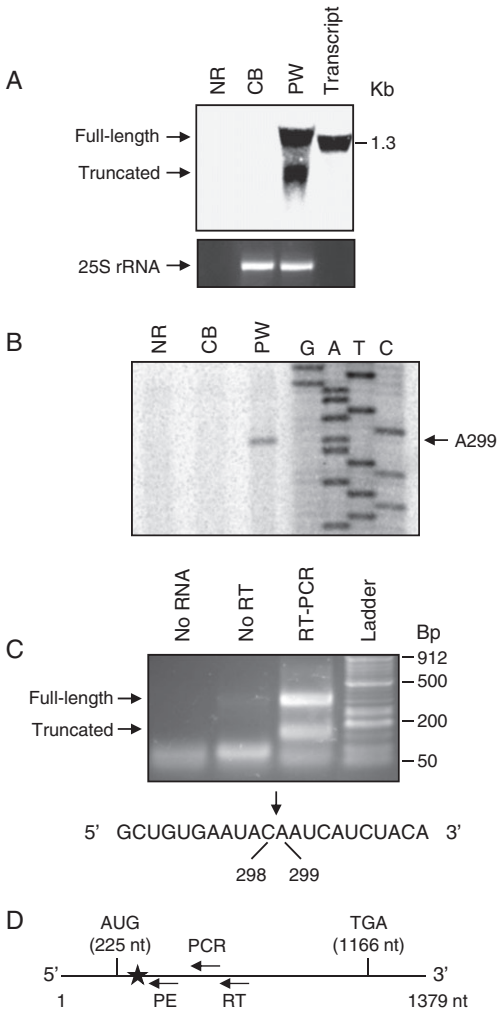

Fig. 1. PAP mRNA is cleaved 298 nucleotides 3′ of the transcription start site. (A) Total PW and CB RNA (38 μg) was separated through 1.5% agarose gel in 7% formamide, transferred to nylon membrane and probed with a PAP-specific minus-strand riboprobe. An in vitro transcript of full-length PAP mRNA (0.5 μg) was also separated as a positive control and a relative size marker. Prior to transfer, the gel was stained with ethidium bromide to visualize the 25S rRNA, as an indication of amount of total RNA loaded. (B) Pokeweed total RNA (PW) was template for primer extension using PAP-specific radiolabeled reverse primer complementary to region 340–361 nt. CB RNA was analyzed as a negative control, and No RNA (NR) indicates radiolabeled reverse primer without extension template. cDNA products were separated through 7 *M* urea/6% acrylamide gel and visualized with a phosphorimager. Dideoxynucleotide sequencing of a PAP cDNA with the same primer was used to identify the reverse transcriptase stall site. (C) 5′ RACE analysis of total PW RNA. PCR products were separated on 1% agarose gel and stained with ethidium bromide. Sequencing results of the cleaved PCR product indicate PAP mRNA cleavage between C298 and A299 (arrow) of full-length PAP mRNA. Both gels are representative of four independent experiments. (D) Diagram of full-length PAP mRNA with nucleotide number (1379 nt), indicating site of cleavage (star; 298–299 nt) and location of reverse primers used for primer extension (PE; 340–361 nt), reverse transcriptase (RT; 607–632 nt) and PCR (532–512 nt) for 5′ RACE analyses.

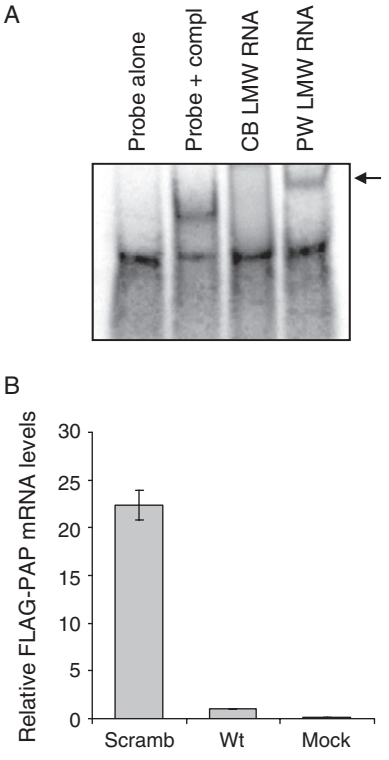

Fig. 2. sRNA299 targets PAP mRNA in vivo. (A) LMW RNA (10 µg) of PW and cabbage (CB) were incubated with a radiolabeled riboprobe of the PAP sequence from nt 286–309 downstream of the transcription start site. Following incubation, samples were separated on a non-denaturing, 15% acrylamide gel and visualized with a phosphorimager. As a positive control for band shift, the riboprobe was incubated with its complementary non-radiolabeled sequence (probe + complement). The gel is representative of three independent experiments. (B) Pokeweed protoplasts were transfected with either FLAG-tagged wild-type PAP cDNA or FLAG-tagged PAP cDNA with scrambled sRNA299 target sequence. Quantitative RT-PCR measurements of scrambled FLAG-PAP mRNA are indicated relative to wild-type FLAG-PAP mRNA. 28S rRNA was used as an internal control. Mock represents cells transfected with vector lacking insert. Values are means ± SE for five independent experiments.

