Transcriptional regulation of pokeweed antiviral protein

Camille A. Diaz

A thesis submitted to the Faculty of Graduate Studies

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Graduate Program in Biology
York University
Toronto, Ontario

September 2019

ABSTRACT

Phytolacca americana (American pokeweed) expresses pokeweed antiviral protein (PAP), a ribosome-inactivating protein (RIP) thought to function in defence. Several PAP isoforms are differentially responsive to jasmonic acid (JA), a phytohormone that mediates resistance against pathogen infections and herbivory. However, few RIP promoters have been characterized and little is known about RIP regulation in their native contexts. In this study, I identified the PAP promoters and discovered long introns in the 5' untranslated regions (5'UTRs) of PAP genes. qRT-PCR and reporter assays revealed that both the PAP-I promoter and intron can independently drive gene expression, leading to the transcription of mRNA variants with distinct 5'UTRs. In addition, a G-BOX promoter element was found to be required for the JA-mediated upregulation of PAP-I. Differential expression analysis confirmed that several PAP isoforms are responsive to other stresses, in addition to JA. This foundational work sheds light on the endogenous regulation of PAP.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor, Dr. Kathi Hudak, for taking me on as an undergrad all those years ago and allowing me to grow as a researcher. Thank you for always encouraging me to think critically and take ownership of my project while providing the most supportive environment a graduate student could hope for.

Next, I would like to thank my advisor Dr. Andy White for all your input and feedback on my project. I would also like to thank Dr. Sam Benchimol and Dr. David Hood for agreeing to be on my examination committee. Finally, I am grateful to members of the White lab and to Mohamed Salem from the Peng lab for training me on various lab equipment.

Over the years I have come to think of the lab as my second home (and on particularly hectic days, I've spent more hours there than in my actual home), largely because of the people in it: I would like to thank past and present members of the Hudak lab, especially Alex (for fixing everything and always being fearless), Jen (for sharing vegan food and getting lost with me on the most random topics), Kass (for being my first mentor), and Kira (for always lending an ear and not getting sick of me at home). As well, I would like to acknowledge the very fluffy Hudak lab hamsters (RIP) for providing comfort and amusement to us all.

Lastly, I would like to thank my family for their unwavering support throughout my studies. Thank you for always encouraging me to pursue my interests—whether it's art or science—even though I know I don't always make it clear to you what it is that I do.

STATEMENT OF CONTRIBUTIONS

All experiments and analyses were performed by Camille Diaz (CD), with the exception of those listed here. Pokeweed genome assembly, RNA-seq expression analysis (**Figure 10A; Figure 11B**), *in silico* PAP gene model generation (**Figure 9A**) and determination of pokeweed intron lengths (**Figure 17**) were performed by Kira Neller. Pokeweed genome walking and PAP-I 5' rapid amplification of cDNA ends (**Figure 5**) were performed by Alexander Klenov (AK). The MashUp reverse transcriptase enzyme used in **2.16** and **2.22** was made by AK. Illustrations were made by CD unless stated otherwise.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF COMMON ABBREVIATIONS

aa	Amino acid
ABA	Abscisic acid
bp	Base pair
BSA	Bovine serum albumin
CaMV	Cauliflower mosaic virus
Carb	Carbenicillin
cDNA	Complementary DNA
CDS	Coding sequence
CRE	Cis-acting regulatory element
CTAB	Cetyl trimethylammonium bromide
dH_2O	Distilled water
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
gDNA	Genomic DNA
GA	Gibberellic acid
GFP	Green fluorescent protein
GUS	β-glucuronidase
HcPro	Helper component protease
IME	Intron-mediated enhancement
JA	Jasmonic acid
Kan	Kanamycin
Kb	Kilobase
LB	Lysogeny Broth
4-MUG	4-methylumbelliferyl-β-D-glucuronide hydrate
nt	Nucleotide
ORF	Open reading frame
PAP	Pokeweed antiviral protein
PEG	Polyethylene glycol
PVP	Polyvinylpyrrolidone
qRT-PCR	Quantitative reverse transcription PCR
RIP	Ribosome-inactivating protein
RNAPII	RNA polymerase II
RT	Reverse transcription or reverse transcriptase
RT-PCR	Reverse transcription PCR
rRNA	Ribosomal RNA
SA	Salicylic acid
SDS	Sodium dodecyl sulphate
SOB	Super Optimal Broth
SLIC	Sequence- and ligation-independent cloning
TF	Transcription factor
TSS	Transcription start site
uORF	Upstream open reading frame
UTR	Untranslated region
X-Gluc	5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid, cyclohexylammonium salt
YEP	Yeast extract peptone

1. INTRODUCTION

1.1 *Phytolacca americana* and pokeweed antiviral protein

1.1.1 General characteristics

Phytolacca americana, commonly known as American pokeweed, is a flowering plant that belongs to the order Caryophyllales and the family Phytolaccaceae. Pokeweed has been studied primarily due to its synthesis of pokeweed antiviral protein (PAP), an N-glycosidase and ribosome-inactivating protein (RIP) that removes a specific adenine residue from the conserved α-sarcin/ricin loop of large ribosomal RNAs (rRNAs) through a process called depurination (Endo et al., 1988). Since damaged ribosomes cannot participate in translation elongation, depurination of rRNA causes an inhibition of global protein synthesis. PAP's antiviral activity is thought to stem from this mechanism of action, as it can limit the synthesis of viral proteins in infected cells and contain the infection (Gessner and Irvin, 1980). In addition to rRNA, PAP can also directly depurinate the genome of some plant and animal RNA viruses, interfering with stages of the viral life cycle (He et al., 2008; Karran and Hudak, 2008; Krivdova and Hudak, 2015; Mansouri et al., 2009). Transgenic plants expressing PAP exhibit resistance against viruses and fungi (Lodge et al., 1993; Zoubenko et al., 1997). Beyond pathogen resistance, pokeweed also shows promise in phytoremediation due to its ability to accumulate high levels of heavy metals that would otherwise be toxic to most plants (Dou et al., 2009; Peng et al., 2008). Despite these useful properties, a wealth of genetic information remains unknown in this non-model plant, as we have only very recently sequenced its genome (Neller et al., 2019).

1.1.2 PAP genes in pokeweed

RIPs are synthesized mainly by plants and are heavily represented in certain taxonomic lineages, including Phytolaccaceae (Di Maro et al., 2014). In pokeweed, several protein isoforms

of PAP have been reported. PAP proteins have been isolated from leaves in the spring (PAP-I, the most well-characterized isoform; Irvin, 1975), in early summer (PAP-II; Barbieri et al., 1982), and in late summer (PAP-III; Kurinov and Uckun, 2003); from seeds (PAP-S1 and S2; Honjo et al., 2002); from cell culture (PAP-C; Barbieri et al., 1989); and from roots (PAP-R and PAP-H; Bolognesi et al., 1990; Park et al., 2002). Through our leaf transcriptome analysis (Neller et al., 2016), we have also identified a novel PAP mRNA (Novel PAP) and demonstrated the expression of PAP-α, which had previously only been detected in genomic DNA (Kataoka et al., 1992). Through our genome annotation, we have confirmed the presence of the following PAP genes in pokeweed: PAP-I, PAP-II, PAP-S1, PAP-S2, PAP-α, Novel PAP, and PAP-H. We did not find genomic evidence for PAP-III, PAP-C, or PAP-R, and comparison with other isoforms revealed that PAP-III has the same amino acid sequence as PAP-II, while PAP-C and PAP-R have the same sequence as PAP-I. Therefore, small biochemical differences reported between these isoforms may have resulted from post-translational modifications, experimental variability, or allelic diversity (Neller et al., 2019).

Due to their cytotoxicity, RIPs usually localize to segregated cellular compartments; indeed, all PAP isoforms contain an N-terminal signal sequence that promotes secretion to the apoplast, the space between the plasma membrane and the cell wall (Ready et al., 1986; our unpublished data). PAP isoforms vary in their ability to depurinate RNA templates (Honjo et al., 2002; Kurinov and Uckun, 2003; Rajamohan et al., 1999) and show different levels of toxicity when expressed constitutively in heterologous systems (Dai et al., 2003; Wang et al., 1998), suggesting that they have different activities in pokeweed. Consistent with a predicted role in pathogen defence, we showed previously through transcriptome analysis that several PAP genes are upregulated in the presence of jasmonic acid (JA), a plant signalling hormone that mediates resistance to insect herbivores, which can act as viral vectors, and necrotrophic pathogens (Neller

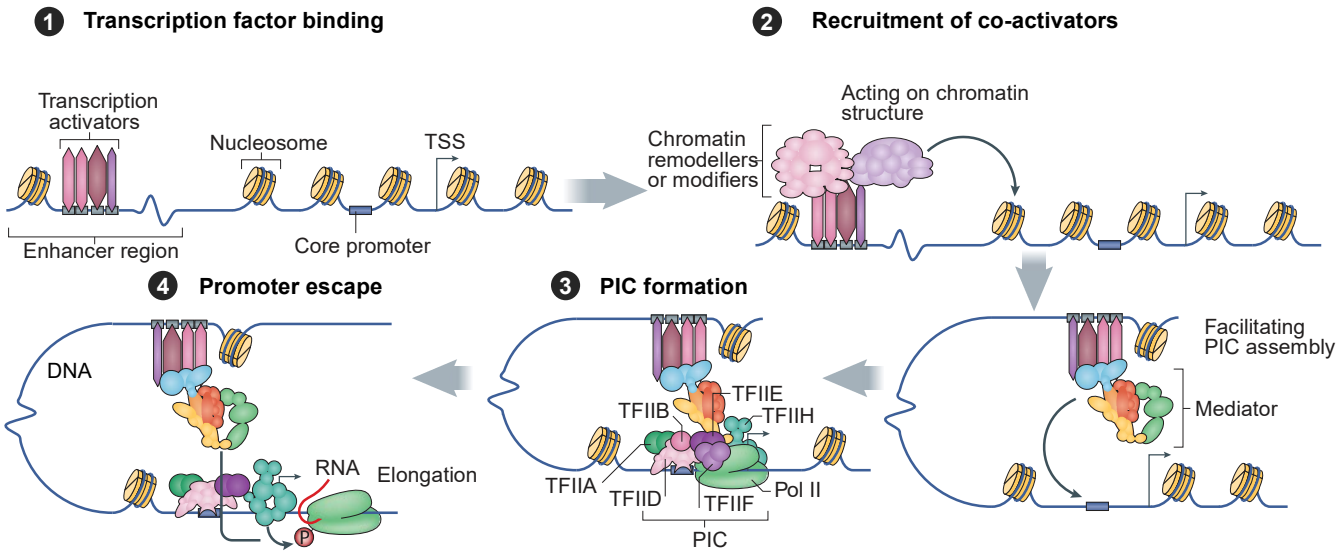
et al., 2016). It is well-known that RIPs can be induced by biotic and abiotic stresses; other RIPs have been found to be responsive to plant hormones (JA, salicyclic acid, abscisic acid) or associated stresses, including insect feeding, cold, heat, drought, high salinity, pathogen infection, and wounding (Chuang et al., 2014; Iglesias et al., 2005; Qin et al., 2009, 2010; Reinbothe et al., 2006; Song et al., 2000). However, little is known about how RIP expression is regulated endogenously, or where RIPs are situated within stress response pathways.

1.2 RNA polymerase II-mediated transcription in eukaryotes

1.2.1 Overview of transcription initiation

In eukaryotes, transcription of all protein-coding genes is governed by the RNA polymerase II (RNAPII) holoenzyme. Control of transcription requires a consortium of transcriptional activators and repressors (collectively known as specific transcription factors) that bind to their target sequences and modulate the activity of the basal transcriptional machinery, through both direct and indirect means (reviewed in Soutourina, 2018; Vernimmen and Bickmore, 2015). A simplified model of eukaryotic transcription initiation is illustrated in **Figure 1A**. Transcription activation starts with the binding of activators to enhancers and proximal promoter regions; these transcription factors (TFs) recruit co-activators that modify chromatin structure and increase accessibility of the region about to be transcribed (Vernimmen et al., 2007; Vierstra et al., 2014). This in turn facilitates the assembly of the basal transcriptional machinery at the core promoter, which is the region surrounding the transcription start site (TSS). The basal transcriptional machinery, or pre-initiation complex (PIC), is comprised of RNAPII and highly-conserved general transcription factors (TFIID, TFIIA, TFIIB, TFIIF, TFIIE, TFIIH/TFIIK) that are present in all eukaryotes. Mediator, a large multiprotein complex that interacts with both specific and general transcription factors, acts as a co-activator by stabilizing the PIC (Esnault et al., 2008; Eyboulet et al., 2015). Mediator also stimulates the phosphorylation of the RNAPII

A


1 Transcription factor binding
Transcription activators
Nucleosome
TSS
Enhancer region
Core promoter
2 Recruitment of co-activators
Acting on chromatin structure
Chromatin remodellers or modifiers
Facilitating PIC assembly
Mediator
3 PIC formation
TFIIE
TFIIB
TFIIH
TFIIA
TFIID
TFIIF
Pol II
PIC
4 Promoter escape
DNA
RNA
Elongation
P

B

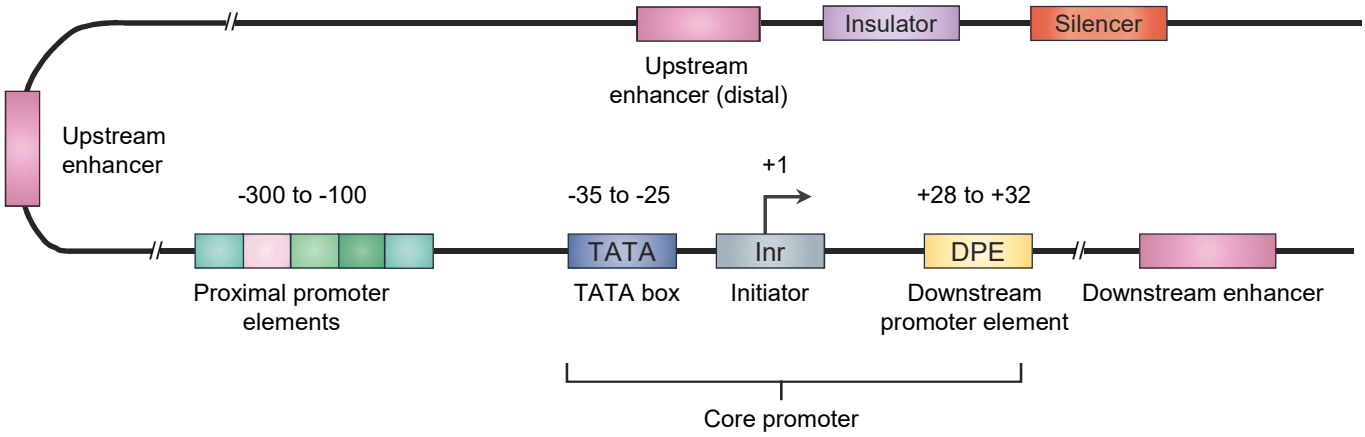

Insulator
Silencer
Upstream enhancer (distal)
Upstream enhancer
-300 to -100
-35 to -25
+1
+28 to +32
TATA
Inr
DPE
Proximal promoter elements
TATA box
Initiator
Downstream promoter element
Downstream enhancer
Core promoter

Figure 1. RNA polymerase II-mediated transcription in eukaryotes. (A) A simplified model of transcription initiation in eukaryotes. **(1)** Transcriptional activators bind to enhancer regions, which can be located upstream or downstream of core promoters. **(2)** Activators recruit chromatin remodelling complexes that increase accessibility of the region about to be transcribed. **(3)** Other transcriptional activators and co-activators facilitate the assembly of the basal transcriptional machinery, or the pre-initiation complex (PIC). The PIC is assembled at the core promoter. It includes RNA polymerase II (RNAPII) and general transcription factors (in order of binding: TFIID, TFIIA, TFIIB, TFIIF, TFIIE, TFIIH/TFIIK, which contains a kinase module known as cyclin-dependent kinase). Mediator of RNA polymerase II transcription (Mediator), a large, multiprotein complex that interacts with both specific and general transcription factors, acts as a co-activator by stabilizing the PIC. **(4)** Cyclin-dependent kinase (CDK) phosphorylates an RNAPII subunit, which is necessary for RNAPII to escape from the promoter and begin transcription elongation. Figure replicated from Soutourina (2018). **(B)** A schematic diagram of common cis-acting regulatory elements in plant promoters. The core promoter region comprises of binding sites for general transcription factors, while proximal promoter elements bind specific transcription factors. Enhancers and silencers, which can be found thousands of kilobases away from their target genes, also bind specific transcription factors. Insulators prevent distal enhancers from acting on the promoters of neighbouring genes.

carboxy-terminal domain by a cyclin-dependent kinase (CDK), allowing RNAPII to be released from the promoter and begin transcription elongation (Kim et al., 1994). A recent study suggests that transcription in plants is mostly regulated at the initiation step, unlike in animals, where the rate of mRNA synthesis is further controlled during elongation through RNA polymerase pausing (Core et al., 2008; Hetzel et al., 2016).

1.2.2 Cis-acting regulatory elements (CREs) in plant promoters

Cis-acting regulatory elements (CREs) are short sequence motifs that bind general and specific TFs during transcription initiation. The positions of common CREs in plant promoters are depicted in **Figure 1B**. The core promoter, which typically extends from -40 to +40 bps around the TSS, contains binding sites for general TFs: both the TATA box and Initiator (Inr) element bind subunits of the TFIID complex, the TF that nucleates PIC assembly. Other motifs may also be present, including the B recognition element (BRE), which binds TFIIB, and the CAAT box, which binds core binding factors (Kumari and Ware, 2013). Initiation may still occur in the absence of some of these motifs; for instance, only 29% of *Arabidopsis* core promoters have been reported to contain TATA elements (Molina and Grotewold, 2005). The downstream promoter element (DPE) is also thought to function cooperatively with Inr to bind TFIID, particularly for TATA-less promoters (Kutach and Kadonaga, 2002).

The proximal promoter is found around 100 to 300 bps upstream of the TSS and contains multiple binding sites for specific TFs (proximal promoter elements). It is in this region that precise control of transcription occurs, as different regulatory elements may bind to a diverse array of plant TFs that are activated only in certain cell and tissue types, developmental stages, and stimuli (reviewed in Yamasaki et al., 2013). Enhancer and silencer elements are similar to proximal promoter elements in that they are cis-acting sequences that are bound by specific TFs; however, they are able to regulate gene expression from a greater distance, and can be located

many kilobases upstream or downstream of their target genes (reviewed in Soutourina, 2018; Vernimmen and Bickmore, 2015). While the onset of high-throughput enhancer screening methods has led to the genome-wide identification of enhancer elements in model mammalian systems (Korkmaz et al., 2016; Vockley et al., 2016), the mapping of distal regulatory elements remains challenging, particularly for organisms that do not have well-assembled genomes. Therefore, classical promoter studies focus on the 1 to 2 Kb region immediately upstream of the TSS, which encompasses the core and proximal promoter elements, as well as any nearby upstream enhancers or silencers.

The high sequence conservation of TF binding motifs means that CREs can be predicted computationally in uncharacterized promoters based on experimentally validated CREs from other plants (Chow et al., 2016; Higo et al., 1999). Through this approach, the general function of a particular gene may be elucidated; for example, a plant promoter that contains motifs known to bind light-responsive TFs may control a gene involved in photosynthesis (López-Ochoa et al., 2007). Since not all predicted motifs may be functional, *in silico* CRE prediction is best paired with wet-lab promoter and TF binding analyses (reviewed in Geertz and Maerkl, 2010; Hernandez-Garcia and Finer, 2014).