A small RNA binds to PAP mRNA in vivo

To test whether a sRNA from pokeweed bound to the observed cleavage site of PAP mRNA, we performed an RNA electrophoretic mobility shift assay. A (+)-strand, radiolabeled in vitro transcript of the 23 nt region of PAP mRNA potentially targeted by the sRNA was synthesized, based on the assumption that cleavage of PAP mRNA at A299 would result from the binding of a reverse complementary sRNA extending approximately 11 nt in the 5′ and 3′ direction from A299. This was reasoned, as plant miRNA-target interactions tend to cause cleavage between positions 10 and 11 of the alignment (German et al. 2008). The 23 nt transcript was used as a probe and incubated with the low molecular weight RNA pool isolated from pokeweed. A resulting shift in migration of the probe was detected (Fig. 2A). This band (indicated by arrow) migrated slightly slower than the migration of the positive control, which was an unlabeled 23 nt in vitro transcript complementary to the probe. Therefore, a sRNA from pokeweed binds to the PAP mRNA sequence and may be longer than 23 nt. We designated this sRNA 'sRNA299', referring to its induced cleavage between nt 298/299 of the PAP mRNA sequence. Incubation of the probe with LMW RNA from cabbage served as a negative control and did not produce a shift in probe migration, indicating some specificity of the probe for a sRNA from pokeweed.

To determine whether sRNA299 reduced levels of PAP mRNA in vivo, constructs of FLAG-tagged PAP, to distinguish them from endogenous PAP, were transfected into pokeweed protoplasts. The constructs contained either the wild-type target sequence for sRNA299, or scrambled target sequence. Quantitative RT-PCR demonstrated that FLAG-PAP mRNA levels bearing the scrambled target sequence had a significant, 22-fold increase compared with FLAG-PAP mRNA levels with wild-type sequence ($P < 0.01$, Student's *t*-test; Fig. 2B). Taken together, these results show that a sRNA binds a specific sequence within the open reading frame of PAP mRNA in vivo, resulting in reduced levels of full-length PAP mRNA.

JA increases the level of PAP mRNA and sRNA299

To investigate the possible physiological role of the sRNA, we treated pokeweed leaves with JA and measured the levels of PAP mRNA and sRNA299. Given that JA signals a defense response to herbivory and pathogen attack, we hypothesized that it would increase the levels of PAP mRNA and resulting protein. This was indeed the case, as application of 5 m*M* JA increased PAP mRNA sevenfold and protein levels beyond control plants without the hormone treatment (Fig. 3A, B). The increase in PAP mRNA with JA, relative to water, was significant ($P < 0.05$, Student's *t*-test). We initially compared the effect of water and 0.5% ethanol on the level of PAP mRNA because JA was prepared in 0.5% ethanol. However, there was no obvious difference between water and 0.5% ethanol on PAP mRNAs levels; therefore, subsequent analyses were performed on plants treated with JA in 0.5% ethanol (+JA) and compared with 0.5% ethanol alone (−JA). We quantified the level of sRNA299 under the same conditions and observed a significant, 3.4-fold increase with JA ($P < 0.01$, Student's *t*-test; Fig. 4A). Treatment of pokeweed plants with JA also increased the amount of PAP mRNA cleavage at the A299 target site, relative to untreated plants (Fig. 4B). These results suggest that PAP levels increase during JA stress, which agrees

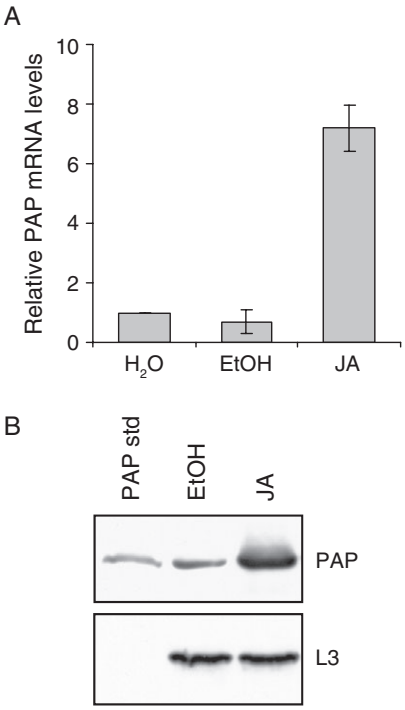

Fig. 3. JA treatment increases levels of PAP mRNA and protein. Pokeweed plants were treated with 5 m*M* JA 24 h prior to RNA or protein extraction. (A) Quantitative RT-PCR measurements of PAP mRNA from pokeweed total RNA are indicated. 28S rRNA was used as an internal control. Values are relative to the water-treated sample and are means ± SE for three independent experiments. (B) Total PW protein (5 μg) was separated by 12% SDS-PAGE, transferred to nitrocellulose and probed with a polyclonal antibody specific to PAP (1:5000). The blot was also probed with a polyclonal antibody specific for L3, a large ribosomal subunit protein, as a loading control (1:5000). PAP standard (std) is purified PAP from pokeweed. The blot is representative of three independent experiments.

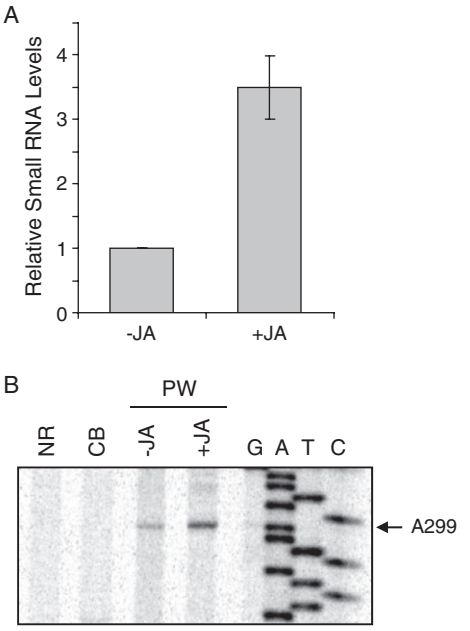

Fig. 4. JA treatment increases levels of sRNA299 and PAP mRNA cleavage. Pokeweed plants were treated with 5 m*M* JA 24 h prior to RNA extraction. (A) Quantitative RT-PCR measurements of sRNA299 from pokeweed low molecular weight RNA are indicated. Measurements used miR156 as an internal control. Values are relative to −JA and are means ± SE for three independent experiments. (B) Pokeweed total RNA (PW) was template for primer extension using PAP-specific radiolabeled reverse primer complementary to region 340–361 nt. CB RNA was analyzed as a negative control, and No RNA (NR) indicates radiolabeled reverse primer without extension template. cDNA products were separated through 7 *M* urea/6% acrylamide gel and visualized with a phosphorimager. Dideoxynucleotide sequencing of a PAP cDNA with the same primer was used to identify the reverse transcriptase stall site. The gel is representative of four independent experiments.

with the hypothesized role of PAP as a defense protein. Moreover, the increase in the sRNA suggests that it offsets a JA-induced transcriptional increase in PAP mRNA.

Pokeweed shows a typical distribution of small RNA sequences

To characterize the sRNA content of pokeweed, we performed sRNA sequencing of plants treated with or without JA. The sRNA libraries illustrated a typical angiosperm-specific size distribution, with 23–24 nt RNAs being most abundant, followed by 20–22 nt RNAs (Fig. S1, Supporting Information). The distributions of unique sequences and total reads were similar. Averaging the results across unique sequences, reads and treatments, the percentages of 20 nt, 21 nt, 22 nt, 23 nt and 24 nt sRNAs were 2.29, 5.50, 5.30, 20.91 and 60.65, respectively. Because of the limited biological variability represented by our RNA-Seq samples, we did not investigate the effect of JA, if any, on the sRNA distribution in pokeweed.

RNA-Seq confirms the existence of a small RNA targeting PAP mRNA

Sequencing of the small RNA pool from pokeweed revealed several small RNAs that aligned to the PAP mRNA, in both the plus- and minus-sense orientation. Based on the observed cleavage at ~300 nt of pokeweed mRNA, we expected to detect small RNAs aligning to this particular region. To identify sRNA299 in pokeweed, libraries were aligned to the reverse complement sequence of PAP mRNA. Between the –JA and +JA libraries, eight reads aligned to the previously characterized PAP mRNA target site. As determined by multiple sequence alignment, the reads ranged in length from 18

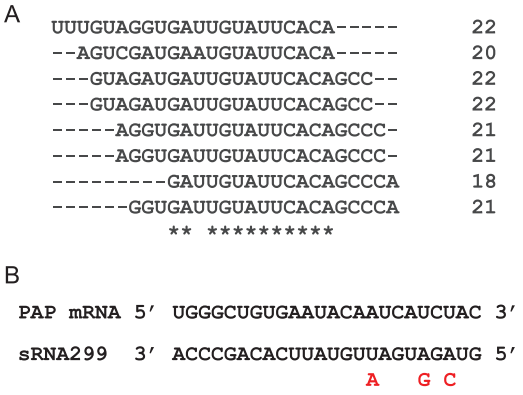

Fig. 5. RNA-Seq confirms the existence of sRNA299 in pokeweed. The +JA and −JA libraries were aligned to the PAP mRNA sequence to identify sRNAs that could regulate the transcript. (A) The multiple sequence alignment for sRNA299 reads present in the +JA and −JA libraries. (B) An illustration of sRNA299 binding to PAP mRNA via complementary base-pairing. Letters in red indicate nucleotides that are not complementary to PAP mRNA but were conserved in some sRNA299 reads.

to 22 nt and were highly conserved (Fig. 5A). The resultant sRNA299 consensus sequence showed high complementarity to the PAP mRNA target site (Fig. 5B). Of the eight reads, three were completely complementary to PAP mRNA and five deviated by only 1–3 nt. These differences could arise from sequencing errors or from different small RNA precursors. Taken together, these results confirm the presence of sRNA299 and its ability to target the PAP mRNA through complementary base-pairing.

Discussion

We show here that PAP mRNA is targeted by a sRNA, which results in cleavage of the message within its open reading frame. Therefore, the sRNA negatively regulates PAP mRNA. Interestingly, JA treatment increases the levels of both PAP mRNA and the sRNA. In most documented cases, the level of sRNA is inversely related to the level of its target RNA; however, recent publications of genome-wide analyses demonstrate that several plant sRNA-target pairs have positively correlated expression patterns (He et al. 2013, Peng et al. 2014). Indeed, an equal number of negatively and positively correlated miRNA-target mRNA pairs have been described during tomato fruit development (Lopez-Gomollon et al. 2012). Positive correlations suggest mechanisms of action that are not limited to classic temporal regulation where negative correlation is evident. For example, miR395 and its target mRNA SULTR2;1 are both induced by low sulfur levels but each is transcribed in a separate neighboring cell type (Kawashima et al. 2009). Therefore, there is an exclusion between the small RNA and its target mRNA, and miR395 may function as a secondary means of preventing SULTR2:1 expression in specific cell types. Other scenarios describe how the mRNA is transcribed equally in two domains; yet, the small RNA is transcribed in only one, with the purpose of excluding the expression of the target in one domain only (Kidner and Martienssen 2004; Levine et al. 2007). Another type of regulation occurs when the small RNA diminishes or modulates the expression of the mRNA target and the degree of target reduction depends on the level of miRNA (Nikovics et al. 2006). Such a mechanism could play an important role in controlling the toxicity of PAP. We suggest that the increase in sRNA may be a means to limit the over-expression of PAP during induction. For example, transgenic plants expressing PAP from a constitutive 35S-CaMV promoter displayed symptoms of toxicity, including stunted growth, mottled leaves and sterility (Lodge et al. 1993). Authors reported correct localization of PAP, indicating deregulated PAP expression likely contributed to the observed toxicity.