1.3 Promoter studies in plants

1.3.1 General overview

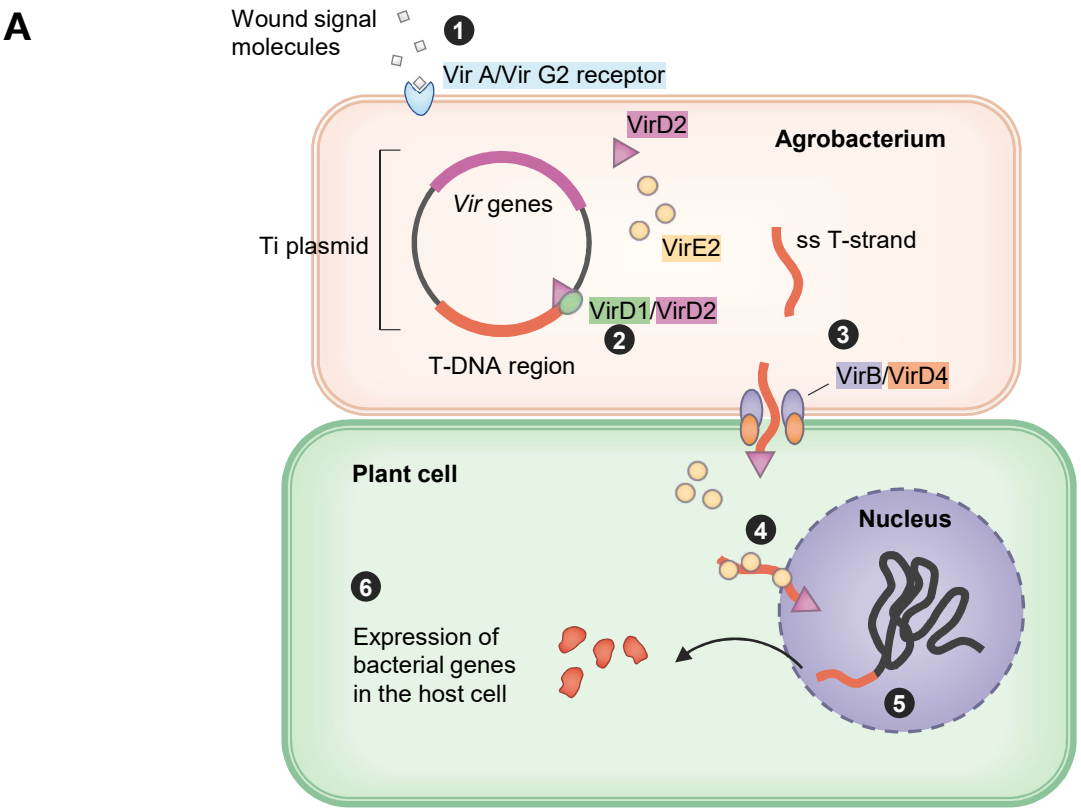
Promoter studies are conducted by placing promoter sequences upstream of reporter genes, expressing the constructs in cells or in whole plants, and assaying for reporter activity as a measure of promoter strength (reviewed in Hernandez-Garcia and Finer, 2014). In plants, the most commonly used reporters are green fluorescent protein (GFP) and related variants (Niedz et al., 1995), luciferase (Ow et al., 1986), and β-glucuronidase (Jefferson et al., 1987). In

particular, the bacterial-derived β-glucuronidase (GUS) has found widespread usage due to its versatility; depending on the type of GUS substrate used, GUS can be detected through histochemical staining, providing a visual indication of reporter expression in intact plant tissue, or through a fluorometric assay, providing a quantitative measurement of promoter strength.

1.3.2 Agrobacterium-mediated transformation

While foreign genes can be expressed in whole plants through biolistic particle bombardment (Kikkert et al., 2005) or rub inoculation of viral vectors (Scholthof, 1999; Seo et al., 2009), the most common and efficient method is through Agrobacterium-mediated transformation (reviewed in Păcurar et al., 2011). *Rhizobium radiobacter* (syn. Agrobacterium) is a gram-negative soil bacterium that causes gall tumours in plants by transferring its own genetic material into the genome of its host (Chilton et al., 1977). Virulent Agrobacterium strains contain a tumour-inducing (Ti) plasmid that encodes the virulence (*vir*) genes required for this lateral genetic transfer, as well as the genes needed for plant tumour formation, which are contained in the T-DNA region (**Figure 2A**). The T-DNA enters plant cells as a single-stranded DNA sequence (T-strand) and integrates randomly into the host genome by taking advantage of naturally-occuring double-stranded DNA breaks (Kim et al., 2007). Upon genome integration, genes encoded by the T-strand direct the synthesis of amino acids and plant growth hormones, resulting in the tumorous proliferation of cells in infected areas (Garfinkel et al., 1981; Zhu et al., 2000).

By removing the native genes in the T-DNA region and replacing them with other sequences, such as reporter genes and plant selectable markers, Agrobacterium can integrate foreign genes into host plant genomes in an asymptomatic manner. The ability of vir proteins to act in trans led to the development of binary transformation vectors (**Figure 2B**), in which the *vir*


A
Wound signal molecules
1
Vir A/Vir G2 receptor
VirD2
Agrobacterium
Vir genes
Ti plasmid
VirE2
ss T-strand
VirD1/VirD2
2
3
T-DNA region
VirB/VirD4
Plant cell
Nucleus
4
6
Expression of bacterial genes in the host cell
5

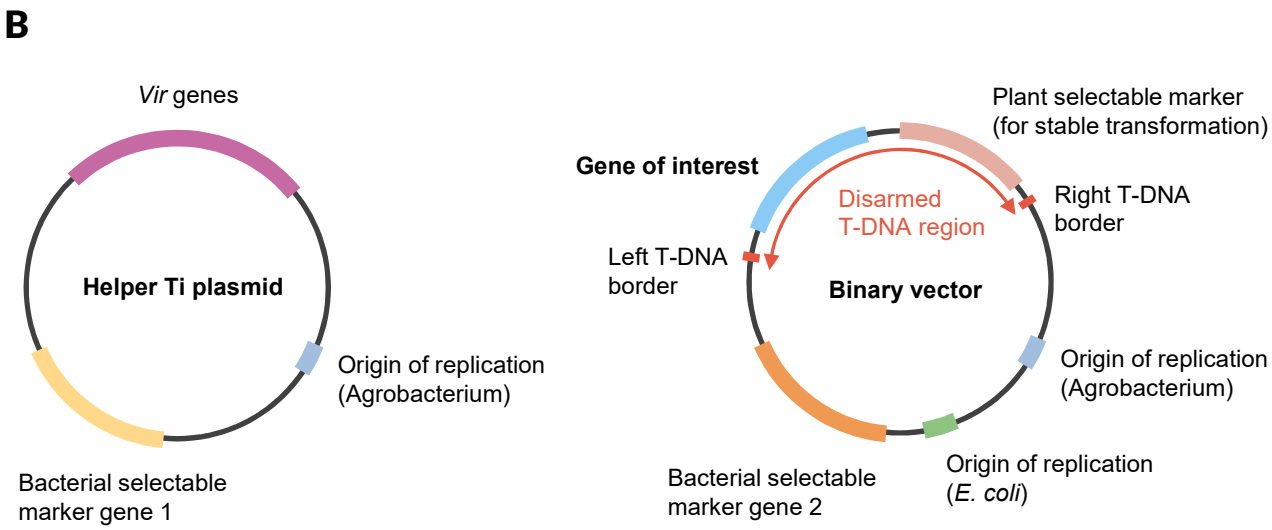

B
Vir genes
Helper Ti plasmid
Origin of replication (Agrobacterium)
Bacterial selectable marker gene 1
Plant selectable marker (for stable transformation)
Gene of interest
Disarmed T-DNA region
Right T-DNA border
Left T-DNA border
Binary vector
Origin of replication (Agrobacterium)
Bacterial selectable marker gene 2
Origin of replication (E. coli)

Figure 2. Agrobacterium-mediated transformation in plants. **(A)** A simplified model of Agrobacterium-mediated transformation. **(1)** Wounded plants release signal molecules that are recognized by bacterial Vir receptors, leading to the activation of *vir* genes on the Ti plasmid of Agrobacterium. **(2)** VirD1/VirD2 endonucleases process the T-DNA region in a sequence-specific manner and release a single-stranded (ss) T-strand. **(3)** VirD2 attaches covalently to the 5' end of the T-strand and the DNA-protein complex is transported into the plant cytoplasm via a bacterial secretion system (VirB/VirD4). Other Vir proteins involved in the plant transformation process, such as VirE2, are exported to the host cell using the same route. **(4)** VirE2 associates with the VirD2-conjugated T-strand and the entire complex localizes to the nucleus. **(5)** The T-strand integrates randomly into the plant genome, leading to the expression of bacterial genes in the host **(6)**. Genes encoded by the T-strand direct the synthesis of amino acids and plant growth hormones, resulting in the tumorous proliferation of cells. **(B)** The binary vector system used in genetic engineering. *Vir* genes are placed in a helper Ti plasmid and act in *trans*, while the disarmed T-DNA region is located in a second binary vector. Bacterial oncogenes have been removed and replaced with genes of interest, such as reporter genes and plant selectable markers; the left and right T-DNA borders allow the region to still be recognized by Vir proteins and processed as wild-type T-DNA. The binary vector can be propagated in both Agrobacterium and *E. coli*, allowing researchers to insert foreign genes with conventional molecular cloning techniques. The figure was illustrated based on information from Păcurar et al. (2011).

genes are placed in a helper Ti plasmid, and the disarmed T-DNA region is placed in a second plasmid (binary vector) that can be propagated in both *E. coli* and Agrobacterium (reviewed in Lee and Gelvin, 2008). This strategy allows researchers to insert foreign genes into the T-DNA region using standard molecular cloning procedures. Foreign genes can be expressed transiently by agroinfiltrating plant tissue with the desired constructs and measuring gene expression after a few days, or stably, by regenerating whole plants from agroinfiltrated tissue and screening for transgenic lines that have successfully integrated the gene of interest (reviewed in Hwang et al., 2017; Krenek et al., 2015). For plant promoter studies, stable transformation is desirable because it allows for the tracking of reporter expression across different developmental stages, tissue types, and stimuli. However, generating stable lines can be time consuming as it is heavily dependent upon the availability of optimized, species-specific plant regeneration techniques, as well as plant generation times.

1.4 Plant hormone-mediated regulation of stress responses

In plants, stress response pathways are regulated primarily by plant hormones (also known as phytohormones), which are signalling molecules produced within cells. Unlike in animals, in which hormone production is restricted to specialized glands (reviewed in Brown-Borg, 2007), each plant cell is capable of synthesizing phytohormones (reviewed in Santner et al., 2009; Verma et al., 2016). **Figure 3** summarizes several stress-associated hormone signalling pathways and their points of crosstalk.

1.4.1 Jasmonic acid (JA)

Jasmonic acid (JA) is a lipid-based phytohormone that plays a major role in defence against insect herbivores and necrotrophic pathogens (reviewed in Wasternack and Hause, 2013). The JA signalling pathway is divided into two mutually antagonistic branches regulated by

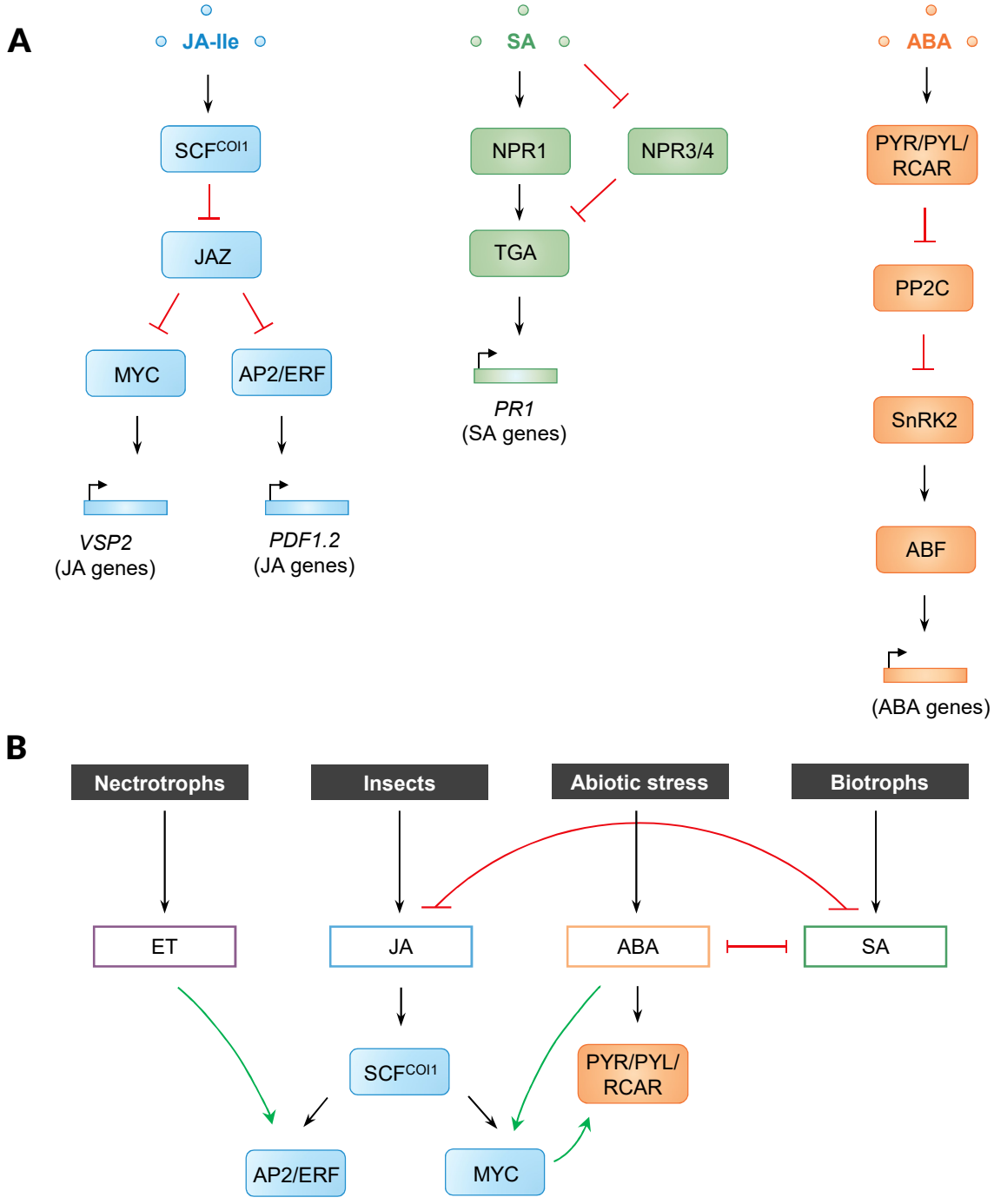


Figure 3. Plant hormone-mediated regulation of stress responses. (A) Summary of individual pathways for jasmonic acid (JA), salicylic acid (SA), and abscisic acid (ABA). **(B)** Key nodes of crosstalk between stress-associated signalling pathways. ET = ethylene. Synergistic interactions are indicated by a green arrow and inhibitory interactions are indicated by a ticked red line. Mutually antagonistic pathways are indicated by a double-ticked red line.

MYC and AP2/ERF TFs: the MYC branch is associated with the wound response and defence against insect herbivores, which can act as viral vectors, while the AP2/ERF branch is associated with resistance to necrotrophic pathogens (Lorenzo et al., 2003, 2004). The AP2/ERF branch requires activation of both the JA and ethylene (ET) pathways (Zhu et al., 2011), and branch signalling specificity is further achieved through interactions with different transcriptional coregulators (reviewed in Pieterse et al., 2011).

The biologically active form required for signalling is a jasmonate-isoleucine conjugate (+)-7-iso-jasmonoyl-L-isoleucine (JA-Ile) that is recognized by the SCF^{COI1}-JAZ receptor complex (Fonseca et al., 2009), where coronatine-insensitive 1 (COI1) is an F-box protein that functions as a ubiquitin ligase (Xu et al., 2002), and jasmonate-ZIM domain (JAZ) is a family of transcriptional repressors (Thines et al., 2007). In the absence of JA-Ile, JAZ proteins repress transcription from JA-associated genes through its direct interactions with the transcription factors MYC and AP2/ERF, which act as master JA signalling activators (Ng et al., 2018). Activation of the SCF^{COI1}-JAZ receptor complex through increased levels of JA-Ile directs the ubiquitination and subsequent degradation of JAZ (Thines et al., 2007), allowing MYC and AP2/ERF to promote the transcription of their downstream target genes, including the marker genes *VSP2* and *PDF1.2* (Lorenzo et al., 2004; Manners et al., 1998).

1.4.2 Salicylic acid (SA)

Salicylic acid (SA) is a phenolic phytohormone that is important for defence against microbial biotrophic pathogens. It is also required for systemic acquired resistance (SAR), a broad-spectrum defence response analogous to the innate immune system in animals. Despite their shared role in pathogen defence, JA and SA signalling responses generally act antagonistically, and this mutual antagonism can occur at multiple points in the two pathways (Kazan and Manners, 2013; Spoel et al., 2003; Yuan et al., 2007; Zander et al., 2010). Since

activating stress-associated pathways can be costly, this JA-SA crosstalk is thought to equip the plant with a more fine-tuned immune response (reviewed in Pieterse et al., 2011).

SA is perceived through receptors belonging to the NPR (non-expressor of pathogenesis-related genes) family, and it was only recently elucidated that NPR1 and NPR3/NPR4 play opposite roles in the transcriptional regulation of SA-associated genes (Ding et al., 2018). At low levels of SA, NPR1 exists in the cytoplasm as an oligomer linked through intermolecular disulphide bonds. When SA levels increase in response to pathogen infection, redox changes in the cytoplasm lead to the monomerization and subsequent nuclear translocation of NPR1 (Cheng et al., 2009; Tada et al., 2008). In the nucleus, NPR1 can act as a transcriptional coactivator and allow TGA TFs to activate transcription of their target genes, including the SA marker gene *PR1* (Rochon et al., 2006; Zhang et al., 2002). In contrast, NPR3/NPR4 are transcriptional co-repressors that bind TGA TFs in the absence of SA. The presence of SA abolishes this interaction and allows NPR1 to bind to TGA TFs (Ding et al., 2018).

1.4.3 Abscisic acid (ABA)

Abscisic acid (ABA) mediates responses to abiotic stresses such as drought, cold, heat, and high salinity mainly by regulating stomatal aperture, allowing plants to control their level of water retention in a dynamic environment (reviewed in Daszkowska-Golec and Szarejko, 2013). ABA and SA response mechanisms act antagonistically at multiple steps, suggesting that plants balance ABA-mediated abiotic stress tolerance and SA-mediated biotic stress resistance through ABA-SA crosstalk (Yasuda et al., 2008). When produced in combination with JA, ABA acts synergistically with JA to activate the MYC branch of the JA response pathway, while antagonizing the AP2/ERF branch (Abe et al., 2003). This results in increased resistance to herbivory at the expense of reduced resistance to necrotrophic pathogens (Anderson et al., 2004;

Fernández-Calvo et al., 2011). Conversely, JA signalling can induce genes that encode ABA receptors (Lackman et al., 2011).

ABA is perceived by receptor proteins in the PYR/PYL/RCAR (Pyrabactin Resistance/Pyrabactin resistance–Like/Regulatory Components of ABA Receptor) family, which are found in both the nuclear and plasma membranes (Rodriguez et al., 2014). In the absence of ABA, protein phosphatases type 2C (PP2C) inhibit SNF1-related protein kinases (SnRK2s) through dephosphorylation, preventing them from activating downstream TFs (Umezawa et al., 2009). In the presence of ABA, PYR/PYL/RCARs bind to and inhibit PP2Cs, allowing the accumulation of active SnRK2s. This enables SnRK2 to phosphorylate ABA-Responsive Element-Binding Factor (ABF) transcriptional activators, thereby promoting the transcription of ABA-associated genes (Furihata et al., 2006).

1.5 Role of leader introns in gene regulation

Introns are a common feature of eukaryotic genes and are found ubiquitously in both coding sequences (CDSs) and untranslated regions (UTRs), although they are more prevalent in CDSs (Hong et al., 2006). Unlike in mammals, where introns frequently span several kilobases (Keane and Seoighe, 2016), introns in flowering plants are generally no more than several hundred nucleotides long. For instance, *Arabidopsis* has a mean intron length of 168 bps (Kaul et al., 2000), while maize (*Zea mays*) has a mean intron length of 516 bps (Schnable et al., 2009). Introns in the 5'UTR, also known as leader introns, exhibit properties that distinguish them from introns in other regions. In *Arabidopsis*, leader introns tend to be longer and have a different nucleotide composition compared to CDS or 3'UTR introns (Chung et al., 2006). Moreover, numerous leader introns have been shown to enhance gene expression in plants and other eukaryotes (reviewed in Gallegos and Rose, 2015; Laxa, 2017). In some instances, leader introns are required for high-level expression; for example, Jeong et al. (2006) found that on its own, the

promoter of the *Arabidopsis* profilin gene *PRF2* could only drive vascular expression, and the *PRF2* leader intron was needed for constitutive expression. Similarly, the leader intron of the *Arabidopsis* metal transporter gene *AtMHX* enhanced reporter expression 272-fold when paired with its weak native promoter (Akua et al., 2010).

Since studies have only been conducted on individual genes, the overall frequency of regulatory leader introns in eukaryotes is unknown. Large-scale studies on regulatory leader introns are further complicated by the fact that introns can influence gene expression through several distinct mechanisms, and so there are presently no clear, conserved sequence motifs that will lead to their mass identification. Leader introns may function simply as classical downstream enhancers by harbouring CREs (Kim et al., 2006), or act as alternative promoters, leading to the transcription of mRNAs that differ only in their 5'UTRs (Morello et al., 2002; Qi et al., 2007). A third mechanism known as intron-mediated enhancement (IME) has also been proposed. The precise mechanism for IME remains unclear, although models at different levels of regulation have been suggested. Leader introns may enhance transcription by creating a local, favourable zone for transcription initiation that is depleted of nucleosomes (Gallegos and Rose, 2015). Additionally, splicing markers in leader introns may promote transcription reinitiation, affect mRNA processing and stability, and increase translation efficiency (reviewed in Shaul, 2017).