sRNAs represent only one level of regulation of gene expression, which is also controlled by transcription factors/promoter sequences and mRNA stability. The contribution of each may mask the negative effect of a small RNA on target mRNA levels. RIP induction by hormones associated with biotic stress has been reported previously. For example, RIPs induced by plant hormones include JIP60, a JA-induced Type III RIP from barley (Dunaeva et al. 1999), beetin, a salicylic acid and hydrogen peroxide-induced Type I RIP from sugar beet (Iglesias et al. 2008) and soRIP2, a salicylic acid-induced Type I RIP from spinach (Kawade and Masuda 2009). Some RIPs are induced directly by stimuli upstream of the hormone signal, providing insight into their functional targets. For example, RIP2, a Type III RIP from maize with insecticidal activity, was induced 100-fold at the RNA level, with a concomitant increase in protein, upon caterpillar feeding but not mechanical wounding (Chuang et al. 2014). Other RIPs have shown pathogen-inducible expression patterns, including beetin and curcin, which are virus and fungus inducible, respectively (Girbés et al. 1996, Qin et al. 2010). Although the PAP gene promoter remains unknown, it is probable that JA treatment induces transcriptional activation of the PAP gene. Our finding that PAP expression is enhanced by JA agrees with previous reports which place RIPs within a signaling cascade mediated by biotic stress. While the antiviral activity of PAP is well established, the induction of PAP by JA suggests that PAP could play a more broad-spectrum role against herbivores and necrotrophic pathogens, well-known elicitors of JA signaling. The targeting of PAP by a small RNA that is also induced by JA may serve to fine-tune PAP levels during stress.

Our finding that PAP mRNA is targeted by a sRNA may explain some discrepancies reported previously between RIP transcript and protein levels. For example, in sugar beet, beetin transcripts were found at similar levels throughout plant development but the protein was only synthesized in adult plants (Iglesias et al. 2008). Upon treatment with mediators of systemic acquired resistance, beetin expression increased at the RNA and protein levels but only in adult plants. Authors suggested that variations in transcript stability, polyadenylation or protein factor recognition could account for developmental regulation of this Type I RIP. In this study, we found 5′ truncated PAP transcripts in pokeweed because of sRNA-associated cleavage. A developmentally regulated sRNA that targets beetin may also contribute to its observed expression pattern in sugar beet.

Although the abundance of sRNA299 in our RNA-Seq libraries was low, averaging less than 1 read per million (data not shown), it is possible that this sRNA shows a developmental or spatial expression pattern that is important for PAP regulation in pokeweed. A recent report found that the vast majority of miRNAs in plants are of low abundance and are species-specific (Montes et al. 2014). In contrast, highly abundant miRNA families tend to be well conserved among plants (Montes et al. 2014). RIPs have been found in less than 20% of angiosperm taxonomic orders, indicating that they are poorly conserved (Di Maro et al. 2014). We searched current plant sRNA databases (plant miRNA database: http://bioinformatics.cau.edu.cn/PMRD/ and tasiRNA database: http://bioinfo.jit.edu.cn/tasiRNADatabase/) for sequences aligning to sRNA299 and did not find any, suggesting that sRNA299 may be unique to pokeweed. Taken together, the low abundance and apparent lack of conservation of sRNA299 reported here does not undermine its functional importance in pokeweed.

Future work will focus on identifying the origin of sRNA299. Presently, our analysis selected for sRNA299 sequences with high complementarity to PAP mRNA and their presence suggests processing from a siRNA pathway. Our electrophoretic mobility shift assay results indicate that the sRNA299 probe annealed to a sRNA larger than 23 nt, although this could be attributed to the use of a non-denaturing gel, or the binding of a third component within the cellular lysate. To our knowledge, this is the first report of a sRNA targeting a RIP mRNA; the expression of other plant RIPs may be controlled in a similar manner. Additionally, this work suggests an endogenous mechanism to modulate the expression of a RIP, possibly limiting its associated toxicity. This may have important implications in transgenic plant systems and agricultural selection programs.

Authors' contributions

L. A. B., K. C. M. N. and K. A. H. conceived the experiments. A. K., K. C. M. N. and G. K. designed and performed the experiments. K. C. M. N. drafted the manuscript. K. A. H. edited the manuscript. All authors approved the final version.

Acknowledgements – The authors are grateful to Dr Yutaka Amemiya, Manager, Genomics Core Facility, Next Generation Sequencing Lab, Sunnybrook Research Institute for small RNA sequencing and for his assistance and advice with sequence analysis. This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to K. A. H. and Canada Graduate Scholarships to K. C. M. N. and G. K.

References

Bologna NG, Voinnet O (2014) The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. Annu Rev Plant Biol 65: 473–503

Chuang W-P, Herde M, Ray S, Castano-Duque L, Howe GA, Luthe DS (2014) Caterpillar attack triggers accumulation of the toxic maize protein RIP2. New Phytol 201: 928–939

Di Maro A, Citores L, Russo R, Iglesias R, Ferreras JM (2014) Sequence comparison and phylogenetic analysis by the maximum likelihood method of ribosome-inactivating proteins from angiosperms. Plant Mol Biol 85: 575–588

Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, Rupp O, Sörensen TR, Stracke R, Reinhardt R, Goesmann A, Kraft T, Schulz B, Stadler PF, Schmidt T, Galadón T, Lehrach H, Weisshaar B, Himmelbauer H (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505: 546–549

Dunaeva M, Goebel C, Wasternack C, Parthier B, Goerschen E (1999) The jasmonate-induced 60 kDa protein of barley exhibits N-glycosidase activity in vivo. FEBS Lett 452: 263–266

Endo Y, Tsurugi K, Lambert JM (1988) The site of action of six different ribosome-inactivating proteins from plants on eukaryotic ribosomes: the RNA N-glycosidase activity of the proteins. Biochem Biophys Res Commun 150: 1032–1036

German MA, Pillay M, Jeong DH, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, De Paoli E, Lu C, Schroth G, Meyers BC, Green PJ (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol 26: 941–946

Gessner SL, Irvin JD (1980) Inhibition of elongation factor 2-dependent translocation by the pokeweed antiviral protein and ricin. J Biol Chem 255: 3251–3253

Girbés T, de Torre C, Iglesias R, Ferrerras JM (1996) RIP for viruses. Nature 379: 777–778

He C, Cui K, Zhang J, Duan A, Zeng Y (2013) Next-generation sequencing-based mRNA and microRNA expression profiling analysis revealed pathways involved in the rapid growth of developing culms in Moso bamboo. BMC Plant Biol 13: 119

Honjo E, Dong D, Motoshima H, Watanabe K (2002) Genomic clones encoding two isoforms of pokeweed antiviral protein in seeds (PAP-S1 and S2) and the N-glycosidase activities of their recombinant proteins on ribosomes and DNA in comparison with other isoforms. J Biochem 131: 225–231

Iglesias R, Pérez Y, Citores L, Ferreras JM, Méndez E, Girbés T (2008) Elicitor-dependent expression of the ribosome-inactivating protein beetin is developmentally regulated. J Exp Bot 59: 1215–1223

Irvin J (1975) Purification and partial characterization of the antiviral protein from *Phytolacca americana* which inhibits eukaryotic protein synthesis. Arch Biochem Biophys 169: 522–528

Irvin JD, Kelly T, Robertus JD (1980) Purification and properties of a second antiviral protein from *Phytolacca americana* which inactivates eukaryotic ribosomes. Arch Biochem Biophys 200: 418–425

Karran RA, Hudak KA (2008) Depurination within the intergenic region of Brome mosaic virus RNA3 inhibits viral replication in vitro and in vivo. Nucleic Acids Res 36: 7230–7239

Kataoka J, Habuka N, Masuta C, Miyano M, Koiwai A (1992) Isolation and analysis of a genomic clone encoding a pokeweed antiviral protein. Plant Mol Biol 20: 879–886

Kaur I, Gupta RC, Puri M (2011) Ribosome inactivating proteins from plants inhibiting viruses. Virol Sin 26: 357–365

Kawade K, Masuda K (2009) Transcriptional control of two ribosome-inactivating protein genes expressed in spinach (*Spinacia oleracea*) embryos. Plant Physiol Biochem 47: 327–334

Kawade K, Ishizaki T, Masuda K (2008) Differential expression of ribosome-inactivating protein genes during somatic embryogenesis in spinach (Spinacia oleracea). Physiol Plant 134: 270–281

Kawashima CG, Yoshimoto N, Maruyama-Nakashita A, Tsuchiya YN, Saito K, Takahashi H, Dalmay T (2009) Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. Plant J 57: 313–321

Khraiwesh B, Zhu J-K, Zhu J (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim Biophys Acta 1819: 137–148

Kidner CA, Martienssen RA (2004) Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature 428: 81–84

Koch PE, Bonness MS, Lu H, Mabry TJ (1996) Protoplasts from *Phytolacca dodecandra* L'Herit (endod) and *P. americana* L. (pokeweed). Plant Cell Rep 15: 824–828

Kulcheski FR, Marcelino-Guimaraes FC, Nepomuceno AL, Abdelnoor RV, Margis R (2010) The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. Anal Biochem 406: 185–192

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25

Levine E, McHale P, Levine H (2007) Small regulatory RNAs may sharpen spatial expression patterns. PLoS Comput Biol 3: e233

Lodge JK, Kaniewski WK, Tumer NE (1993) Broad-spectrum virus resistance in transgenic plants expressing pokeweed antiviral protein. Proc Natl Acad Sci USA 90: 7089–7093

Lopez-Gomollon S, Mohorianu I, Szittya G, Moulton V, Dalmay T (2012) Diverse correlation patterns between microRNAs and their targets during tomato fruit development indicates different modes of microRNA actions. Planta 236: 1875–1887

Loss-Morais G, Turchetto-Zolet AC, Etges M, Cagliari A, Körbes AP, Maraschin Fdos S, Margis-Pinheiro M, Margis R (2013) Analysis of castor bean ribosome-inactivating proteins and their gene expression during seed development. Genet Mol Biol 36: 74–86

Montes RA, de Fátima R-CF, De Paoli E, Accerbi M, Rymarquis LA, Mahalingam G, Marsch-Martínez N, Meyers BC, Green PJ, de Folter S (2014) Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nat Commun 5: 3722

Nikovics K, Blein T, Peaucelle A, Ishida T, Morin H, Aida M, Laufs P (2006) The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. Plant Cell 18: 2929–2945

Parente A, Conforto B, Di Maro A, Chambery A, De Luca P, Bolognesi A, Iriti M, Faoro F (2008) Type 1 ribosome-inactivating proteins from *Phytolacca dioica* L. leaves: differential seasonal and age expression, and cellular localization. Planta 228: 963–975

Peng Z, He S, Gong W, Sun J, Pan Z, Xu F, Lu Y, Du X (2014) Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes. BMC Genomics 15: 760

Qin W, Ming-Xing H, Ying X, Xin-Shen Z, Fang C (2005) Expression of a ribosome inactivating protein (curcin 2) in *Jatropha curcas* is induced by stress. J Biosci 30: 351–357