1.6 Research objectives

It is well-established that RIPs can be induced by stresses (reviewed in Zhu et al., 2018), and several PAP isoforms have been shown to be responsive to JA (Neller et al., 2016). However, little is known about the endogenous regulation of RIP expression and few RIP promoters have been characterized. Previous attempts to isolate the PAP-I promoter in our lab had been unsuccessful, and the 5'UTRs of PAP genes had not been clearly delineated. Therefore, the goal of this study was to identify and functionally characterize PAP gene promoters using both *in*

silico methods and reporter gene assays. Emphasis was placed on PAP-I, the most well-studied and highly expressed isoform; more specifically, I aimed to identify cis-acting regulatory elements within the PAP-I promoter that are responsible for the JA-mediated upregulation of the PAP-I transcript. To complement promoter analyses, pokeweed plants were subjected to four stress treatments (JA, SA, mechanical wounding, and PEG, which simulates drought) and the expression profiles of PAP transcripts present in the leaf were determined through RNA-seq and qRT-PCR. This study also led to the discovery of a novel, conserved feature in PAP genes: a long leader intron. The PAP-I leader intron's potential role in influencing gene expression and the mechanism by which it acts was explored in parallel. This foundational work sheds light into the distinct roles of the PAP isoforms in the plant and provides novel insight into the endogenous regulation of ribosome-inactivating proteins.

2. MATERIALS AND METHODS

2.1 Plant cultivation

Seeds – *Nicotiana tabacum* seeds (cv. 'Samsun') were provided by the Canadian Tobacco Research Foundation (Delhi, ON). *Phytolacca americana* seeds originated in New Jersey, USA, and plants have been grown and maintained at York University, Toronto, ON, Canada since 2004.

Sowing – Tobacco seeds were sown in trays containing rich soil mixture (1 part garden soil: 1 part Pro-Mix all-purpose soil mix: 0.5 part cattle manure: 0.5 part sand) and seedlings were transplanted to individual 2.5-inch square pots 1.5 weeks after germination. Plants were covered with clear plastic domes until the two-leaf stage. To promote germination, pokeweed seeds were gently swirled in absolute sulphuric acid for 4 minutes and rinsed under running tap water for 15 minutes. Seeds were kept submerged in tap water for 2-4 days until ≥ 95% of seed coats were lightly cracked. Seeds were sown in individual 2.5-inch square pots containing rich soil mixture. Trays containing pots were covered with clear plastic domes and placed on heated electric mats until cotyledons had emerged fully.

Growth conditions – All tobacco and pokeweed plants were kept in growth chambers at 24°C under 14 hours of light (9:30 AM to 11:30 PM) and 10 hours of darkness. Chamber lighting was comprised of 75% fluorescent and 25% incandescent bulbs (180 $\mu E/m^2/s$) and fan speed was set to 65%. Plants were watered approximately every 3 days and fertilized weekly with a 2-1-6 liquid fertilizer (Advanced Nutrients).

2.2 Isolation of pokeweed genomic DNA

Genomic DNA (gDNA) was isolated from pokeweed leaves using a CTAB-based extraction method. Leaf tissue (2 g) was homogenized in a prewarmed mortar (60°C) with a pestle and 20 mL of prewarmed (60°C) CTAB buffer (1.4 M NaCl, 100 mM Tris-HCl, pH 8.0, 20 mM EDTA, 2% CTAB, 0.2% β-mercaptoethanol, 2% PVP). The slurry was transferred into a 50 mL conical polypropylene tube and incubated at 60°C for 1.5 hours. DNA was extracted twice with equal volume of chloroform:isoamyl alcohol (24:1) and centrifuged at 3275 x g in a Beckman-Coulter Allegra X-12 swinging bucket centrifuge. The upper aqueous phase was transferred into a new 50 mL tube each time. DNA was precipitated by adding 0.7 volume cold isopropanol to the upper aqueous phase and centrifuging at 3275 x g for 30 minutes. Following centrifugation, the supernatant was discarded and the resulting pellet was washed twice in 70% ethanol. The DNA pellet was then resuspended in 500 µL dH_2O and treated with 10 µL of 10 mg/mL RNAse A for 1 hour at 37°C. DNA was extracted twice with equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, pH 8.0) and precipitated with the addition of 2.5 volumes 100% ethanol and 0.10 volume of 3 M NaOAc. DNA was pelleted by centrifuging at 14 000 x g for 20 minutes. The pellet was washed twice with 200 µL of 75% ethanol and resuspended in 200-500 µL of dH_2O, depending on pellet size. DNA concentration was measured using a spectrophotometer (A_{260} nm) and gDNA quality was evaluated by separating 2 µg of sample on a 0.7% agarose gel at 110 V for 30 minutes.

2.3 Polymerase chain reaction (PCR)

PCR was used for cloning and validation of computational PAP gene models. Depending on the experiment, the template was either genomic DNA (250 ng), plasmid DNA (5-15 ng), or complementary DNA (cDNA) directly from a reverse transcription reaction (0.5-1 µL). Each reaction had the following components: DNA template, 5 µL of 5X Q5 Reaction Buffer, 0.5 µL of

10 mM dNTPs, 1.25 µL of 10 µM forward primer, 1.25 µL of 10 µM reverse primer, 0.5 unit of Q5 High-Fidelity DNA Polymerase (New England Biolabs #M0491S), 5 µL of 5X Q5 GC Enhancer Buffer (for templates with high-GC content and/or secondary structure) and dH_2O to a final volume of 25 µL. PCR amplification was initiated by denaturation at 95°C for 90 seconds and was followed by 35 cycles alternating between denaturation (95°C; 30 seconds), annealing (61-70°C on an Eppendorf gradient thermocycler; 30 seconds), and extension (72°C; 45 s per Kb of DNA). A final extension was carried out at 72 °C (5 minutes for PCR products ≤ 4 Kb; 10 minutes for longer PCR products). To check for correct amplification, PCR products (8 µL) were separated on an agarose gel (1-2.0%, depending on insert size) at 110 V for 30 minutes.

2.4 Generation of PAP promoter-reporter gene constructs

A schematic diagram for generating promoter-reporter gene constructs is illustrated in **Figure 4**. Primers used for cloning are listed in **Table 2** (Appendix B). The 1262-bp region upstream of the PAP-I transcription start site (TSS) was considered the PAP-I promoter. This 1.3 Kb PAP-I promoter, along with the 5'UTR amd 1.6 Kb PAP-I intron, were PCR amplified from gDNA with the primers PAP-I-prom-SLIC-FOR and PAP-I-prom-SLIC-REV, gel purified (see **2.5**), and cloned by one-step SLIC (see **2.6**) into the pUC-based vector pHSG299. The pHSG299 plasmid containing the PAP-I promoter and intron then served as the PCR template for all downstream PCRs of PAP-I promoter fragments. The promoter was serially truncated from the 5'end in approximately 250 bp increments to give rise to several truncation fragments: 1124 bp, 711 bp, 584 bp, 432 bp, 296 bp, and 102 bp. Promoter-intron fragments were generated through PCR as described in **2.3** by pairing the same reverse primer with different forward primers spanning the promoter.

To produce intronless versions of promoter constructs, a reverse primer (5-UTR-no-int-REV) was designed to connect the two portions of the PAP-I 5'UTR that were originally interrupted by the intron. This reverse primer was paired with different forward primers used above. PCR products were gel purified as described in **2.5** and used as templates for a second round of PCR, this time with forward and reverse primers that contained additional sequences for cloning. Other PAP promoter-reporter constructs were generated in the same manner as the intronless PAP-I promoter-reporter constructs. Each promoter fragment was then cloned into a reporter vector (pCambia 0305.2 GFP-GUS) using either one-step SLIC (see **2.6**) or Gibson assembly (see **2.7**). These methods were chosen over traditional restriction enzyme cloning because they are sequence independent, and PCR fragments could be inserted to the vectors without the addition of restriction enzyme sites. Therefore, the PAP-I promoter and 5'UTR sequences were inserted into the reporter vector exactly as they exist in pokeweed. Since SLIC and Gibson assembly required the same input PCR fragments, the two methods were used simultaneously and putative constructs were taken from reactions that yielded the highest cloning efficiency. Linear vectors were also made through PCR as described in **2.3** (Primer pair: pCambia-prom-FOR and pCambia-prom-REV).

To produce translocated intron-promoter constructs, linear 1262 and 102 PAP-I promoter vectors were generated through PCR (Primer pairs: P1-transl-FOR and pCambia-prom-REV; P7-transl-FOR and pCambia-prom-REV) and the PAP-I intron was inserted upstream of each fragment using Gibson assembly. Minus catalase intron constructs were made by excluding the intron through PCR (Primer pairs: pCambia-1-no-cat-FOR and pCambia-1-no-cat-REV; pCambia-2-no-cat-FOR and pCambia-2-no-cat-REV) and reassembling the two fragments through Gibson assembly. T/GBOXATPIN2 was deleted in the 1262/-GBOX construct by excluding the element through PCR and reassembling the two fragments through Gibson

assembly (Primer pairs: pCambia-GBOX-mut-1-FOR and pCambia-GBOX-mut-1-REV; Primer pairs: pCambia-GBOX-mut-2-FOR and pCambia-GBOX-mut-2-REV).

All plasmids were sent to Bio Basic for sequence confirmation using the custom primers listed in **Table 3** (Appendix B).

2.5 Low-melt agarose gel extraction

Gel-purified DNA was used to generate all constructs. First, DNA fragments were separated at 70 V for 45 minutes on a low-melt agarose gel (0.7-2.0% depending on fragment size) precast with ethidium bromide. After electrophoresis, the gel was placed on a UV illuminator and the band of the correct size was excised. DNA was extracted from the gel slice with the EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic #BS353) following the manufacturer's protocol. DNA concentration was measured using a spectrophotometer (A_{260} nm).

2.6 One-step sequence- and ligation-independent cloning (SLIC)

One-step SLIC was performed according to Jeong et al. (2012) with minor modifications. In a 0.2 mL microtube, gel-purified insert and vector fragments were combined at a 1:4 insert to vector molar ratio (vector molar amount = 16-40 fmol, depending on availability) with 1 µL of 10X NEBuffer 2.1 (500 mM NaCl, 100 mM Tris-HCl, pH 7.9, 100 mM $MgCl_2$, 10 mg/ml BSA), 1.5 units of T4 DNA polymerase (New England Biolabs #M0203S) and dH_2O to a final volume of 10 µL. The reaction was incubated at room temperature for 2.5 minutes and stopped by incubating on ice for 10 minutes. SLIC reaction products (5 µL) were transformed into chemically competent DH5α *E. coli* cells as described in **2.8**.

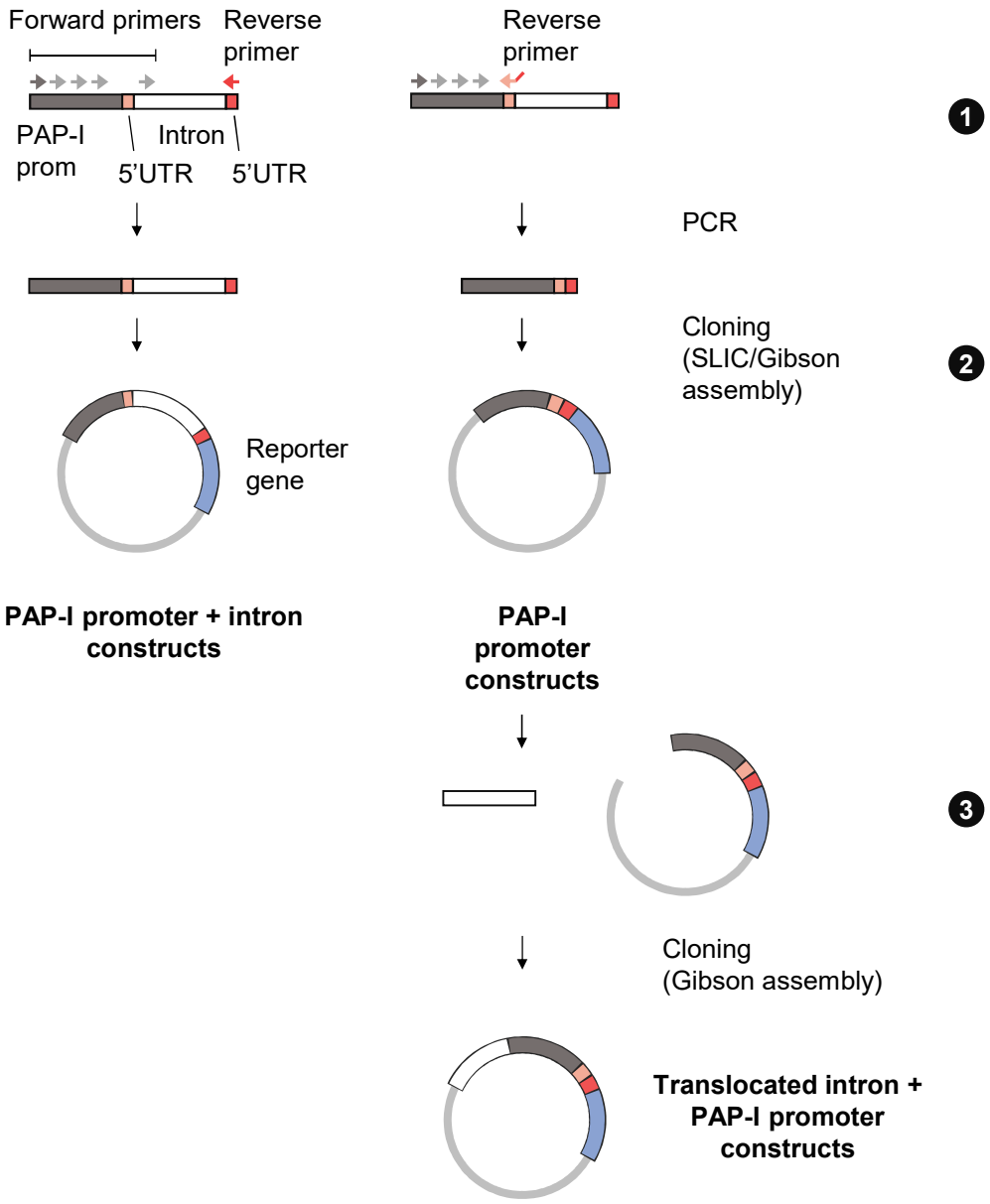


Figure 4. Generation of PAP-I promoter-reporter gene constructs. (1) PAP-I promoter-intron fragments were generated through PCR by pairing the same reverse primer (red arrow) with different forward primers (black and grey arrows) spanning the promoter. To produce intronless versions of promoter constructs, a reverse primer connecting the two portions of the 5'UTR (fused pink and red arrow) was paired with the same forward primers described above. **(2)** PCR fragments were cloned into a GFP-GUS reporter gene vector using one-step SLIC or Gibson assembly, which are sequence-independent cloning strategies. **(3)** To produce translocated intron-promoter constructs, the PAP-I intron was cloned upstream of the PAP-I promoter using Gibson assembly.

2.7 Gibson assembly

In a 0.2 mL microtube, gel-purified insert and vector fragments were combined at a 1:3 insert to vector molar ratio (vector molar amount = 16-40 fmol, depending on availability) with 6 µL of 2X Gibson Master Mix (New England Biolabs #E2611, or a homemade mix with similar components; see Appendix B) and dH_2O to a final volume of 12 µL. The reaction was incubated at 50°C for 45 minutes, followed by enzyme heat inactivation at 70°C for 10 minutes. Gibson reaction products (2-5 µL) were transformed into chemically competent DH5α *E. coli* cells as described in **2.8**.

2.8 *E. coli* plasmid transformation

Plasmid DNA (50-100 ng purified plasmid, or 2-5 µL from the SLIC/Gibson reaction) was added to 120 µL of thawed, chemically competent DH5α *E. coli* cells and the mixture was incubated on ice for 30 minutes. Cells were heat shocked at 42°C for 30 seconds and placed again on ice for 2 minutes. Next, cells were mixed with 900 µL of Super Optimal Broth (SOB) and incubated at 37°C for 1 hour with shaking at 250 rpm. Following incubation, cells were pelleted by centrifuging at 14 000 x g for 60 seconds. The top 800 µL supernatant was discarded and cells were resuspended in the remaining medium. Resuspended bacteria were plated on selective LB agar plates and incubated overnight at 37°C.

2.9 Colony PCR

Colony PCR was used to screen individual colonies for the correct insert size following *E. coli* transformation. First, single *E. coli* colonies were picked from a freshly grown selective LB plate and resuspended in separate 0.2 mL microtubes containing 40 µL of dH_2O. Depending on the number of colonies on each plate, 8-20 unique colonies were screened per construct. To maintain unique colonies, 2 µL of resuspended cells were streaked on a new selective LB agar

plate and incubated overnight at 37°C. The remaining 38 µL of cells were incubated at 95°C for 10 minutes to promote lysis. Lysates were snap cooled on ice for 2 minutes and centrifuged at 14 000 x g for 5 minutes to pellet cell debris. Each colony PCR had the following components*: 5* µL of cleared cell supernatant*, 5* µL of 5X Phusion HF Buffer *(*New England Biolabs #B0518S*), 0.5* µL of 10 mM dNTPs, 1.25 µL of 10 µM forward primer, 1.25 µL of 10 µM reverse primer, 0.25 µL of home-purified Phusion DNA polymerase, and dH_2O to a final volume of 25 µL. PCR amplification was initiated by denaturation at 95°C for 90 seconds and was followed by 35 cycles alternating between denaturation (95°C; 30 seconds), annealing (64°C on an Eppendorf gradient thermocycler; 30 seconds), and extension (72°C; 45 s per Kb of DNA). A final extension was carried out at 72 °C for 5 minutes. Approximate insert sizes were determined by separating PCR products (8 µL) on an agarose gel (1-2.0%, depending on insert size) precast with ethidium bromide at 110 V for 30 minutes.

2.10 Small-scale plasmid isolation (Miniprep)

Minipreps were conducted when small amounts of plasmid DNA were required. E. coli cells carrying the plasmid of interest were grown in 5 mL of Super Broth with the appropriate selective antibiotic. Cultures were incubated overnight at 37°C with shaking at 250 rpm. The following day, cells were pelleted by centrifuging at 3275 x g for 10 minutes. Plasmid DNA was obtained with the EZ-10 Spin Column Plasmid DNA Miniprep Kit (Bio Basic #BS413) following the manufacturer's protocol. DNA concentration was measured using a spectrophotometer (A_{260} nm) and plasmid quality was evaluated by separating 0.5 µg of sample on an agarose gel at 110 V for 30 minutes.

2.11 Large scale plasmid isolation (Maxiprep)

E. coli cells carrying the plasmid of interest were grown in 5 mL of Super Broth with the appropriate selective antibiotic. Cultures were incubated overnight at 37°C with shaking at 250 rpm. The following day, 1 mL of culture was transferred to 200 mL of selective LB medium and incubated with shaking at 300 rpm for an additional 12-18 hours. Cultures were then transferred into multiple 50 mL conical polypropylene tubes and centrifuged at 3275 x g for 10 minutes to pellet cells. Plasmid DNA was obtained with the QIAGEN Plasmid Maxi Kit following the manufacturer's protocol (QIAGEN #12162). DNA concentration was measured using a spectrophotometer (A_{260} nm) and plasmid quality was evaluated by separating 0.5 µg of sample on an agarose gel at 110 V for 30 minutes.

2.12 Agrobacterium-mediated transient expression in tobacco leaves

Agrobacterium transformation – Plasmid DNA (100-150 ng) was added to 20-40 µL of thawed *Rhizobium radiobacter* (syn. Agrobacterium) AGL1 electrocompetent cells and the mixture was transferred into 2 mm cuvettes (BioRad #1652082). Cells were electroporated (2.5 kV, 25 µF capacitance, and 400 Ω resistance) and allowed to recover in 2 mL of non-selective YEP medium for 2 hours at 28°C with shaking at 250 rpm. After recovery, 120 µL of cells were plated on YEP agar plates containing 50 µg/mL carbenicillin and 50 µg/mL kanamycin (YEP Carb/Kan) and incubated at 28°C for 3 days.