Qin X, Shao C, Hou P, Gao J, Lei N, Jiang L, Ye S, Gou C, Luo S, Zheng X, Gu X, Zhu X, Xu Y, Chen F (2010) Different functions and expression profiles of curcin and

curcin-L in *Jatropha curcas* L. Z Naturforsch C 65: 355–362
Rajamohan F, Venkatachalam TK, Irvin JD, Uckun FM (1999) Pokeweed antiviral protein isoforms PAP-I, PAP-II, and PAP-III depurinate RNA of human immunodeficiency virus (HIV)-1. Biochem Biophys Res Commun 260: 453–458
Song SK, Choi Y, Moon YH, Kim S-G, Choi YD, Lee JS (2000) Systemic induction of a *Phytolacca insularis* antiviral protein gene by mechanical wounding, jasmonic acid, and abscisic acid. Plant Mol Biol 43: 439–450
Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 3: 12
Vepachedu R, Bais HP, Vivanco JM (2003) Molecular characterization and post-transcriptional regulation of ME1, a type-I ribosome-inactivating protein from *Mirabilis expansa*. Planta 217: 498–506

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Pokeweed shows a typical small RNA size distribution. RNA-Seq was performed from low molecular weight RNA extracted 24 h after treatment of pokeweed plants with or without 5 m*M* JA. (A) Size distribution of unique sequences. (B) Size distribution of reads. Values are means ± SE for two technical replicates from one plant (–JA) or two technical replicates from two plants (+JA).

Edited by H. Saitoh

5.2 - List of additional contributions

Alekseenko, A., Barrett, D., Pareja-Sanchez, Y., Howard, R. J., Strandback, E., Ampah-Korsah, H., Rovšnik, U., Zuniga-Veliz, S., **Klenov, A.**, Malloo, J., Ye, S., Liu, X., Reinius, B., Elsässer, S., Nyman, T., Sandh, G., Yin, X., & Pelechano, V. (2020). Detection of SARS-CoV-2 using non-commercial RT-LAMP reagents and raw samples. *MedRxiv*, 2020.08.22.20179507. https://doi.org/10.1101/2020.08.22.20179507

Jobst, K. A., **Klenov, A.**, Neller, K. C. M., & Hudak, K. A. (2016). Effect of depurination on cellular and viral RNA. In S. Jurga, V. A. Erdmann, & J. Barciszewski (Eds.), *Modified Nucleic Acids in Biology and Medicine* (pp. 273–297). Springer. https://doi.org/10.1007/978-3-319-34175-0_12

Neller, K. C. M., **Klenov, A.**, Guzman, J. C., & Hudak, K. A. (2018). Integration of the pokeweed miRNA and mRNA transcriptomes reveals targeting of jasmonic acid-responsive genes. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00589

Neller, K. C. M., **Klenov, A.**, & Hudak, K. A. (2016). The pokeweed leaf mRNA transcriptome and its regulation by jasmonic acid. *Frontiers in Plant Science*, *7*(March), 1–13. https://doi.org/10.3389/fpls.2016.00283

Neller, K. C. M., **Klenov, A.**, & Hudak, K. A. (2019). Prediction and characterization of miRNA/target pairs in non-model plants using RNA-seq. *Current Protocols in Plant Biology*, *4*(2), e20090. https://doi.org/10.1002/cppb.20090

5.3 – Supplemental Table for Chapter 2

Supplementary Table 1 – Primers used for sequencing and cloning of PkMV-AR

Primer Name	Sequence (5' to 3')	Purpose
A1	GCTGTCAACGATACGCTACGTAACGGCATGACAGTG-d(T)24	RT for sequencing PkMV
A2	TGTGCCTGAGGAGACTTTCAAAATCACG	PCR for Seq3
A3	GCTGTCAACGATACGCTACGTAACG	PCR for Seq3
A4	GCGAACAGTGAACACTTACCCTTG	PCR for Seq2
A5	CTCTCGAGTCCTGTCTGC	PCR for Seq2
A6	GCCATCTCTTCCACCAGTGATGATT	PCR for Seq1
A7	AAAATAAAACAAAACATCACAACATAACAC	PCR for Seq1
B1	d(T)21-CCTCTTTTACCTCTGTACATGTGTTCCACGAGGTAAAACC	RT for cloning, PCR for PkMV-3'
B2	GTAGATTTTGGGATTAAAGTGCAACCCACACTAGATTGCGACAACCGCATGATATCGTACC	PCR for PkMV-3'
B3	AAAATAAAACAAAACATCACAACATAACACAAAACAATCTATCAACTCTCAAGCTCTCAG	PCR for PkMV-5'
B4	GGTACGATATCATGCGGTTGTCGCAATCTAGTGTGGGTTGCACTTTAATCCCAAAATCTAC	PCR for PkMV-5'
B5	CACATTTCCCCGAAAAGTGCCACCTGAACG	PCR for PkMV-Vec
B6	CCTGATGCGGTATTTTCTCCTTACGCATCTGTGC	PCR for PkMV-Vec
C1	ACGTGGAGACAACATTCTTGTGTACGTTGC	PCR for PkMV-Check
C2	GTTTAACCGTGCACTGCGATAGAAGACTC	PCR for PkMV-Check
D1	TCAGTTGCTGAGACGTTCTGGAGGGGTTTTAATAG	PCR for PkMV-eGFP Fragment 1
D2	GATAGCACATAACTTGTAGAGGTGCTATAATG	PCR for PkMV-eGFP Fragment 1
D3	CATTATAGCACCTCTACAAGTTATGTGCTATC	PCR for PkMV-eGFP Fragment 2
D4	ATAATGATTCATTGTGCGACGGACTCCATCTGTTACAC	PCR for PkMV-eGFP Fragment 2
D5	GTCGCACAATGAATCATTATTCAATGGTGAGCAAGGGCGAGGAGCTGTTCACCG	PCR for eGFP insert
D6	CCAGAACGTCTCAGCAACTGACTGGAAATGCACTTCAATCTTGTACAGCTCGTCCATG	PCR for eGFP insert

5.4 – Supplemental Tables for Chapter 3

Supplementary Table 2 – Primers used for generation of viral libraries, scrambled/positive controls and NGS preparation.

Primer Name	Sequence 5' to 3'	Purpose
Lib-For	ATCGAGGTCTCAATGGNNNNNN	Generation of dsDNA viral library for insertion into reporter plasmid
Lib-Rev	TGAGCGGTCTCTAGGTNNNNNN	Generation of dsDNA viral library for insertion into reporter plasmid
Scr-Top	ATGGGATCGGATACGTACCTCGAGTTAATGCTACGCTAT	Scrambled positive control for fluorescence measurements
Scr-Bott	AGGTATAGCGTAGCATTAACTCGAGGTACGTATCCGATC	Scrambled positive control for fluorescence measurements
Pos-Top	ATGGTTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGC	Positive control for fluorescence measurements, strong Sigma70 promoter
Pos-Bott	AGGTGCTAGCATTATACCTAGGACTGAGCTAGCTGTCAA	Positive control for fluorescence measurements, strong Sigma70 promoter
Lib-Seq-For	GATCAGCTCGAGTGCCACCTG	Sanger sequencing of eGFP library
Lib-Seq-Rev	TTGTGCCCATTAACATCACCA	Sanger sequencing of eGFP library
NGS-For	TAGAGACTAGTGGAAGACATGGAG	Amplification of chloramphenicol library prior to NGS
NGS-Rev	TTAAGTGAACTTGGGCCCAGTA	Amplification of chloramphenicol library prior to NGS
Poly-d(T)23-VN	TTTTTTTTTTTTTTTTTTTTTTTVN	Reverse transcription of viral RNA for library generation

Supplementary Table 3 – Primers used to amplify PkMV PPRs for fluorescence measurement and to obtain +1TSS with TSO-5'RACE.

Primer Name	Sequence 5' to 3'	Purpose
PPR6F-F	ATCGAGGTCTCAATGGTTCATTTGAAGAGTTCG TGA	PkMV specifc putative promoter region 6F primer
PPR6F-R	TGAGCGGTCTCTAGGTGTGGAGCCAATTTTCAT GGT	PkMV specifc putative promoter region 6F primer
PPR6R-F	TGAGCGGTCTCTAGGTTTCATTTGAAGAGTTCG TGA	PkMV specifc putative promoter region 6R primer
PPR6R-R	ATCGAGGTCTCAATGGGTGGAGCCAATTTTCAT GGT	PkMV specifc putative promoter region 6R primer
PPR7F-F	ATCGAGGTCTCAATGGTGTAGTCGTAGATTTTG GGA	PkMV specifc putative promoter region 7F primer
PPR7F-R	TGAGCGGTCTCTAGGTAAGAAAGCAGCCTCTGT TGC	PkMV specifc putative promoter region 7F primer
PPR7R-F	TGAGCGGTCTCTAGGTTGTAGTCGTAGATTTTG GGA	PkMV specifc putative promoter region 7R primer
PPR7R-R	ATCGAGGTCTCAATGGAAGAAAGCAGCCTCTG TTGC	PkMV specifc putative promoter region 7R primer
PPR8F-F	ATCGAGGTCTCAATGGACAAGCACGAACAATG GTGC	PkMV specifc putative promoter region 8F primer
PPR8F-R	TGAGCGGTCTCTAGGTGTTTATTGAGGATTGTT TCG	PkMV specifc putative promoter region 8F primer
PPR8R-F	TGAGCGGTCTCTAGGTACAAGCACGAACAATG GTGC	PkMV specifc putative promoter region 8R primer
PPR8R-R	ATCGAGGTCTCAATGGGTTTATTGAGGATTGTT TCG	PkMV specifc putative promoter region 8R primer
PPR9F-F	ATCGAGGTCTCAATGGAGCAAACAGCTGTCTG GTTG	PkMV specifc putative promoter region 9F primer
PPR9F-R	TGAGCGGTCTCTAGGTAAAGAAAGGGATCTTG ACAT	PkMV specifc putative promoter region 9F primer
PPR9R-F	TGAGCGGTCTCTAGGTAGCAAACAGCTGTCTGG TTG	PkMV specifc putative promoter region 9R primer
PPR9R-R	ATCGAGGTCTCAATGGAAAGAAAGGGATCTTG ACAT	PkMV specifc putative promoter region 9R primer
PPR10F-F	ATCGAGGTCTCAATGGGCCAGACAAGATGCAT GAGG	PkMV specifc putative promoter region 10F primer
PPR10F-R	TGAGCGGTCTCTAGGTGTACGCGGTATGGCGTG GAT	PkMV specifc putative promoter region 10F primer
PPR10R-F	TGAGCGGTCTCTAGGTGCCAGACAAGATGCAT GAGG	PkMV specifc putative promoter region 10R primer
PPR10R-R	ATCGAGGTCTCAATGGGTACGCGGTATGGCGTG GAT	PkMV specifc putative promoter region 10R primer
PPR11F-F	ATCGAGGTCTCAATGGTCTCTGTCATCCATAAT GAC	PkMV specifc putative promoter region 11F primer