Agroinfiltration in tobacco leaves – Constructs were expressed in tobacco as pokeweed is at least partially resistant to agroinfiltration (our unpublished data; Kanzaki et al., 2005). Agrobacterium cultures harbouring promoter::GFP-GUS constructs were used to agroinfiltrate four-leaf *Nicotiana tabacum* leaves as previously described (Zhao et al., 2017). Freshly transformed (< 1.5 weeks) Agrobacterium AGL1 cells were grown overnight (28°C, 250 rpm) in 5

mL of YEP Carb/Kan liquid medium. Agrobacterium cells that had previously been transformed with a helper component protease (HcPro) plasmid were grown overnight under the same conditions. The following day, 100 µL of culture was transferred to 25 mL of YEP Carb/Kan medium and incubated with shaking at 300 rpm for an additional 12-14 hours, until late log phase (OD_{600} = 0.7 – 1.2). Cells were pelleted by centrifugation at 3275 x g for 10 minutes and washed with 15 mL of agroinfiltration solution (10 mM MES-KOH, pH 5.6, 10 mM $MgCl_2$, 200 µM acetosyringone). After washing, cells were resuspended in agroinfiltration solution to a final OD_{600} of 0.5 and incubated at room temperature for 3 hours. Prior to agroinfiltration, cultures transformed with the promoter::GFP-GUS constructs were each mixed in a 2:1 ratio with the HcPro culture. Agrobacterium cells were injected into the abaxial surface of leaves using a needleless 5 mL syringe. For the JA experiment, leaves were sprayed with either 0 mM JA (0.5% ethanol) or 5 mM JA 24 hours after agroinfiltration. For all experiments, direct comparisons were only made between the same batches of agroinfiltrated plants.

2.13 Histochemical GUS staining

GUS histochemical assays were performed according to Jefferson et al. (1987) with some modifications. Leaf discs (0.5 cm diameter) from inoculated plants (minimum of three plants per construct) were harvested 72 hours post agroinfiltration, vacuum-infiltrated with 5-bromo-4-chloro-3-indolyl-b-D-glucuronic acid, cyclohexylammonium salt (X-Gluc) solution (1.2 mM X-Gluc, 100 mM sodium phosphate buffer, pH 7.0, 10 mM EDTA, 2 mM potassium ferricyanide, 2 mM potassium ferrocyanide, 0.1% Triton X-100), and incubated overnight at 37°C. After GUS staining, leaf discs were cleared of chlorophyll by washing in increasing concentrations of ethanol (50%-100%) for 48 hours.

2.14 GUS protein extraction and fluorometric GUS assay

GUS protein extraction – Leaf discs (1 cm diameter; 2 discs per sample) from inoculated plants (minimum of three plants per construct) were harvested into 2 mL screw cap tubes (Sarstedt #72.693) 72 hours post agroinfiltration, flash frozen in liquid nitrogen, and stored in liquid nitrogen until processing. To homogenize tissue, approximately 200 mg of 1.0 mm glass beads (BioSpec #11079110) and 300 µL of cold GUS extraction buffer (50 mM phosphate buffer, pH 7.0, 10 mM DTT, 10 mM EDTA, 0.1% SDS, 0.1% Triton X-100, 10 mM β-mercaptoethanol) were added to frozen leaf discs and the mixture was placed into a BioSpec 3110BX Mini-BeadBeater for 1 minute at setting 48. Tubes were placed on ice for 1 minute to cool before the pulse was repeated. Following tissue homogenization, samples were centrifuged for 15 minutes (14 000 x g, 4°C) and supernatants were transferred into new 1.5 mL microtubes. The centrifugation step was repeated once more to fully clear extracts. Aliquots were taken for protein quantification and the 4-methylumbelliferyl-β-D-glucuronide hydrate (4-MUG) fluorometric GUS assay.

Protein concentrations were determined using a BCA Reducing Agent Compatible Protein Assay Kit (G-Biosciences #786-573) following the microplate version of the manufacturer's protocol. To ensure buffer compatibility with the BCA reagent, samples were diluted by half with dH_2O to a final volume of 25 µL. Diluted protein samples were incubated with 25 µL of working RACA solution at 37°C for 15 minutes, followed by 200 µL of working BCA solution at 37°C for 30 minutes. A_{562} nm was measured at room temperature using a Synergy H4 Hybrid microplate reader and compared to a previously determined BSA standard curve.

Fluorometric GUS assay – GUS fluorometric assays were performed in black, clear bottom 96-well plates according to Côté and Rutledge (2003), with some modifications. In a black 1.5 mL microtube, 10 µL of cleared plant extract was mixed with 720 µL of 0.1 mM 4-MUG substrate and incubated at 37°C. Beginning at 0 minutes, 10-µL aliquots were taken from each reaction every 15 minutes for a total of 60 minutes and pipetted into wells containing 180 µL of stop buffer (0.2 M Na_2CO_3). Three technical replicates per time point were taken for each sample. Fluorescence values were measured at room temperature using a Synergy H4 Hybrid microplate reader (excitation at 365 nm; emission at 455 nm) and compared to a previously determined 4-MU standard curve. GUS activity was calculated from the linear slope of the fluorescence readings and normalized to the total protein concentration.

2.15 Isolation of total tobacco RNA from agroinfiltrated leaves

Leaf discs (1 cm diameter; 2 discs per sample) from inoculated plants (minimum of four plants per construct) were harvested into 2 mL screw cap tubes (Sarstedt #72.693) 72 hours post agroinfiltration, flash frozen in liquid nitrogen, and stored in liquid nitrogen until processing. Frozen tissue was homogenized in a 2 mL screw cap tube with a pestle and total RNA was extracted with the Monarch Total RNA Miniprep Kit (New England Biolabs #T2010) according to the manufacturer's protocol, except that samples were eluted with 50 µL of dH_2O instead of 100 µL. In-column DNAse treatment was performed according to the manufacturer's instructions. RNA concentration was measured using a spectrophotometer (A_{260} nm) and RNA quality was evaluated by separating approximately 400 ng of sample on a 2% agarose gel at 110 V for 30 minutes.

2.16 Reverse transcription (RT) of cDNA from total tobacco RNA

Total tobacco RNA (250 ng) from agroinfiltrated leaves was combined with 1 µL of 10 µM gene-specific reverse primer to a final volume of 10 µL and denatured at 65°C for 10 minutes. Following denaturation, RNA samples were snap cooled on ice for 2 minutes and mixed with MashUp reverse transcriptase (Klenov, unpublished) master mix, which had the following components: 0.5 µL of purified MashUp reverse transcriptase, 20 units of murine RNAse inhibitor (New England Biolabs #M0314), 4 µL of 5X First Strand Buffer, 1 µL of 0.1 M DTT, 1 µL of 10 mM dNTPs, and dH_2O to a final volume of 10 µL. Samples (20 µL) were incubated at 50°C for 60 minutes, followed by heat inactivation of the RT enzyme at 70°C for 10 minutes. cDNAs were stored at -40°C until further quantification.

2.17 Quantitative RT-PCR (qRT-PCR) of GFP-GUS transcripts from agroinfiltrated leaves

RT reaction product (2.5 µL) was combined with forward and reverse primers (refer to **Table 4** in Appendix B for sequences and final concentrations), 33 µL of 2X SYBR Green qPCR Master Mix (Bimake #B21202), and dH_2O to a final volume of 66 µL. Each sample was mixed by vortexing and divided into three 0.2 mL microtubes to make 20 µL triplicates. qRT-PCRs were conducted in a QIAGEN Rotor-Gene Q thermocycler with the following settings: hold at 50°C for 20 seconds, initial denaturation and hot-start DNA polymerase activation for 10 minutes, followed by 40 cycles alternating between denaturation (95°C; 15 seconds) and combined annealing/extension (68°C, 45 seconds). RNA levels were quantified using the ΔΔCt method (Livak and Schmittgen, 2001). Ribosomal protein L25 and ubiquitin-conjugating enzyme E2 (Ntubc2) transcripts served as internal controls. Melting curve analyses were conducted to ensure that only one PCR product was present for each reaction.

2.18 Stress treatments of pokeweed plants and isolation of total pokeweed RNA

Stress treatments – Four-leaf pokeweed plants were subjected to the following treatments: sprayed with 5 mM JA or SA (solubilized in 0.5% ethanol), watered every 3 days for a seven-day period with 10% polyethylene glycol (PEG), or wounded with forceps (WND). Plants sprayed with 0.5% ethanol (ET) or watered normally (WT) served as controls for JA/SA and PEG/WND, respectively. Leaf tissue (2 g) was flash-frozen and stored in liquid nitrogen 24 hours following treatment for JA, SA, ET, and WND samples, and 3 days after the final treatment for PEG and WT samples. Leaves from three independent plants were pooled for each biological replicate.

Isolation of total pokeweed RNA – Total RNA was isolated from frozen pokeweed leaves using an RNAzol-based method. Frozen leaf tissue (2 g) was homogenized in a chilled mortar with a pestle and transferred into a 50 mL conical polypropylene tube containing 4.5 mL of RNA pre-extraction buffer (20 mM citrate buffer, pH 4.0, 100 mM NaCl, 10 mM EDTA, 1% SDS, 2% PVP, 2% β-mercaptoethanol) and 4.5 mL of acid phenol:chloroform:isoamyl alcohol (25:24:1, pH 4.3). The mixture was vortexed well and centrifuged at 3275 x g in a Beckman-Coulter Allegra X-12 swinging bucket centrifuge for 10 minutes. The upper layer was transferred into a new 50 mL tube, mixed well with equal volume acid phenol:chloroform:isoamyl alcohol, and aliquoted into multiple 1.5 mL microtubes. Samples were centrifuged at 14 000 x g for 10 minutes and upper layers were pooled into a single 50 mL tube. RNAzol (Sigma-Aldrich #R4533; 2.5 volumes) was added to this pooled sample and the resulting mixture was vortexed and incubated on ice for 15 minutes. Following incubation, the mixture was aliquoted into multiple 1.5 mL microtubes and centrifuged at 14 000 x g for 10 minutes. Supernatants (1 mL per microtube) were transferred into a single 50 mL tube and mixed with equal volume 100% isopropanol. RNA was pelleted by centrifuging at 3275 x g for 20 minutes. The pellet was washed twice with 500 µL of 75% ethanol and resuspended in 100-300 µL of 1 mM citrate buffer, pH 6.4, depending on pellet size. RNA

concentration was measured using a spectrophotometer (A_{260} nm) and RNA quality was evaluated by separating 2 µg of sample on a 2% agarose gel at 110 V for 30 minutes. RNA samples from stress-treated plants (n = 4 pooled biological replicates) were sent to The Centre for Applied Genomics (The Hospital for Sick Children, Toronto, ON, Canada) for cDNA library preparation. RNA-seq libraries were constructed with the TruSeq Stranded mRNA Library Preparation Kit (Illumina #RS-122-2101).

2.19 Reverse transcription (RT) of cDNA from total pokeweed RNA

Total pokeweed RNA (500 ng) was combined with 1 µL of 10 µM gene-specific reverse primer to a final volume of 10 µL and denatured at 65°C for 10 minutes. Following denaturation, RNA samples were snap cooled on ice for 2 minutes and mixed with SuperScript III reverse transcriptase (Thermo Fisher #18080093) master mix, which had the following components: 25 units of SuperScript III reverse transcriptase, 20 units of murine RNAse inhibitor (New England Biolabs #M0314), 4 µL of 5X First Strand Buffer, 1 µL of 0.1 M DTT, 1 µL of 10 mM dNTPs, and dH_2O to a final volume of 10 µL. Samples (20 µL) were incubated at 50°C for 60 minutes, followed by heat inactivation of the RT enzyme at 70°C for 10 minutes. cDNAs were stored in -40°C until further quantification.

2.20 Quantitative RT-PCR (qRT-PCR) of PAP transcripts

qRT-PCR was performed according to **2.17** but with pokeweed-specific reference genes and primers (refer to **Table 5** in Appendix B for sequences and final concentrations). Elongation factor-1-gamma (EF1G) and the cell wall protein BIIDXI (BDX) served as internal controls for pokeweed RNA-seq validations, as these transcripts were stably expressed under our four stress treatments. Gene-specific PAP isoform primers were designed to anneal to nucleotide sequences

that are unique to each isoform and melting curve analyses were conducted to ensure that only one PCR product was present for each reaction.

2.21 Validation of PAP gene models

To validate PAP gene models at both the genomic and mRNA levels, PCR was performed as described in **2.3** using either pokeweed gDNA (see **2.2**) or pokeweed cDNA (see **2.19**) as the starting template. PCR products were gel-purified as described in **2.5** and used as templates for a second round of PCR, this time with forward and reverse primers that contained additional sequences for cloning. Amplicons from the second round of PCR were gel-purified and cloned into the multiple cloning site of pHSG299 vector using one-step SLIC (see **2.6**). All constructs were sequenced and compared with computationally-derived gene and mRNA models. Primers used for gene model validations are listed in **Table 6** (Appendix B).

2.22 Validation of the PAP-I alternative transcript in pokeweed

Reverse transcriptase was performed as described in **2.19**, except that 0.5 μL of MashUp RT (Klenov, unpublished) was used instead of SuperScript III. Using 0.5 μL of RT product as template, PCR was performed as described in **2.3** with the primers listed in **Table 7** (Appendix B).

2.23 Identification of putative stress-responsive CREs in PAP promoters

The 1.3 Kb sequence upstream of the predicted TSS was considered the proximal promoter for each PAP isoform. Promoter sequence identity analysis was performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters. Putative plant-specific cis-regulatory elements (CREs) were identified using PLACE (http://www.dna.affrc.go.jp/htdocs/PLACE/; currently accessible at

http://www.dna.affrc.go.jp/PLACE/?action=newplace) and PlantPan 2.0 (http://plantpan2.itps.ncku.edu.tw/) web interfaces (Chow et al., 2016; Higo et al., 1999). To reduce false positives, only sequences with ≥ 90% identity to the published motifs were included.

2.24 Statistical analysis

Comparisons between means of two groups were performed using two-tailed t-tests ($p \leq 0.05$), while comparisons between more than two groups were performed using ANOVAs followed by Tukey's post-hoc multiple pairwise comparisons tests ($p \leq 0.05$). All statistical analyses were conducted in R (R Core Team, 2017). Figures were made in Microsoft Excel 2016 or in R version 3.5.0 with the package ggplot2 (Wickham, 2011).

3. RESULTS

3.1 DISCOVERY OF THE PAP-I LEADER INTRON AND PROMOTER

3.1.1 PAP-I contains a 1.6 Kb leader intron

The pokeweed genome had not yet been sequenced at the beginning of this project. Previously, however, approximately 1 Kb of the putative PAP-I promoter had been isolated through genome walking (our unpublished data). Only 36 bps of this sequence aligned with the 117-bp PAP-I 5'UTR region that had been determined computationally through RNA-seq (Neller et al., 2016) and experimentally through 5'RACE (**Figure 5**). Since there are multiple PAP genes and pseudogenes with high sequence similarity in pokeweed, I first sought to determine through PCR of pokeweed genomic DNA (gDNA) if the sequence obtained through genome walking is upstream of the PAP-I coding sequence. Forward primers that align to different regions of the PAP-I gene (A to F; black arrows) were designed from the genome walking and PAP-I mRNA sequences, and paired with a single reverse primer ('Reverse'; green arrow) that aligns to the 3' end of the PAP-I coding region (**Figure 6A**). PCR with genome walking forward primers C and D produced amplicons that were 1.1 and 1.7 Kb in length (**Figure 6B**), respectively, and sequencing of these PCR products confirmed that the 1 Kb genome walking fragment is upstream of the PAP-I coding region. Unexpectedly, PCR with PAP-I 5'UTR forward primers B, E, or F produced either a 1 Kb (B) or a 2.7 Kb (E and F) amplicon. In contrast, RT-PCR from total pokeweed RNA with the same primers all produced 1 Kb bands, corresponding to the size of the mature PAP-I transcript (**Figure 6C**). Sequencing of the 2.7 Kb long amplicons revealed that there is a 1.6 Kb intron flanked by two portions of the 5'UTR (**Figure 7A**) and that the genome walking sequence that was previously thought to be the PAP-I promoter is part of the novel leader intron. Although atypical for its length, as plant introns are usually no more than

A

Top: PAP-I genome walking sequence
Bottom: PAP-I mRNA sequence

```
*        *        *        *        *        *        *        *        *        *
    1 AAAAGGGTGTGACTTTCTAGTATATATTATACTCGTAACATAACTTTTTTGTTTAAAAAGATACTCTTGCTAGTTTCCATGCGTTTAATTTAGATAAAAA 100
    1 A------T---A------------------ACT-G---CA---------TGTT----------CTC-----A-T----A-------AA---A-A-A---- 24
                                                *                              *
              *        *        *        *        *        *        *        *        *        *
  101 AAATAAAGGGTATATTTTGCATTTCTAACAGATTTGGAAGGCTTCTTCTTATTGACTGTTTTTTTTTTACTTTGGATTATAATTGAATAAAAATAAATAT 200
   25 ------------------GC----CT--CA----------GC----------TG-C-----------TA------A--A-AA---A-------------- 42
                                *                                        *
              *        *        *        *        *        *        *        *        *        *
  201 TGATGTTAATGGATAATAAGGGACTAAAATAAAAATAACTTAATATATTATATGTATTTTTTCGCTCCATAAAGCTACATAAAGTCAATGCATTTCAAAT 300
   43 ----------------------AC----------------------------GT-T------G-T-----AAG--A-A-AAA---AA-G-A----AA-- 63
                                                                          *
              *        *        *        *        *        *        *        *        *        *
  301 GAACGATTGGTGTATCAGTCCTCAAGGATGTTTAATGGATAATAAGGGACTAAATAAAAATAACTTAATATATTATATGTATTTTTTCGCTCTATAAAGC 400
   64 G-----TT-GTG----AGT--T-A------------------------ACT-----------AC----------------------------------- 79
                    *
              *        *        *        *        *        *        *        *        *        *
  401 TACATAAAGTCAATGCATTTCAAATGAATGATTAGTGTATCAATCCTCAAGGATGTTTAATGGATAATAAGGGACTAAATAAAAATAACTTAATATATTA 500
   80 ---------------------------------------------------------------------------------------------------- 80

              *        *        *        *        *        *        *        *        *        *
  501 TATATATTTTTTTCGCTCTATAAAGCTACATAAAGTCAATGCATTTCAAATAAACGATTGGTGTATCAGTCCTAAGGATGTATGAATCCATAATTCAGTT 600
   80 ---------------------------------------------------------------------------------------------------- 80

              *        *        *        *        *        *        *        *        *        *
  601 ACAACAAGGTTGTGGTTATAAGAGTATTCAGTTGTCTTTCCTTTTTTGGTTGCTTGATATATTATATATATATATAGAGTATTACCAAGGCTCGACACGA 700
   80 ---------------------------------------------------------------------------------------------------- 80

              *        *        *        *        *        *        *        *        *        *
  701 CTCGAGTTGACTCGGTTACAAACCTAATTGGTCTAAGGAGAACAACTCTCTATATAGTATTACAAGGATACGCATACATCGTACATGTAATCCAGGTTCT 800
   80 ---------------------------------------------------------------------------------------------------- 80

              *        *        *        *        *        *        *        *        *        *
  801 TACCTTGTTTTTCATTGATAGGGCGAAAGTATTGGAACTAGCTAGTAGGAAGGGAAGATGAAGTCGATGCTTGTGGTGACAATATCAATATGGCTCATTC 900
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   80 -------------------AGGGCGAAAGTATTGGAACTAGCTAGTAGGAAGGGAAGATGAAGTCGATGCTTGTGGTGACAATATCAATATGGCTCATTC 160
                         *        *        *        *        *        *        *        *        *
              *        *        *        *        *        *        *        *        *
  901 TTGCACCAACTTCAACTTGGGCTGTGAATACAATCATCTACAATGTTGGAAGTACCACCATTAGCAAATACGCCACTTTTCTGAATGATC~~~~~~~~~~ 990
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||.........
  161 TTGCACCAACTTCAACTTGGGCTGTGAATACAATCATCTACAATGTTGGAAGTACCACCATTAGCAAATACGCCACTTTTCTGAATGATCTTCGTAATGA 260
              *        *        *        *        *        *        *        *        *        *
```

B

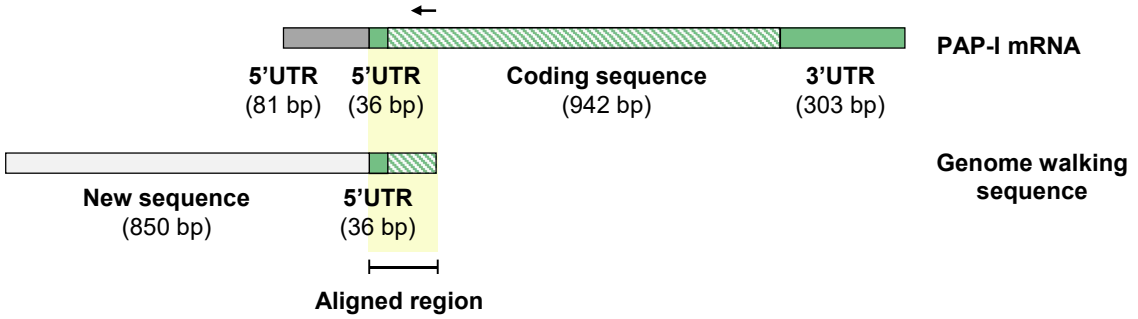


Figure 5. The sequence obtained through genome walking does not fully match the PAP-I mRNA sequence. (A) Alignment of the PAP-I genome walking sequence (top) with the PAP-I mRNA (bottom). The starting methionine (ATG) is highlighted in green. **(B)** Schematic diagram of the PAP-I transcript and the genome walking sequence. The position of the reverse primer (black arrow) used for genome walking is shown. The regions that align (shown in **A**) are in green and highlighted in yellow. The regions that show no alignment are in grey.

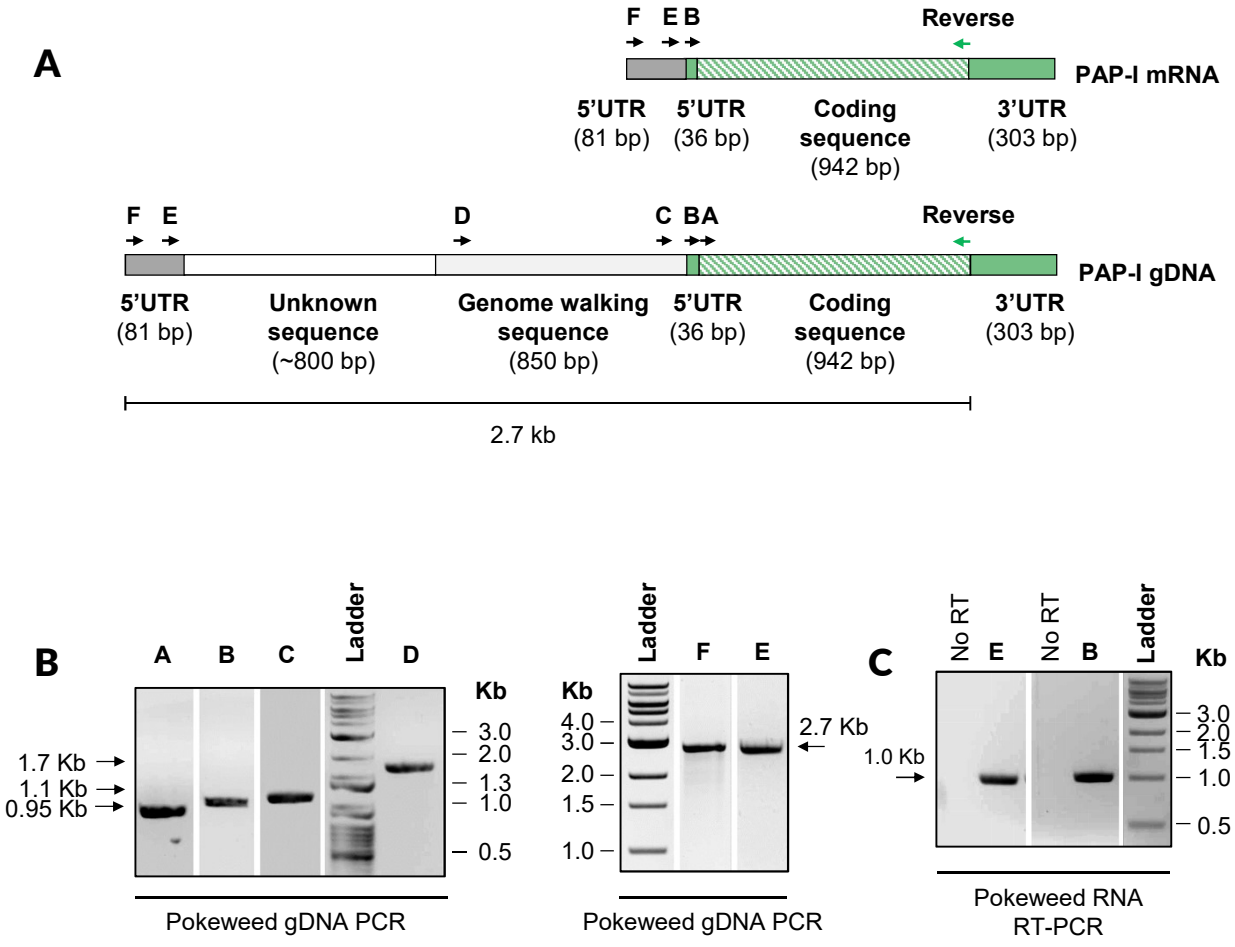


Figure 6. A 2.7 Kb PCR product is generated with some PAP-I 5'UTR forward primers. (A) Schematic diagram of the PAP-I gene. For PCR, the same reverse primer (green arrow) was paired with several forward primers (black arrow; A – F) that anneal to different regions of PAP-I. Primers were designed from the PAP-I mRNA and genome walking sequences. **(B)** PCR of pokeweed genomic DNA with the primer pairs listed in **A**. **(C)** RT-PCR from total pokeweed RNA with two of the primer pairs listed in **A**. The size of the PCR product corresponds to the length of the mature PAP-I transcript. All PCR products were sequenced.

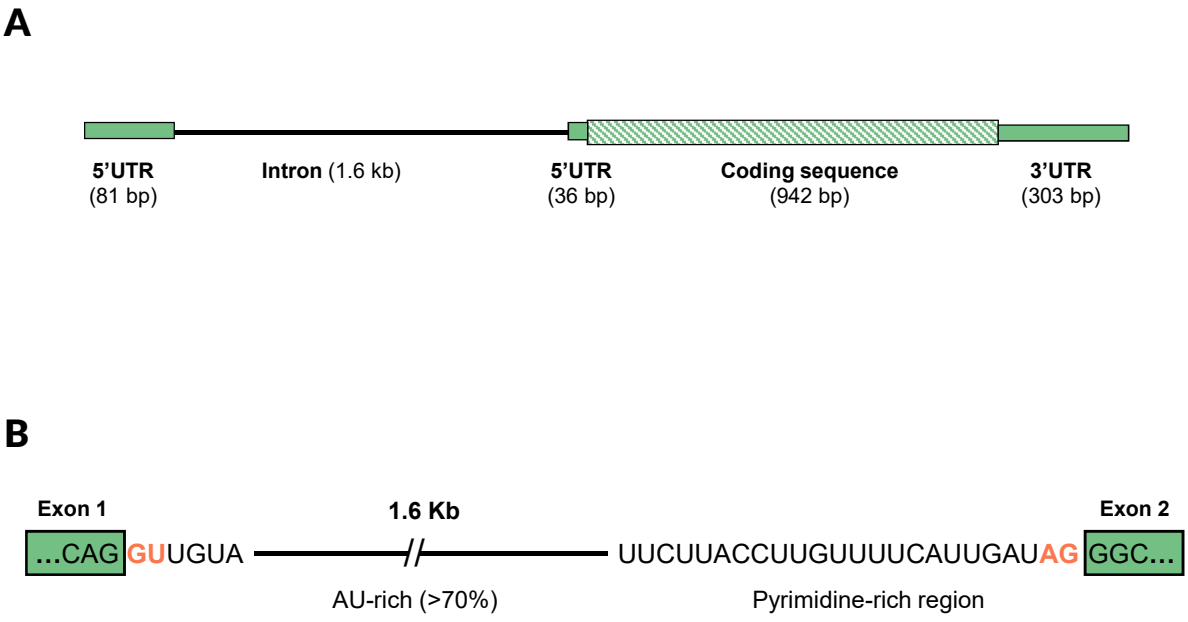
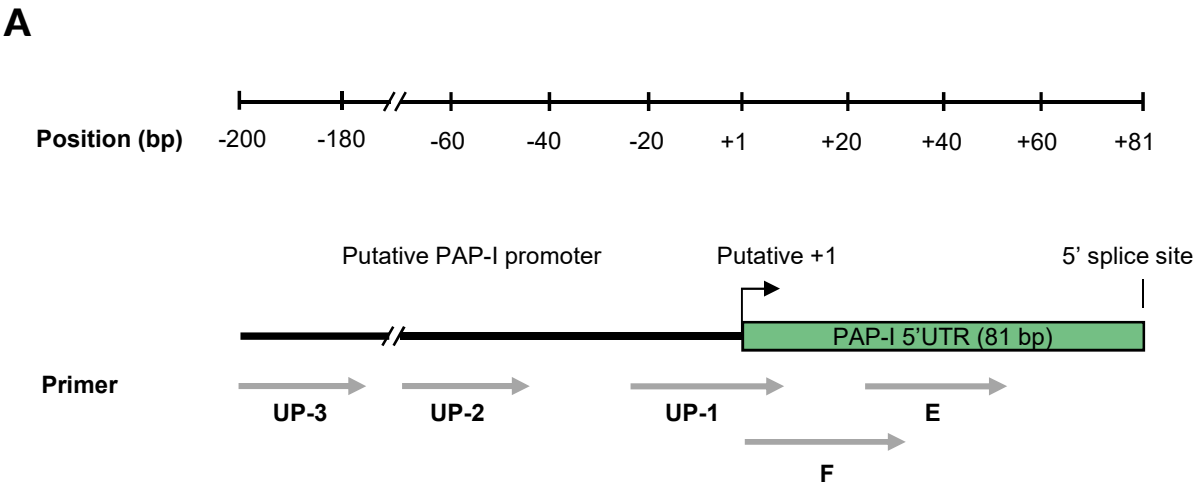


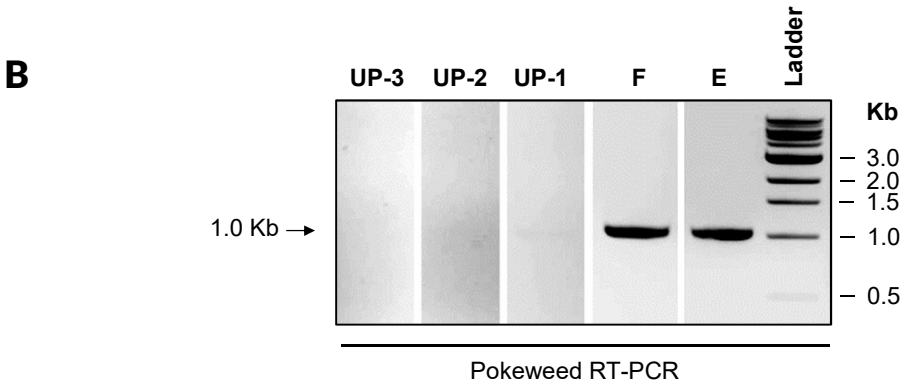
Figure 7. PAP-I contains a 1.6 Kb leader intron. (A) Schematic diagram of the PAP-I gene, showing two exons and a novel intron in the 5'UTR. The genome walking sequence previously thought to be the PAP-I promoter is part of the PAP-I intron. **(B)** The 1.6 Kb sequence has features that are common to plant introns, such as canonical 5'GU/3'AG borders (shown in red), a polypyrimidine tract, and an AU-rich sequence.

several hundred nucleotides long (Wu et al., 2013), the 1.6 Kb PAP-I leader intron contains canonical plant intron features, such as 5'GU/3'AG splice borders, a polypyrimidine tract at the 3'end, and an AT-rich (>70%) sequence (**Figure 7B**).

3.1.2 Identification of the PAP-I promoter

To identify the PAP-I promoter, I first needed to validate the sequence of the PAP-I 5'UTR that we had previously determined, and from this, map the approximate position of the +1 transcription start site (TSS). Using our recently assembled pokeweed genome, RT-PCRs were conducted with forward primers that align to regions near or within the putative 5'UTR (**Figure 8A**) and a reverse primer that aligns to the 3' end of the PAP-I coding sequence. To ensure that the lack of PCR products was not due to inefficient primer annealing, all RT-PCRs were performed at four different annealing temperatures using a gradient thermocycler. No clear bands were observed when the forward primers were positioned upstream of the putative TSS (**Figure 8B**). Since our 5'UTR does not fully agree with a published PAP-I mRNA sequence (accession number DL213291.1), it was initially thought that our PAP-I 5'UTR sequence was incomplete. However, a BLASTN against our pokeweed genome showed that the PAP-I DL213291.1 5'UTR does not align to any portion of the 18 Kb contig that contains PAP-I, and instead aligns to the coding sequence of an unrelated gene (pokeweed homologue of VIP5, an RNA polymerase II-associated protein). Taken together, these results suggest that we had previously identified the full-length PAP-I 5'UTR both through RNA-seq and 5'RACE, and that the PAP-I DL213291.1 mRNA contains erroneous sequence. Therefore, the 1.3 Kb region immediately upstream of the TSS was identified as the PAP-I promoter. The PAP-I core promoter was predicted to contain three well-conserved CREs: the CAAT box, the TATA box, and the Inr element (**Figure 8C**).





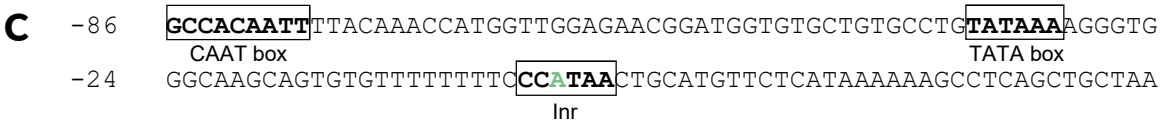


Figure 8. Verification of the PAP-I +1 transcription start site through RT-PCR. (A) Schematic diagram of the PAP-I 5'UTR with the putative transcription start site (TSS) as determined through de novo transcript assembly and 5'RACE. Different forward primers (grey arrow) were paired with a reverse primer that aligns to the 3' end of the PAP-I coding sequence (not indicated). Grey arrows are drawn to scale. **(B)** RT-PCR from pokeweed total RNA with the primer pairs described in **A**. No clear bands were observed when the PAP-I forward primers were positioned upstream of the putative +1 TSS. At least two biological replicates were performed for each primer pair, and yielded the same results. **(C)** Putative cis-acting regulatory elements identified in the PAP-I core promoter (CAAT box, TATA box, Initiator). The +1 nucleotide is in green.

3.1.3 PAP genes contain long leader introns and alternative promoters

After the discovery of the leader intron in PAP-I, I wanted to determine if leader introns were also present in lesser-characterized PAP genes. Additionally, it was necessary to map the 5'UTRs of the other PAP genes in order to identify their promoters. PAP gene models were generated through the program MAKER (**Figure 9A**) by aligning transcriptome and protein evidence to the assembled pokeweed genome, and computational models were then validated through PCR of genomic DNA and RT-PCR of total RNA (**Figure 9B**). Since this method required RNA-seq reads, only the PAP genes that were expressed in leaf tissue were included. PCR from the same PAP primer pairs produced either a short or long product depending on the starting DNA template. When the templates were cDNAs from total pokeweed RNA, all PCR products were approximately 1 Kb long, corresponding to the size of the processed PAP transcripts; when the template was gDNA, however, PCR products ranged in size from 2.7 kb to 7.5 kb, corresponding to the different sizes of the PAP genes. None of the isoforms had introns in their CDSs, with the exception of PAP-II, which contained a 736-bp intron that had been reported previously (Poyet and Hoeveler, 1997). Interestingly, two of the PAP isoforms (PAP-II and PAP-S2) each had two distinct gene models that differed only in their 5'UTRs, suggesting the use of alternative promoters. The gene model of the putative novel isoform was not included because our genome assembly was highly fragmented and its 5'UTR was not found in the same genomic contig as its CDS and 3'UTR. Nevertheless, these results indicate that the long leader intron is a conserved feature in PAP genes.

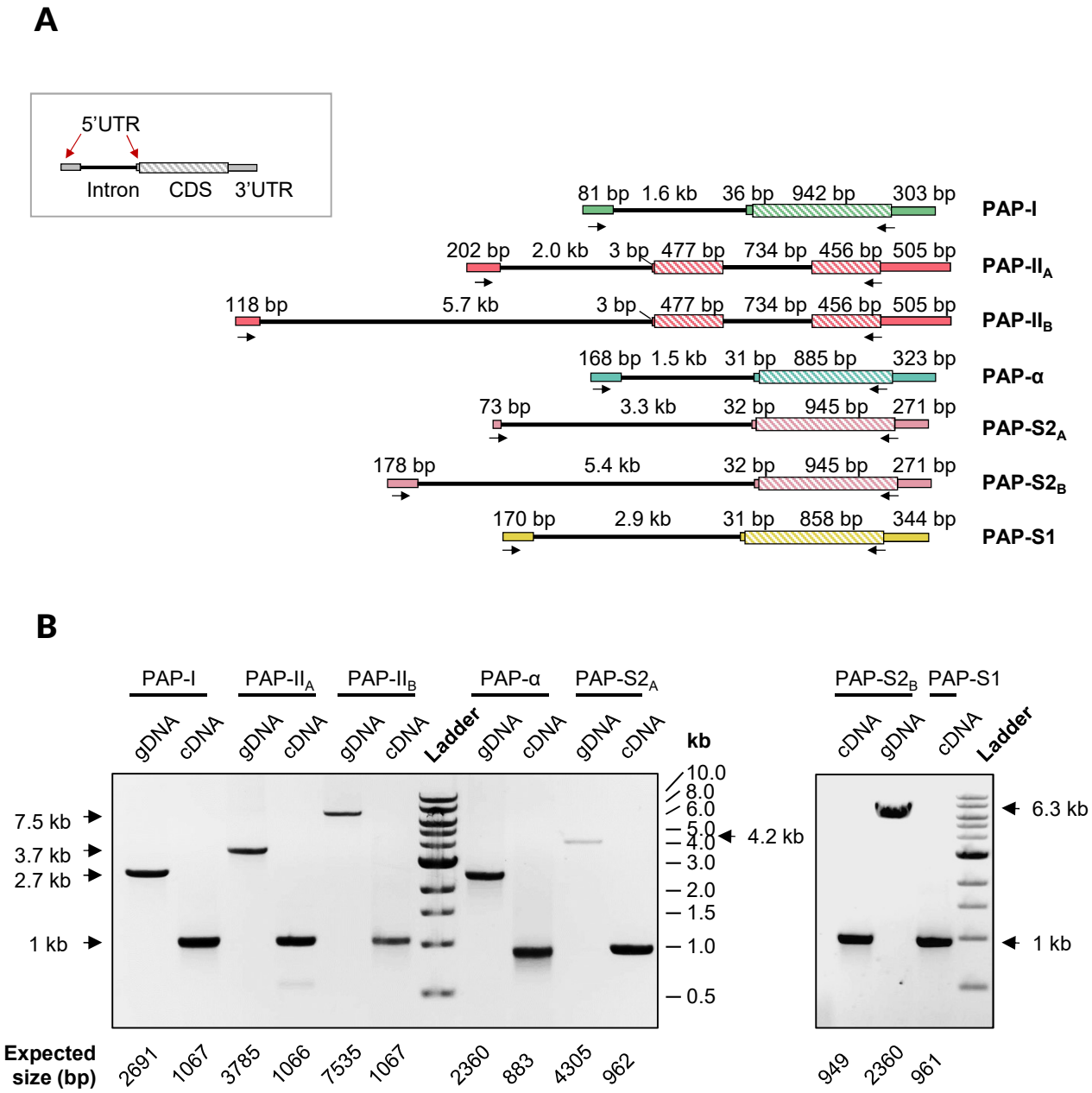


Figure 9. PAP genes contain long leader introns and alternative promoters. **(A)** PAP gene models were generated by the program MAKER using transcriptome and protein evidence. Arrows indicate primer binding sites used for gene model validations. **(B)** Validation of PAP gene models through PCR of pokeweed genomic DNA and PCR of cDNA from total pokeweed RNA. PCR products were validated by sequencing.

3.2 CHARACTERIZATION OF PAP PROMOTERS

3.2.1 PAP genes have distinct expression profiles in response to biotic and abiotic stresses

To investigate the potential roles of the PAP genes in pokeweed, plants were subjected to several biotic and abiotic stresses: jasmonic acid (JA), salicylic acid (SA), mechanical wounding (WND), and PEG, which simulates drought. PAP mRNA expression levels were then determined through RNA-seq (**Figure 10A**) and validated through qRT-PCR (**Figure 10B**). RNA-seq differential expression analysis was conducted at the gene level, meaning that transcript variants were not distinguished. Linear regression analysis between qRT-PCR and RNA-seq differential expression results had an R^2 of 0.8807, showing high correlation between the two methods of transcript quantification. In addition to showing differences in stress-induced expression change, the PAP transcripts varied greatly in abundance; average abundance in transcripts per million (TPM) across all samples ranged from 146 (PAP-α) to 9911 (PAP-I). Among the four stress treatments, the PAPs were most responsive to JA (**Figure 10C**), with four being significantly upregulated (PAP-I, PAP-II, PAP-S1, PAP-α) and one being slightly but significantly downregulated (PAP-S2). Novel PAP and PAP-S2, which exhibited only slight responses to PEG ($\log_2$ FC = -0.72) and JA ($\log_2$ FC = -0.47), respectively, had the most divergent expression profiles.