PPR11F-R	TGAGCGGTCTCTAGGTCCATTGCGCTGGGATCATGC	PkMV specifc putative promoter region 11F primer
PPR11R-F	TGAGCGGTCTCTAGGTTCTCTGTCATCCATAATGAC	PkMV specifc putative promoter region 11R primer
PPR11R-R	ATCGAGGTCTCAATGGCCATTGCGCTGGGATCATGC	PkMV specifc putative promoter region 11R primer
PPR12F-F	ATCGAGGTCTCAATGGACAAGAACCAGCGTCTCAGA	PkMV specifc putative promoter region 12F primer
PPR12F-R	TGAGCGGTCTCTAGGTCCAGTCAACGGATCCACAAA	PkMV specifc putative promoter region 12F primer
PPR12R-F	TGAGCGGTCTCTAGGTACAAGAACCAGCGTCTCAGA	PkMV specifc putative promoter region 12R primer
PPR12R-R	ATCGAGGTCTCAATGGCCAGTCAACGGATCCACAAA	PkMV specifc putative promoter region 12R primer
TSO	TACTTCCAATCCAATGRGRGR	Template switching oligo, DNA/RNA hybrid
TSO-For	TACTTCCAATCCAATGGG	TSO specific forward primer
Lib-5R-RT	CGTGTCTTGTAGTTCCCGTCA	RT primer for TSO-5'RACE of eGFP
Lib-5R-Rev1	TCTCGCAAAGCATTGAACACC	eGFP specific reverse primer for PCR amplification
Lib-5R-RevA1	TTATCCACTTCCAATTCTCGCAAAGCATTGAACACC	eGFP specific reverse primer with adapter for cloning into sequencing vector
Lib-5R-Rev2	TAACCTTCGGGCATGGCACTCT	Nested eGFP specific reverse primer for PCR amplification
Lib-5R-RevA2	TTATCCACTTCCAATTAACCTTCGGGCATGGCACTCT	Nested eGFP specific reverse primer with adapter for cloning into sequencing vector
pLK-For	ATTGGAAGTGGATAAGCATG	Linearizing sequencing vector
pLK-Rev	ATTGGATTGGAAGTACTATAGTGA	Linearizing sequencing vector

Supplementary Table 4 – Cryptic promoters as top and bottom strand oligos for fluorescence measurements

Primer Name	Sequence 5' to 3'	Purpose
9R-Top	ATGGGGGATCTTGACATTTTCTGGTATTGTTGTACCA ATCAAATTCCTGTAATT	9R promoter top strand
9R-Bott	AGGTAATTACAGGAATTTGATTGGTACAACAATACC AGAAAATGTCAAGATCCC	9R promoter bottom strand
9R-SM-Top	ATGGGGGATCTTGACATTTTCTGGGATCGTTGTACC AATCAAATTCCGGTAGTT	9R promoter with silent mutations, top strand
9R-SM-Bott	AGGTAACTACCGGAATTTGATTGGTACAACGATCCC AGAAAATGTCAAGATCCC	9R promoter with silent mutations, bottom strand
10F-Top	ATGGTAGGCAGATTGACTATGGCTCAGGCCACCAA AGTAGCATACACTTTGCAA	10F promoter top strand
10-Bott	AGGTTTGCAAAGTGTATGCTACTTTGGTGGCCTGAG CCATAGTCAATCTGCCTA	10F promoter bottom strand
10F-SM-Top	ATGGTAGGCAGACTAACGATGGCTCAGGCCACCAA AGTCGCTTACACTTTGCAA	10F promoter with silent mutations, top strand
10F-SM-Bott	AGGTTTGCAAAGTGTAAGCGACTTTGGTGGCCTGAG CCATCGTTAGTCTGCCTA	10F promoter with silent mutations, bottom strand
11R-Top	ATGGCTGAAGAATTTCAATATTTTCTGCCGTGTGATT CCGTGTATAATGACTTC	11R promoter top strand
11R-Bott	AGGTGAAGTCATTATACACGGAATCACACGGCAGA AAATATTGAAATTCTTCAG	11R promoter bottom strand
11R-SM-Top	ATGGCTGAAGAATTTCAATATTCTCCGCCGTGTGAT TCCGTGTGTAATGACTTC	11R promoter with silent mutations, top strand
11R-SM-Bott	AGGTGAAGTCATTACACACGGAATCACACGGCGGA GAATATTGAAATTCTTCAG	11R promoter with silent mutations, bottom strand
12F-Top	ATGGAGTTAGGCTATGAGGTGCACGCTGATGATGAT ACAATAGAACATTTCTTT	12F promoter top strand
12F-Bott	AGGTAAAGAAATGTTCTATTGTATCATCATCAGCGT GCACCTCATAGCCTAACT	12F promoter bottom strand
12F-SM-Top	ATGGAGTTAGGCTACGAGGTGCACGCTGATGATGAC ACCATAGAACATTTCTTT	12F promoter with silent mutations, top strand
12F-SM-Bott	AGGTAAAGAAATGTTCTATGGTGTCATCATCAGCGT GCACCTCGTAGCCTAACT	12F promoter with silent mutations, bottom strand

Supplementary Table 5 – Cloning PkMV into pLX-B2 backbone with silent mutations

Primer Name	Sequence 5' to 3'	Purpose
pLX-PkMV-F1-For	AAAATAAAACAAAACATCACAACATAACAC	PkMV specific primer to amplify first fragment
pLX-PkMV-F1-10F-SM-Rev	CTTTGGTGGCCTGAGCCATCGTTAGTCTG	PkMV specific primer to amplify first fragment with silent mutation in 10F
pLX-PkMV-F1-11R-SM-Rev	CTCCGCCGTGTGATTCCGTGTGTAA	PkMV specific primer to amplify first fragment with silent mutation in 11R
pLX-PkMV-F2-10F-SM-For	GCTCAGGCCACCAAAGTCGCTTACACT	PkMV specific primer to amplify second fragment with silent mutation in 10F
pLX-PkMV-F2-11R-SM-For	TACACACGGAATCACACGGCGGAGA	PkMV specific primer to amplify second fragment with silent mutation in 11R
pLX-PkMV-F2-Rev	TTTTTCCTCTTTTACCTCTGTACATGTGTT	PkMV specific primer to amplify second fragment
pLX-PkMV-Vec-For	AGAGGTAAAAGAGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCAGATCGTTCAAA	Linearizing pLX-B2-RFP vector with overlaps for second fragment
pLX-PkMV-Vec-Rev	GTGATGTTTTGTTTTATTTTCCTCTCCAAATGAAATGAACTTCCT	Linearizing pLX-B2-RFP vector with overlaps for first fragment