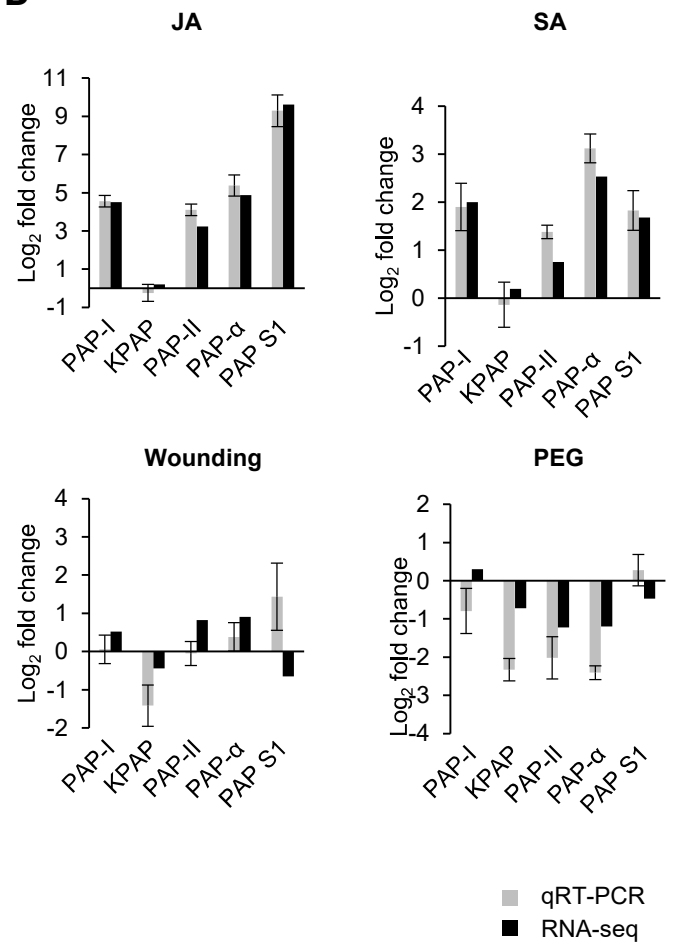
3.2.2 Putative stress-responsive cis-regulatory elements in the PAP promoters

To gain further insight into the distinct transcript expression profiles of the PAP genes, putative stress-associated CREs were annotated in the PAP proximal promoters, including the alternative promoters of PAP-II and PAP-S2 (**Table 1**). As with PAP-I, the 1.3 kb sequence upstream of the validated TSSs was considered the promoter for each PAP isoform. For the novel PAP gene, the sequence upstream of its experimentally validated 5'UTR (our unpublished data)

A

Gene ID	Isoform	log_2FC JA	log_2FC SA	log_2FC PEG	log_2FC WND	Avg. Expr. (TPM)
PHYAM_020596	PAP-I	4.50	2.00	0.30	0.52	9911
PHYAM_028184	PAP-II	3.24	0.75	-1.22	0.82	1834
PHYAM_021314	PAP-S1	9.62	1.68	-0.65	-0.47	886
PHYAM_010467	PAP-S2	-0.47	-0.06	-0.08	-0.04	495
PHYAM_022058	PAP-α	4.87	2.53	-1.20	0.91	146
PHYAM_012451	Novel	0.21	0.19	-0.72	-0.44	1092

B



C

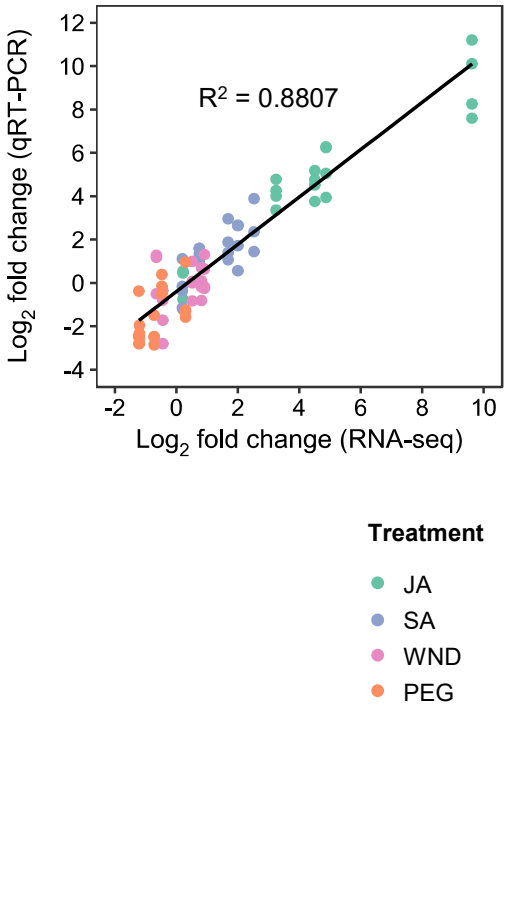


Figure 10. PAP genes have distinct stress-associated profiles. Pokeweed plants were subjected to the following treatments: sprayed with 5 mM JA or SA, watered every three days for a seven-day period with 10% polyethylene glycol (PEG), or wounded with forceps (WND). **(A)** Expression changes of all PAP genes in response to stress treatments using RNA-seq data (n = 4). TPM = transcripts per million; FC = fold change. Green and red indicate significant up- or down-regulation, respectively. **(B)** qRT-PCR validation of RNA-seq differential expression results for select PAP genes (n = 4). Elongation factor-1-gamma (EF1G) and the cell wall protein BIIDXI (BDX) served as reference genes for normalization. Error bars represent the SEM. **(C)** Linear regression analysis between qRT-PCR and RNA-seq differential expression results (log_2 fold change) for PAP genes. Results for qRT-PCR are from at least three independent biological replicates for each transcript.

Table 1. Putative stress-responsive cis-regulatory elements (CREs) in the PAP promoters.

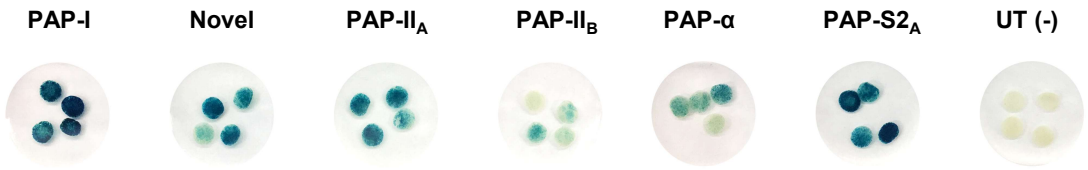
CRE	Motif	Response	Copy number in PAP promoter								Reference
			I	II_A	II_B	N	α	S1	$S2_A$	$S2_B$	
ABRELATERD1	ACGTG	ABA; Dehydration	2	1	3	3	1	-	1	2	(Simpson et al., 2003)
ABREMOTIFAOSOSEM	TACGTGTC	ABA	-	1	-	-	-	-	-	-	(Hobo et al., 1999)
ABREOSRAB21	ACGTSSSC	ABA	1	-	-	-	-	-	-	-	(Marcotte et al., 1989)
ABRERATCAL	MACGYGB	ABA; Ca^{2+}	2	-	1	2	-	-	-	-	(Kaplan et al., 2006)
ACGTATERD1	ACGT	Dehydration	6	2	4	8	4	-	2	14	(Simpson et al., 2003)
ANAERO1CONSENSUS	AAACAAA	Low oxygen	1	2	1	1	1	-	1	2	(Mohanty et al., 2005)
ANAERO2CONSENSUS	AGCAGC	Low oxygen	-	-	2	-	-	-	-	-	(Mohanty et al., 2005)
BIHD1OS	TGTCA	Disease	1	2	4	2	6	5	4	6	(Luo et al., 2005)
CACGTGMOTIF	CACGTG	Defense-related	-	-	2	2	-	-	-	-	(Gu et al., 2002)
CAREOSREP1	CAACTC	GA	2	1	1	1	1	1	1	-	(Sutoh and Yamauchi, 2003)
CCAATBOX1	CCAAT	Heat stress	4	6	2	-	-	1	-	2	(Rieping and Schöffl, 1992)
DPBFCOREDCDC3	ACACNNG	ABA	-	2	1	2	-	2	-	1	(Kim et al., 1997)
DRE2COREZMRAB17	ACCGAC	ABA; Dehydration	-	2	-	1	-	1	1	-	(Kizis and Pagès, 2002)
DRECRTCOREAT	RCCGAC	Cold; Dehydration; High salt	-	3	-	1	-	1	1	-	(Dubouzet et al., 2003)
EBOXBNNAPA	CANNTG	Cold	6	12	8	8	8	2	6	12	(Agarwal et al., 2006)
ELRECOREPCRP1	TTGACC	Pathogen; SA; Wounding	-	-	2	3	-	-	-	-	(Laloi et al., 2004)
GADOWNAT	ACGTGTC	GA	-	1	-	-	-	-	-	-	(Ogawa et al., 2003)
GARE1OSREP1	TAACAGA	GA	-	-	1	1	1	-	-	-	(Sutoh and Yamauchi, 2003)
GARE2OSREP1	TAACGTA	GA	-	-	-	-	-	-	-	2	(Sutoh and Yamauchi, 2003)
GAREAT	TAACAAR	GA	2	1	1	1	2	2	2	-	(Ogawa et al., 2003)
GT1CONSENSUS	GRWAAW	Light-inducible; SA	17	16	16	11	12	13	14	14	(Buchel et al., 1999)
GT1GMSCAM4	GAAAAA	Pathogen; High salt	6	6	5	3	3	4	5	2	(Park et al., 2004)
LTRE1HVBLT49	CCGAAA	Cold	1	1	-	3	-	-	-	1	(Dunn et al., 1998)
LTRECOREATCOR15	CCGAC	Cold; Dehydration	1	6	1	3	-	2	2	-	(Kim et al., 2002)
MYB1AT	CANNTG	ABA; Dehydration	5	4	2	2	6	6	6	3	(Abe et al., 2003)
MYB1LEPR	GTTAGTT	Defense-related	-	1	-	-	-	1	1	1	(Chakravarthy et al., 2003)
MYB2CONSENSUSAT	YAACKG	ABA; Dehydration	-	1	1	-	-	-	1	-	(Abe et al., 2003)
MYBCORE	CNGTTR	ABA; Dehydration	-	3	3	2	1	1	1	1	(Abe et al., 2003)
MYBGAHV	TAACAAA	GA	1	1	1	1	1	1	1	-	(Gubler et al., 1995)
MYCATERD1	CATGTG	Dehydration	-	-	1	-	-	-	1	1	(Simpson et al., 2003)
MYCATRD22	CACATG	Dehydration	-	-	1	-	-	-	1	1	(Simpson et al., 2003)
PREATPRODH	ACTCAT	Hypoosmolarity	-	2	1	1	2	-	-	1	(Satoh et al., 2004)
PYRIMIDINEBOXHVEPB1	TTTTTTCC	ABA; GA	1	-	-	-	-	1	-	1	(Cercós et al., 1999)
PYRIMIDINEBOXOSRAMY1A	CCTTTT	Sugar repression; GA	4	-	-	3	2	1	1	-	(Mena et al., 2002)
RYREPEATBNNAPA	CATGCA	ABA	-	1	-	-	2	5	4	-	(Ezcurra et al., 1999)
SURECOREATSULTR11	GAGAC	Sulphur	4	1	2	-	2	1	1	1	(Maruyama-Nakashita et al., 2005)
T/GBOXATPIN2	AACGTG	JA	2	-	-	-	1	-	-	1	(Boter et al., 2004)
WBOXATNPR1	TTGAC	SA	2	2	5	6	6	4	3	3	(Yu et al., 2001)
WBOXNTCHN48	CTGACY	Defense-related	-	-	2	3	-	-	-	-	(Yamamoto et al., 2004)
WBOXNTERF3	TGACY	Wounding	4	4	8	12	5	4	3	3	(Nishiuchi et al., 2004)

was considered the putative promoter, even though this region was not found in the same contig as its CDS and 3'UTR due to the fragmented nature of our genome assembly. Consistent with other eukaryotic promoters, all promoters contained TATA boxes -35 to -25 bps upstream of the putative TSSs. I also identified CREs associated with diverse biotic and abiotic stresses, such as T/GBOXATPIN2 (JA), W-boxes (SA), ABRE motifs (ABA), CCAAT boxes (heat stress), GARE motifs (gibberellic acid, GA), and MYB motifs (drought). Although some CREs were present in all PAP promoters (e.g. EBOXBNNAPA, GT1CONSENSUS, MYB1AT, and WBOXATNPR1), most elements differed in abundance and distribution. For instance, most ABRE motifs were absent in the PAP-α, PAP-S1, and PAP-$S2_A$ promoters, while T/GBOXATPIN2 was only present in PAP-I, PAP-α, and PAP-$S2_B$, and WBOXNTCHN48 was only present in PAP-II_B and Novel PAP. The two PAP-II promoters had only 50.5% sequence identity to each other, and the two PAP-S2 promoters had only 47.2% identity, suggesting that each isoform promoter contains binding sites for different sets of transcription factors.

3.2.3 PAP gene promoters can drive reporter expression

While bioinformatic analysis revealed the presence of putative CREs in the PAP promoters, I wanted to confirm that these promoter sequences were capable of driving gene expression. PAP promoter fragments, along with their validated 5'UTRs, were isolated from gDNA, placed upstream of the GFP-GUS reporter gene, and expressed in tobacco leaves through agroinfiltration. Although PCR was attempted for all putative PAP isoform promoters, not all fragments were amplified successfully, likely due to very high nucleotide sequence similarity between some gDNA regions. Nevertheless, all tested promoter fragments, including the two promoters of PAP-II, were able to drive reporter gene expression to varying degrees (**Figure 11A**), and the strengths of these promoters (high, medium, low) correlated well with the

A



B

Isoform	Expression (TPM)	Promoter strength
PAP-I	3359	High
Novel	869	Medium
PAP-II$_A$	713	Medium
PAP-II$_B$	7.7	Low
PAP-α	44	Medium-Low
PAP-S2$_A$	929	High

Figure 11. PAP isoform promoters can drive reporter gene expression. (A) The 1.3 Kb regions upstream of the validated PAP TSSs were inserted upstream of the GFP-GUS reporter gene. Tobacco leaves were agroinfiltrated with PAP promoter::GFP-GUS constructs and stained for GUS. Untransformed (UT) Agrobacterium = negative control. At least four independent plants per construct were tested (1 leaf disc is 1 biological replicate). **(B)** Abundances of PAP isoform transcripts in untreated pokeweed plants as determined through RNA-seq. TPM = transcripts per million.

abundances of their corresponding transcripts in pokeweed plants watered normally (**Figure 11B**).

3.2.4 JA-responsiveness in the PAP-I promoter is mediated by a T/GBOXATPIN2 cis-regulatory element

Since the PAP-I transcript was highly upregulated with JA, I hypothesized that CREs in the PAP-I promoter could mediate this response. Serially truncated PAP-I promoter fragments (ranging from 1262 bp to 102 bp) were placed upstream of GFP-GUS (**Figure 12A**), transiently expressed in tobacco leaves through agroinfiltration, and treated with either 0 mM or 5 mM JA. Apart from those expressing 102::GFP-GUS, all plants bearing the PAP-I promoter::GFP-GUS constructs showed higher GUS activity following JA treatment (**Figure 12B**), as indicated by the intensity of blue staining. Importantly, these results demonstrate that the -296 to -103 proximal promoter region is sufficient for the JA-responsiveness of the PAP-I promoter. As shown in **Figure 12C**, bioinformatic annotation of CREs in this region revealed an element (T/GBOXATPIN2; sequence in green text) that binds the master JA signalling regulator MYC (Boter et al., 2004). To examine the specific effect of T/GBOXATPIN2 on the JA-responsiveness of the PAP-I promoter, the element was deleted in the 1262 promoter fragment (1262/-GBOX), and leaves expressing wild-type and mutated promoter constructs were treated with JA. Deletion of T/GBOXATPIN2 altered the response of the PAP-I promoter to JA so that reporter expression now decreased significantly ($p < 0.05$) following JA treatment (**Figure 12D**). These findings suggest that the T/GBOXATPIN2 element (-180 to -175 relative to the TSS) is required for the JA-mediated upregulation of PAP-I.

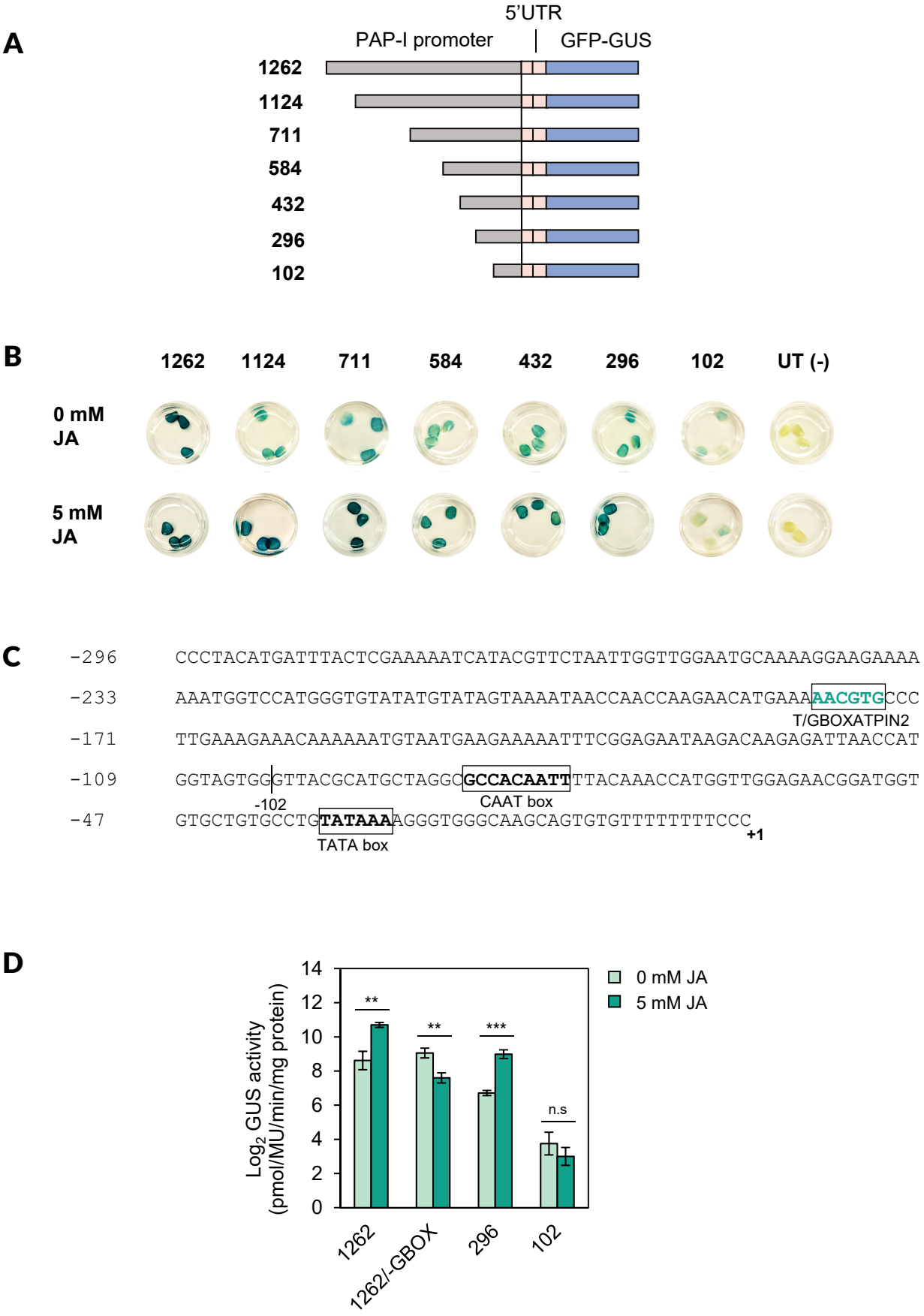

A
5'UTR
PAP-I promoter
GFP-GUS
1262
1124
711
584
432
296
102
B
1262
1124
711
584
432
296
102
UT (-)
0 mM JA
5 mM JA
C
-296 CCCTACATGATTTACTCGAAAAATCATACGTTCTAATTGGTTGGAATGCAAAAGGAAGAAAA
-233 AAATGGTCCATGGGTGTATATGTATAGTAAAATAACCAACCAAGAACATGAAAAACGTGCCC
T/GBOXATPIN2
-171 TTGAAAGAAACAAAAAATGTAATGAAGAAAAATTTCGGAGAATAAGACAAGAGATTAACCAT
-109 GGTAGTGGGTTACGCATGCTAGGCGCCACAATTTTACAAACCATGGTTGGAGAACGGATGGT
-102
CAAT box
-47 GTGCTGTGCCTGTATAAAAGGGTGGGCAAGCAGTGTGTTTTTTTTCCC
TATA box
+1
D
Log2 GUS activity (pmol/MU/min/mg protein)
14
12
10
8
6
4
2
0
**
**
***
n.s
0 mM JA
5 mM JA
1262
1262/-GBOX
296
102

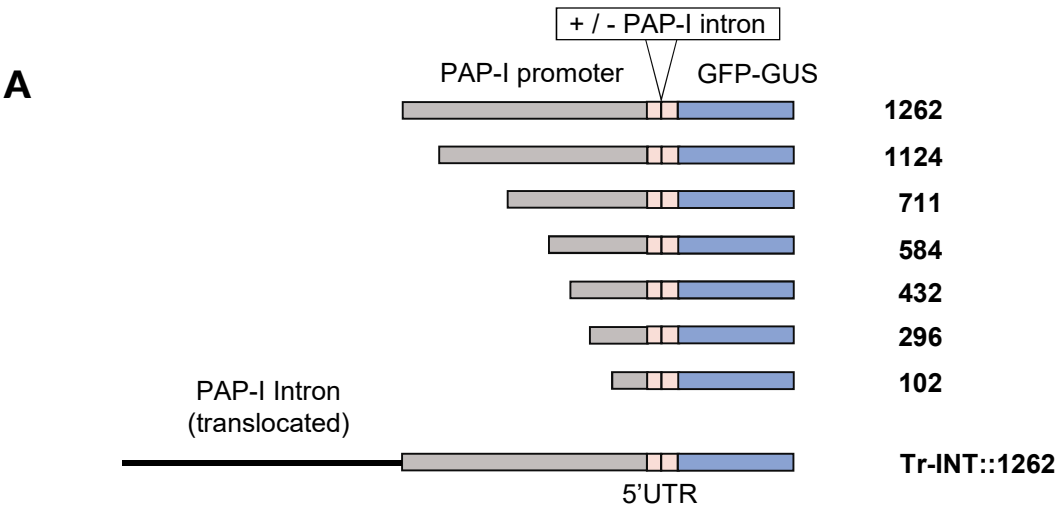
Figure 12. The PAP-I promoter is responsive to jasmonic acid. (A) Schematic diagram of the PAP-I promoter::GFP-GUS constructs. The PAP-I promoter was serially truncated from the 5' end. The PAP-I intron was excluded to examine the effect of the promoter alone. **(B)** Tobacco leaves agroinfiltrated with PAP-I promoter::GFP-GUS constructs were treated with either 0 mM (mock) or 5 mM JA and stained for GUS. Three independent plants per construct were tested (1 leaf disc is 1 biological replicate). UT = untransformed Agrobacterium. **(C)** Sequence of the -296 to +1 region of the PAP-I promoter. Nucleotide position is indicated on the left, relative to the validated TSS (+1). The CAAT and TATA boxes are shown, along with the JA-associated element T/GBOXATPIN2 (green text). **(D)** GUS fluorometric assay in tobacco leaves agroinfiltrated with PAP-I promoter::GFP-GUS constructs and treated with either 0 mM (mock) or 5 mM JA. At least four independent plants per constructs were tested. The T/GBOXPINAT2 element was deleted in the 1262/–GBOX construct. Error bars represent the SEM. Comparisons between mock-treated and JA-treated samples were performed for each promoter construct using two-tailed t-tests, $p < 0.05$. '*' = $p < 0.05$; '**' = $p < 0.01$; '***' = $p < 0.001$; 'n.s.' = not significant.

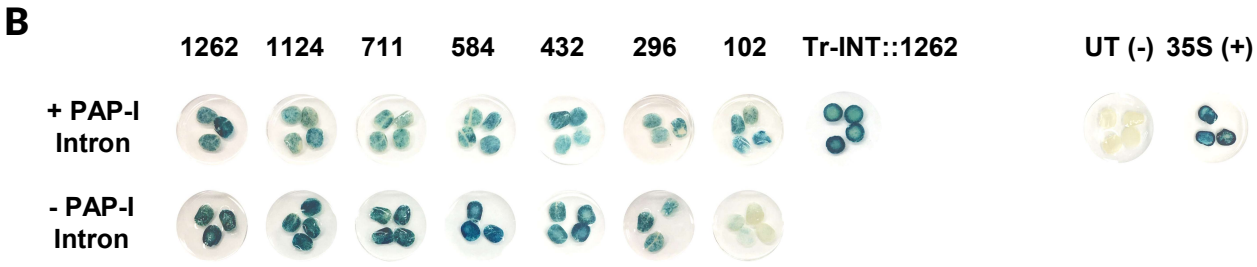
3.3 EFFECT OF THE PAP-I LEADER INTRON ON REPORTER GENE EXPRESSION

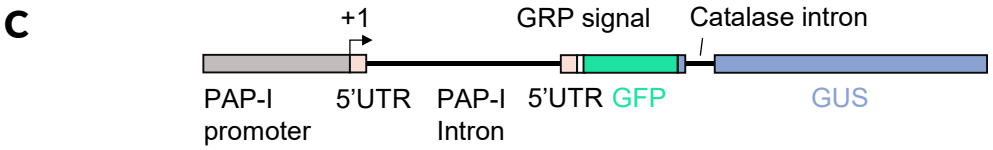
3.3.1 The effect of the PAP-I intron on reporter gene expression is dependent on promoter length

To examine the effect of the PAP-I intron on gene expression, tobacco leaves were agroinfiltrated with serially truncated PAP-I promoter::GFP-GUS constructs (with and without the intron) and stained for GUS (**Figure 13A** and **13B**). The translocation construct (Tr-INT::1262) was included to determine if the intron could influence expression in a position-independent manner. Contrary to my expectations, all three 1262 promoter fragments (1262::INT, 1262, and Tr-INT::1262) showed equally high levels of GUS staining, and the other PAP-I promoter fragments (1124, 711, 584, 432, and 296) drove higher reporter expression in the absence of the PAP-I leader intron. The one exception was the 102::INT construct, which had higher reporter expression than its intronless counterpart.

One key difference between the PAP-I gene as it is found in pokeweed and the GFP-GUS reporter gene is that GFP-GUS contains a second intron in its coding sequence (**Figure 13C**): the modified catalase intron, which is derived from the first intron of the castor bean catalase gene *CAT-1*, was originally inserted in GUS to prevent leaky reporter gene expression in Agrobacterium (Ohta et al., 1990). However, introns that are normally enhancing may inhibit gene expression in a position-dependent manner when artificially paired with another intron, possibly due to an unknown mechanism that affects translational efficiency (Bourdon et al., 2001). To ensure that the reduction in reporter expression was not due to an unexpected interaction between the two introns, the catalase intron was removed from select constructs (1262 and 102) and reporter expression was quantified in agroinfiltrated leaves through a GUS fluorometric assay (**Figure 13D**). Although the removal of the catalase intron increased GUS


A
+ / - PAP-I intron
PAP-I promoter
GFP-GUS
1262
1124
711
584
432
296
102
PAP-I Intron
(translocated)
Tr-INT::1262
5'UTR


B
1262
1124
711
584
432
296
102
Tr-INT::1262
UT (-)
35S (+)
+ PAP-I Intron
- PAP-I Intron


C
+1
GRP signal
Catalase intron
PAP-I promoter
5'UTR
PAP-I Intron
5'UTR
GFP
GUS

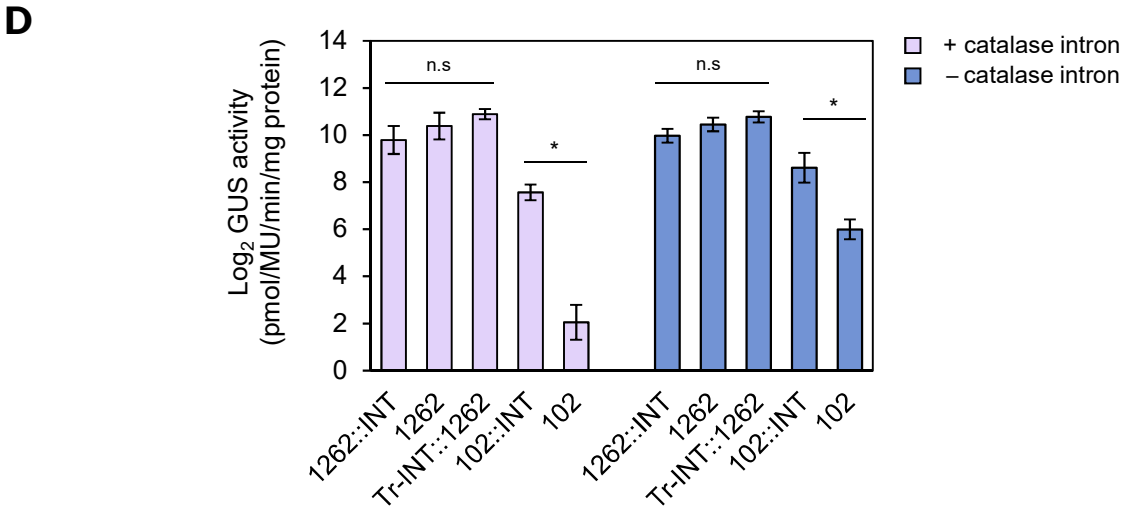

D
Log2 GUS activity (pmol/MU/min/mg protein)
14
12
10
8
6
4
2
0
n.s
n.s
*
*
+ catalase intron
– catalase intron
1262::INT
1262
Tr-INT::1262
102::INT
102
1262::INT
1262
Tr-INT::1262
102::INT
102

Figure 13. The effect of the PAP-I intron on reporter gene expression is dependent on PAP-I promoter length. (A) Schematic representation of PAP-I promoter::GFP-GUS constructs. The PAP-I promoter was serially truncated from the 5' end and placed upstream of the GFP-GUS reporter gene, with and without the PAP-I intron (+/- PAP-I intron). The PAP-I intron was also translocated upstream of the 1262 bp promoter fragment (tr-INT::1262). **(B)** Effect of the PAP-I intron on GUS expression. Tobacco leaves were agroinfiltrated with PAP-I promoter::GFP-GUS constructs (+/- PAP-I intron) and stained for GUS. At least three independent plants per construct were tested (1 leaf disc is 1 biological replicate). 35S (+) = GFP-GUS under the control of the CaMV 35S promoter. UT = untransformed Agrobacterium. **(C)** Schematic of the GFP-GUS reporter gene showing the castor bean catalase intron in the GUS coding sequence. +1 denotes the transcription start site. **(D)** GUS fluorometric assay of tobacco leaves agroinfiltrated with PAP-I promoter::GFP-GUS constructs (+/- PAP-I intron; +/- catalase intron). At least three independent plants per construct were tested. Error bars represent the SEM. Comparisons between promoter constructs were performed using a one-way ANOVA followed by Tukey's post-hoc tests, $p < 0.05$. '*' = $p < 0.05$; 'n.s.' = not significant.

activity slightly, particularly for the lower-expressing 102 promoter constructs (factorial ANOVA, $p < 0.05$), the trend observed in the GUS histochemical experiment remained the same for the constructs tested. Therefore, the effect of the PAP-I intron on reporter expression is dependent on promoter length. Importantly, it is not required for high-level expression when a strong PAP-I promoter is present.

3.3.2 The PAP-I intron has intrinsic promoter activity

Although the PAP-I intron appears to reduce reporter gene expression when paired with some PAP-I promoter fragments (1124, 711, 584, 432, and 296), it increased the level of GFP-GUS significantly when paired with the minimal promoter (102), which was otherwise unable to drive high reporter expression on its own. To investigate the mechanism by which the PAP-I intron is acting, two additional reporter constructs were made (**Figure 14A**): the PAP-I intron was either inserted into the GFP-GUS vector in the absence of any promoter (INT), or translocated upstream of the 102 fragment (Tr-INT::102). The translocated construct was made to determine if the influence of the PAP-I intron is position-dependent. Histochemical and fluorometric GUS assays (**Figure 14B** and **14C**) revealed that INT and Tr-INT::102 were able to drive reporter expression at levels similar to the 102::INT construct ($p > 0.05$, one-way ANOVA). More importantly, these results show that the PAP-I intron can drive reporter expression even in the complete absence of the promoter.

If the PAP-I intron has intrinsic promoter activity and transcription can initiate within the intron region, a PAP-I transcript variant should contain a partially retained intron sequence and lack the 81-nt 5'UTR that is found upstream of the intron. To test this hypothesis, RT-PCRs were conducted from total pokeweed RNA with several forward primers that span the length of the intron (black arrows) and a single reverse primer (green arrow; **Figure 14D**). Since intron

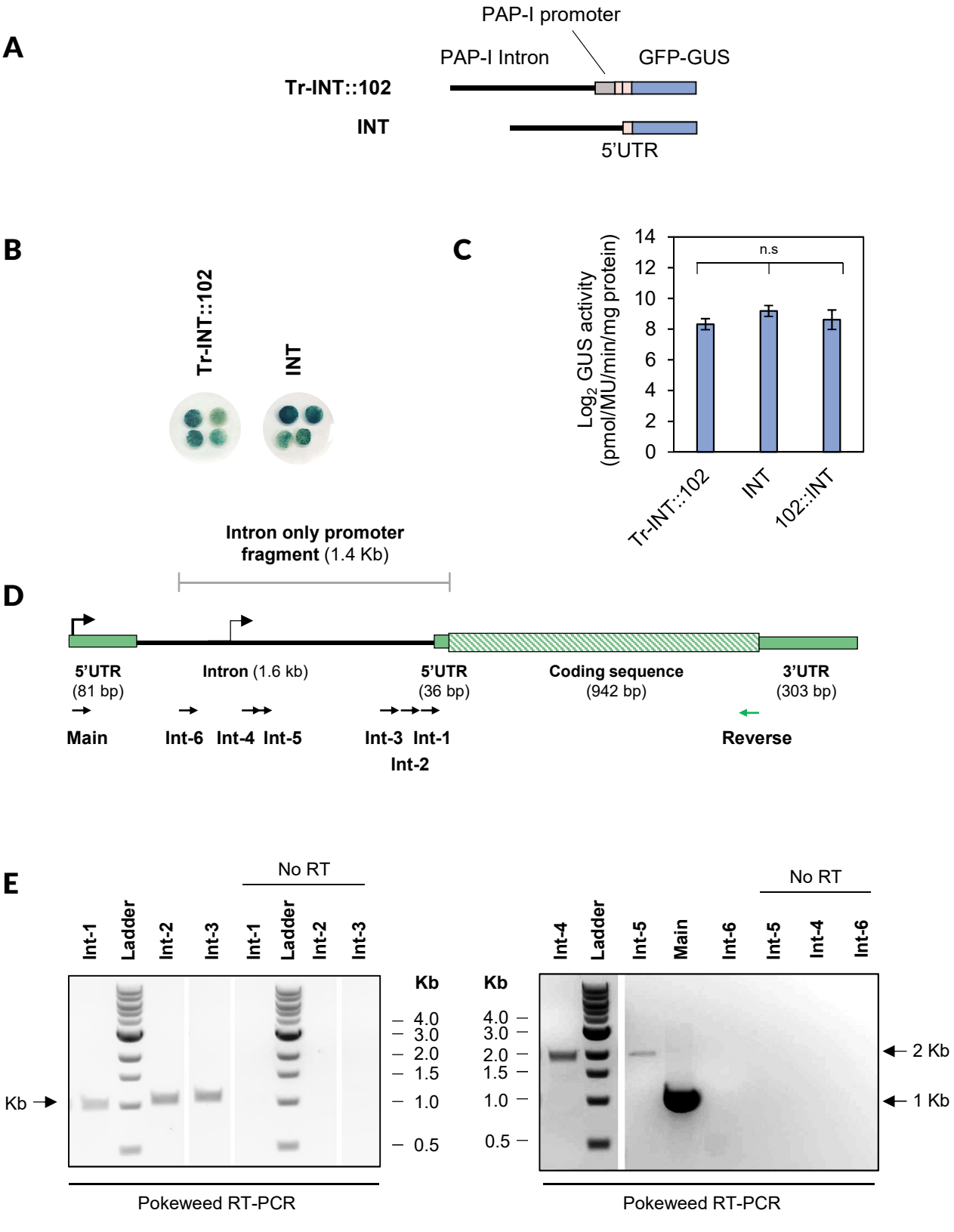

A
PAP-I promoter
PAP-I Intron
GFP-GUS
Tr-INT::102
INT
5'UTR
B
Tr-INT::102
INT
C
n.s
Log2 GUS activity (pmol/MU/min/mg protein)
14
12
10
8
6
4
2
0
Tr-INT::102
INT
102::INT
D
Intron only promoter fragment (1.4 Kb)
5'UTR (81 bp)
Intron (1.6 kb)
5'UTR (36 bp)
Coding sequence (942 bp)
3'UTR (303 bp)
Main
Int-6
Int-4
Int-5
Int-3
Int-1
Int-2
Reverse
E
No RT
Int-1
Ladder
Int-2
Int-3
Int-1
Ladder
Int-2
Int-3
Kb
4.0
3.0
2.0
1.5
1.0
0.5
1.1 Kb
Pokeweed RT-PCR
No RT
Int-4
Ladder
Int-5
Main
Int-6
Int-5
Int-4
Int-6
Kb
4.0
3.0
2.0
1.5
1.0
0.5
2 Kb
1 Kb
Pokeweed RT-PCR

Figure 14. The PAP-I intron acts as an alternative promoter. (A) Schematic representation of translocated intron::102 (tr-INT::102) and intron only (INT) GFP-GUS constructs. **(B)** Histochemical and **(C)** fluorometric assays of GUS activity in tobacco leaves agroinfiltrated with PAP-I promoter::GFP-GUS constructs. At least three independent plants per construct were tested for each experiment (1 leaf disc is 1 biological replicate). Error bars represent the SEM. **(D)** Schematic diagram of the PAP-I gene showing the two transcription start sites. For RT-PCR, the same reverse primer (green arrow) was paired with several forward primers (black arrows) that anneal to different regions of the PAP-I intron using the same complementary DNA (cDNA) as the PCR template. **(E)** RT-PCR from total pokeweed RNA with the primer pairs listed in **C**. Transcription initiation within the PAP-I intron gives rise to a longer PAP-I transcript that contains a portion of the intron sequence.

splicing primarily occurs co-transcriptionally in both plants and animals (Herzel et al., 2017; Jabre et al., 2019), the reverse primer was designed to span the end of the PAP-I CDS and 3'UTR. This strategy ensured that only full-length transcripts were captured instead of nascent transcripts that had not yet undergone splicing. With the exception of Int-6, all forward primers produced a band of the anticipated size, suggesting that a portion of the intron is retained in some PAP-I transcripts, and that the intronic TSS is somewhere between the binding sites for the Int-4 and Int-6 primers (**Figure 14E**). No bands were observed in any of the -RT controls, indicating that the PCR products were not due to gDNA contamination. Based on the intensity of the bands, the alternative transcript is lower in abundance than the main PAP-I transcript. Curiously, the long PAP-I 5'UTR contains multiple predicted upstream open reading frames (uORFs) that range in length from 10 to 49 amino acids (**Figure 15**). Although RT-PCR does not allow for the precise mapping of the TSS within the intron, it demonstrates that the TSS is located at least 1 Kb away from the first AUG of the PAP-I coding sequence, and that this PAP-I mRNA variant therefore contains a long 5'UTR.

3.3.3 The two PAP-I promoters do not work additively to increase total GFP-GUS transcript levels

After discovering that the PAP-I intron can act as a promoter, I wanted to determine how the two PAP-I promoters were contributing to reporter expression at the transcriptional level. GFP-GUS mRNA levels were measured from tobacco leaves expressing different PAP-I::GFP-GUS constructs. To differentiate between GFP-GUS transcript variants, two sets of primers ('GUS' and 'UTR'; **Figure 16A**) that anneal to different regions of the GFP-GUS transcript were used: the first primer set ('GUS') was designed to capture all GFP-GUS transcripts irrespective of their 5'UTRs, while the second primer set ('UTR') was designed to only capture transcripts that had initiated from the canonical PAP-I promoter, and therefore contained a 117-nt 5'UTR. To

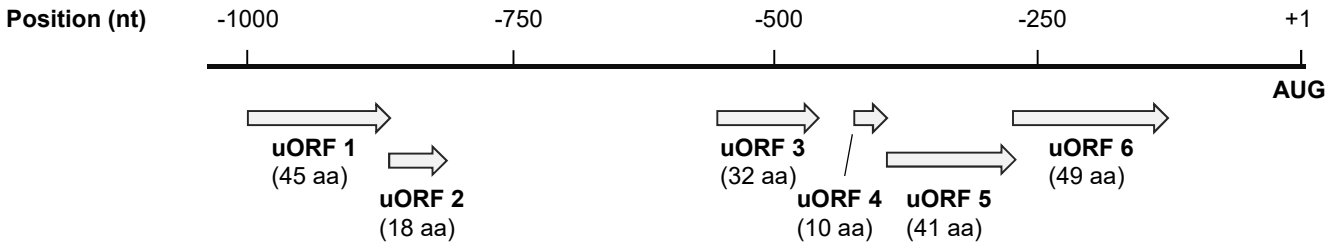


Figure 15. The long PAP-I transcript contains multiple putative upstream open reading frames (uORFs). Schematic diagram of the long PAP-I 5'UTR showing putative uORFs (grey arrows) within the partially retained intron sequence, as predicted by ORFfinder. The length of each uORF (number of amino acids) in indicated within parentheses. Position labels are relative to the first AUG of the main ORF (denoted as +1).

ensure that any differences observed were not due to unequal primer annealing, qPCR was performed with a plasmid template (1262::GFP-GUS-no catalase) that contained the sequences to both qPCR amplicons. The plasmid was serially diluted to 5 orders of magnitude beginning at 0.5 ng. Both amplicons had similar C_q values at all plasmid concentrations tested ($p > 0.05$), indicating that qPCR reactions with the two primer sets have comparable efficiencies when equal amounts of template DNA are present (**Figure 16B**). As shown in **Figure 16C**, mRNA levels differed significantly depending on which primer pair was used, but only for the constructs in which the PAP-I intron was present in its native position. For example, total GFP-GUS transcript levels were similar ($p > 0.05$) in the three 1262 promoter constructs, but transcripts with the 117-nt 5'UTR decreased significantly in 1262::INT, suggesting that a substantial fraction of the mRNA population did not contain this full sequence. Similar trends were observed with 711::INT and 102::INT, although differences between the levels of 'GUS' and 'UTR' amplicons were even more substantial for these constructs. When the intron was either removed (1262, 711) or translocated upstream of the PAP-I promoter (Tr-INT::1262, Tr-INT::102), levels of both the 'GUS' and 'UTR' amplicons were similar, suggesting that most GFP-GUS mRNAs in the pool contained the 117-nt 5'UTR. When only the intronic promoter was present (INT), total GFP-GUS levels were similar to 102::INT and Tr-INT::102. GFP-GUS protein levels (as measured through GUS fluorometric assays) correlated well with total GFP-GUS mRNA levels for most constructs (**Figure 16D**), suggesting that both transcript variants could be translated into functional reporter protein despite the presence of putative uORFs in the 5'UTR of the long GFP-GUS mRNA.

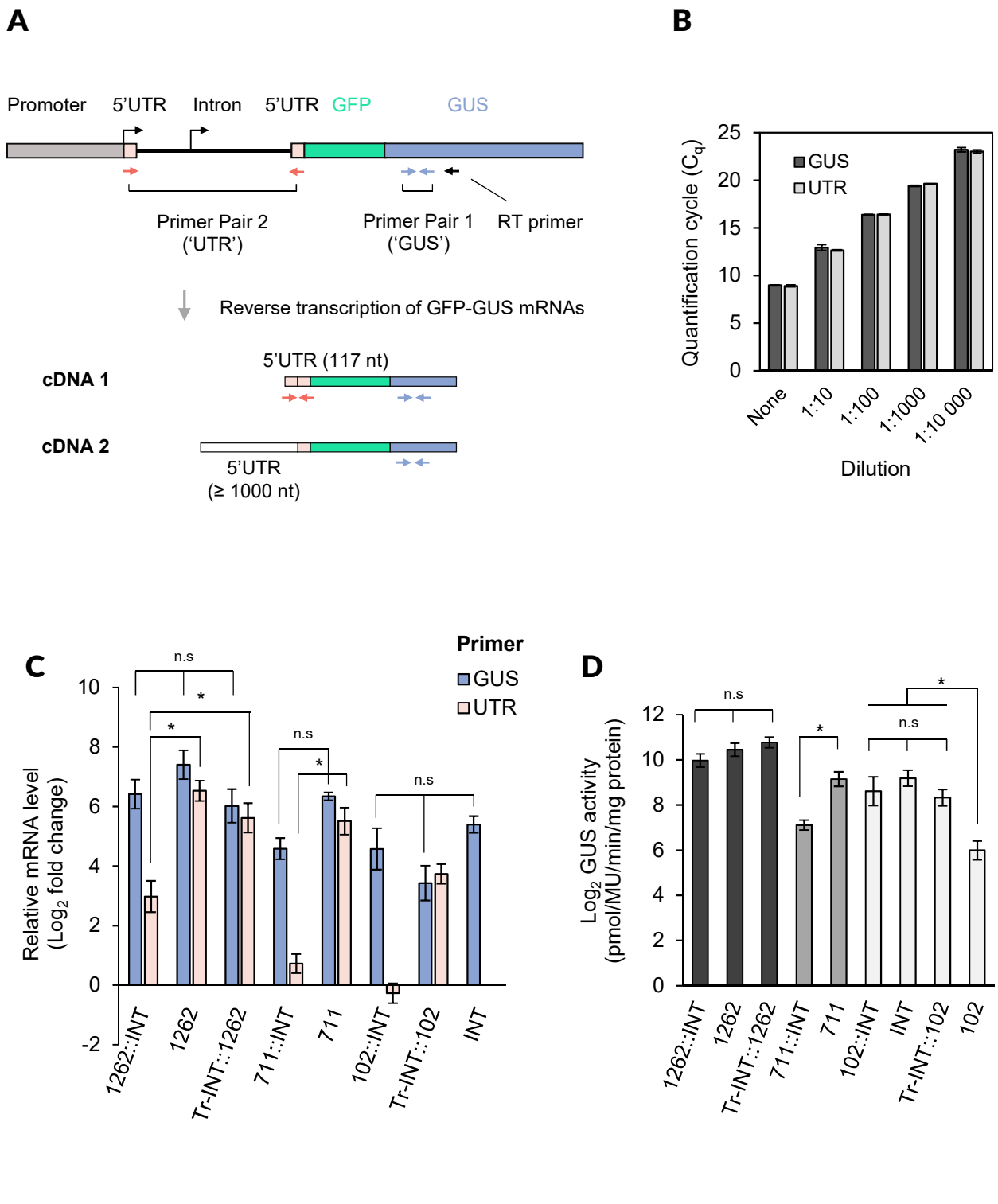

A
Promoter
5'UTR
Intron
5'UTR
GFP
GUS
Primer Pair 2
('UTR')
Primer Pair 1
('GUS')
RT primer
Reverse transcription of GFP-GUS mRNAs
cDNA 1
5'UTR (117 nt)
cDNA 2
5'UTR
(≥ 1000 nt)
B
Quantification cycle (Cq)
GUS
UTR
None
1:10
1:100
1:1000
1:10 000
Dilution
C
Primer
GUS
UTR
Relative mRNA level
(Log2 fold change)
n.s
*
1262::INT
1262
Tr-INT::1262
711::INT
711
102::INT
Tr-INT::102
INT
D
Log2 GUS activity
(pmol/MU/min/mg protein)
n.s
*
1262::INT
1262
Tr-INT::1262
711::INT
711
102::INT
INT
Tr-INT::102
102

Figure 16. The two PAP-I promoters do not work additively to increase total GFP-GUS transcript levels. (A) Schematic of the GFP-GUS reporter gene, showing the positions of the primer pairs ('UTR' and 'GUS', pink and blue arrows, respectively) used in qRT-PCR. The primer used in reverse transcriptase (RT) reactions is depicted as a black arrow. The 'GUS' primer pair can bind to all GFP-GUS cDNAs, while the 'UTR' primer pair can only bind to cDNAs that have the full 117-nt PAP-I 5'UTR. **(B)** qPCR cycle quantification values (C_q) for the two primer pairs listed in **A**. A plasmid that contains the sequences to both amplicons was serially diluted and used as a qPCR template to test primer efficiencies. **(C)** qRT-PCR of tobacco leaves expressing PAP-I promoter::GFP-GUS constructs with the primer pairs listed in **A**. At least four independent plants were tested. Error bars represent the SEM. Ribosomal protein L25 and ubiquitin-conjugating enzyme E2 (Ntubc2) transcripts served as internal controls. Transcripts are expressed as $\log_2$ fold change relative to transcript levels in the 102 construct. **(D)** GUS fluorometric assay of tobacco leaves agroinfiltrated with PAP-I promoter::GFP-GUS constructs (+/- PAP-I intron). At least three independent plants were tested. Error bars represent the SEM. For **C** and **D**, comparisons were made using a one-way ANOVA followed by Tukey's post-hoc tests, $p < 0.05$. '*' = $p < 0.05$; 'n.s.' = not significant.

4. DISCUSSION

4.1 FUNCTIONAL ROLE OF PAP IN POKEWEED

4.1.1 Expression of PAP genes in response to biotic and abiotic stresses

PAP isoforms vary in their ability to depurinate RNA templates (Honjo et al., 2002; Kurinov and Uckun, 2003; Rajamohan et al., 1999) and show different levels of toxicity when expressed in other plants (Dai et al., 2003; Wang et al., 1998), suggesting that they have different activities in pokeweed. PAP isoforms present in the leaf had distinct expression profiles in response to four different types of stresses (JA, SA, PEG, and wounding), and four of the isoforms (PAP-I, PAP-II, PAP-S1, PAP-α) were highly upregulated with JA. In Neller et al. (2019), we gained insight into the functional roles of the PAP isoforms by grouping differentially-expressed pokeweed genes into separate clusters based on their expression profiles. The assignment of the isoforms into separate clusters reinforces their distinct roles in the plant; for example, PAP-II clustered with genes associated with JA biosynthesis and wounding, while PAP-S2 and PAP-α clustered with genes associated with JA signalling and terpenoid biosynthesis. Curiously, the PAP-I cluster contained only nine genes, including two transcription factors from the HD-ZIP homeobox family that regulate plant growth and leaf development in response to abiotic stresses (Aoyama et al., 2007; Söderman et al., 1996). Increased tolerance to biotic or abiotic stresses usually comes at the expense of plant growth (Bechtold and Field, 2018; Pandey et al., 2017). Therefore, PAP-I may be more broadly involved in regulating the balance between defence and growth.

As in pokeweed, RIPs in other plants are typically encoded as distinct protein isoforms by multi-gene families (Chan et al., 2010; Dohm et al., 2014; Urasaki et al., 2017; Wu et al., 2015), and several studies have shown that RIP isoforms within a single species often have distinct

expression profiles. In *Jatropha curcas* (physic nut), the two isoforms that were studied, curcin and curcin-L, had tissue-specific expression profiles; curcin was expressed primarily in the endosperm, while curcin-L was induced in the leaf after PEG treatment (Qin et al., 2010). The RIPs of *Ricinus communis* (castor bean) are endosperm-specific but show distinct expression profiles in the different stages of seed development (Loss-Morais et al., 2013). Furthermore, a genome-wide analysis of RIP expression in *Oryza sativa* (rice) revealed that many of its 31 RIPs were scattered across different tissue types and differentially responsive to an array of stresses, including pathogen infection, PEG, high salinity, and cold (Jiang et al., 2008).

4.1.2 Transcriptional control of PAP expression

The promoters of all PAP genes contained CREs associated with biotic and abiotic stresses, suggesting that PAP is broadly implicated in plant defence. In addition, all PAP promoters contained putative TATA boxes -35 to -25 bps upstream of their TSSs; in *Arabidopsis*, genes associated with biotic and abiotic stresses are more likely to contain TATA elements in their promoters (Kumari and Ware, 2013). When placed upstream of a reporter gene and expressed in tobacco, all tested PAP promoters were able to drive reporter expression to varying degrees, and the strengths of these promoters correlated well with the abundances of their corresponding transcripts in pokeweed. In PAP-II, one promoter (PAP-II_A) appears to be stronger than the other, and the two PAP-II transcripts are not expressed equally in untreated plants (PAP-II_A = 713 TPM; PAP-II_B = 7.7 TPM).

Since PAP genes were most responsive to JA among the four stress treatments tested, I aimed to identify CREs that could mediate this response. Promoter truncation constructs of the PAP-I promoter revealed that a region close to the TSS (-296 to -103) was sufficient for JA-responsiveness. More specifically, a conserved T/GBOXATPIN2 element (-180 to -175 bps relative to the TSS), which is thought to bind the master JA signalling activator MYC, was

required for the JA-mediated upregulation of PAP-I. The presence of a putative MYC binding site in its promoter suggests that PAP-I may be involved in the MYC branch of the JA pathway. Mutation of this element abolished JA-responsiveness of genes in tomato, *Arabidopsis*, and barley (Boter et al., 2004; Rouster et al., 1997). In the case of the PAP-I promoter, reporter expression in the 1262/-GBOX deletion mutant decreased significantly upon JA treatment, suggesting that there may be transcriptional repressors or co-repressors involved in fine-tuning the response. In the absence of the binding site for MYC, however, only negative regulation would occur. Consistent with this hypothesis, genome-wide studies in other eukaryotes have shown that repressors can be associated with actively transcribed regions (Wang et al., 2009; Wirén et al., 2005), and may work alongside other TFs to modulate transcription (reviewed in Reynolds et al., 2013).

Analysis of JA-associated CREs in the promoters of other PAP isoforms revealed that the T/GBOXATPIN2 element was also present in PAP-α and PAP-S2$_B$, but not in PAP-II and PAP-S1, which were highly JA-responsive. The lack of this element in some promoters of JA-responsive isoforms may be compensated by the presence of W-boxes; this element binds WRKY TFs, which are primarily SA-responsive (Dong et al., 2003). However, a substantial fraction of WRKYs (at least 30%) are JA-responsive in *Arabidopsis* (Schluttenhofer et al., 2014), and individual JA-responsive WRKY TFs have been characterized in other plants (Cui et al., 2018; Suttipanta et al., 2011). In pokeweed, we identified several WRKY TFs that were upregulated with JA (homologues of WRKY3, WRKY4, WRKY22, WRKY23, WRKY24, WRKY33, WRKY40, WRKY41, WRKY49, WRKY70, and WRKY75; Neller et al., 2019). Therefore, W-boxes in the promoters of PAP isoforms may contribute to their JA-responsiveness. Additionally, PAP isoforms that were not highly responsive to JA but contained putative JA-associated CREs in their promoters may have expression profiles that were not accurately captured within our single

time point. For example, an RNA-seq time course of the JA response in *Arabidopsis* revealed dynamic expression patterns over the first 16 hours following treatment, including distinct early and late responses (Hickman et al., 2017). In addition to CREs associated with JA, I identified CREs associated with the hormones SA, ABA, and GA. JA and SA have well-established roles in plant defence against pathogens and insect herbivores, while ABA contributes to the resistance of abiotic stresses such as drought, salinity, cold, and heat stress (Verma et al., 2016). GA, through cross-talk with ABA pathways, helps mediate the balance between dormancy and plant maturation during stress (Weiss and Ori, 2007).

4.2 THE PAP-I LEADER INTRON INFLUENCES GENE EXPRESSION

4.2.1f The PAP-I leader intron acts as an alternative promoter

Annotation of the pokeweed genome revealed that the PAP genes contain previously unreported leader introns. Based on the distribution of predicted intron lengths in pokeweed, the PAP leader introns are longer than 83% (PAP-α) to 94% (PAP-II_B) of all introns (**Figure 17**). Interestingly, visual inspection of RIP gene models in sugar beet and quinoa genomes revealed that long leader introns are also present in other RIPs within the same taxonomic order (not shown), although these were not noted in the corresponding publications (Dohm et al., 2014; Jarvis et al., 2017). Therefore, leader introns may be a general feature not only of PAP genes, but of RIPs in other species.

Numerous leader introns have been shown to enhance gene expression in plants and other eukaryotes (reviewed in Gallegos and Rose, 2015; Laxa, 2017). To investigate the effect of the leader intron on PAP-I expression, PAP-I promoter truncations were fused to the GFP-GUS reporter gene either with or without the full-length PAP-I intron. Contrary to my expectations, the leader intron either had no detectable effect on reporter protein expression (1262) or a

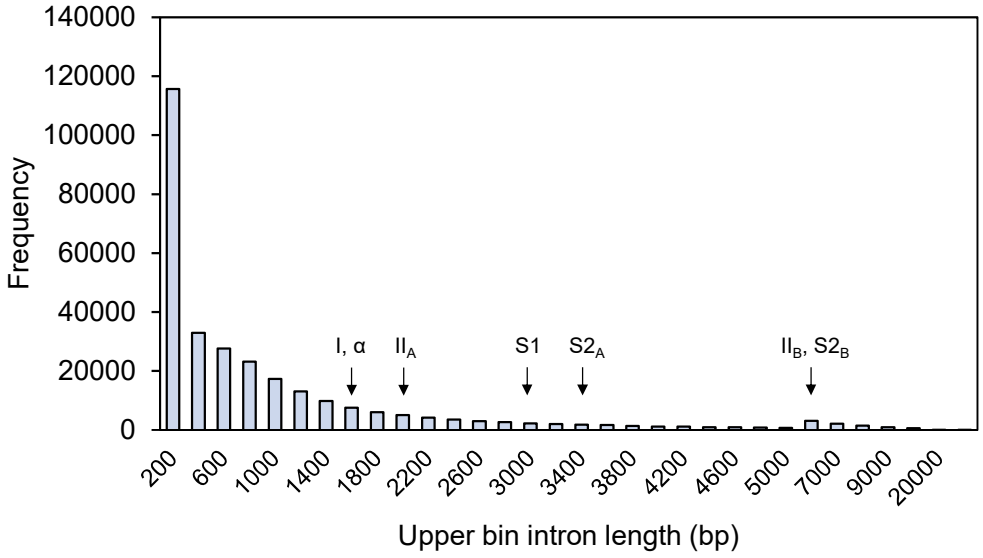


Figure 17. Intron length distribution in pokeweed genes. Pokeweed gene models were generated by the program MAKER using transcriptome and protein evidence. Each bar represents the number of pokeweed introns within an intron length bin (mean = 898 bp; median = 389 bp). Bins containing the PAP leader introns are indicated with black arrows. Letters indicate the names of the PAP genes.

negative effect (1124, 711, 584, 432, and 296). The sole exception was the core promoter fragment (102), which was able to drive higher reporter expression when paired with the intron. This enhancement persisted when the intron was either translocated upstream of the core promoter (Tr-INT::102) or fused to the reporter gene on its own (INT). From these findings, several properties of the PAP-I intron can be inferred: 1) it is not required for high-level expression when longer PAP-I promoter fragments are present; 2) it enhances reporter expression in a position-independent manner only when paired with the weak PAP-I core promoter; 3) it can drive reporter expression in the complete absence of any promoter. Taken together, these results suggest that the PAP-I intron acts as an alternative promoter. While classical enhancers can also promote gene expression in a position-independent manner, as was observed with the 102::INT and Tr-INT::102 fragments, they would not be able to drive the expression of a functional protein in the complete absence of a promoter (Kulaeva et al., 2012), and would have a consistent enhancing effect when paired with variations of the same promoter. In this case, selective enhancement occurs because the PAP-I core promoter is otherwise unable to drive high-level expression on its own, and transcription is driven almost solely by the intronic promoter. When the PAP-I proximal promoter region is included, however, the two promoters do not appear to work concurrently to increase transcription. Potential mechanisms for transcriptional interference have been proposed. If two tandem promoters are in close proximity, as they are in PAP-I, a high density of regulatory factors and transcriptional machinery at one promoter can block transcription from the other due to steric hindrance (Palmer et al., 2011). Additionally, two promoters may compete for a shared distal enhancer (Conte et al., 2002; Schambach et al., 2006), although this is unlikely to be a factor in my experiments because only 1.3 Kb of the PAP-I promoter was present, and both promoters were able to drive high reporter expression when expressed individually.

A model of the interaction between the two promoters is depicted in **Figure 18**. Transcription initiation from the main PAP-I promoter results in the production of a PAP-I transcript with a 117-nt 5'UTR (short PAP-I transcript) in which the 1.6 Kb leader intron is spliced. However, initiation from the intronic promoter results in the production of a PAP-I mRNA with a longer (~1 Kb) 5'UTR. In this case, a portion of the intron sequence is retained because the transcript lacks the 5' sequence needed for splicing. The two promoters do not act together to increase mRNA levels, and the ability of the main promoter to drive expression may be negatively affected by the intronic promoter. Despite the presence of putative uORFs in the long transcript, both transcripts can be translated into functional protein in tobacco, as GUS activity was detected even in leaves that were only expressing the long transcript. uORFs are usually associated with translational inhibition because the presence of upstream start and stop codons can stall ribosomes or trigger the nonsense-mediated decay pathway, resulting in transcript degradation (reviewed in Peccarelli and Kebaara, 2014). However, they can also have a neutral effect in translation, such as when the upstream AUGs are in a weak Kozak context (Wang and Rothnagel, 2004), or even a positive effect, through interactions with the translational machinery that lead to increased translation at the main ORF (Barbosa et al., 2013). Therefore, it is unknown if the two PAP-I transcripts differ in mRNA stability and translational efficiency.

4.2.2 Contribution of the leader intron to PAP-I regulation in pokeweed

In pokeweed leaf tissue, the short PAP-I transcript is present at a much higher abundance than the alternative long transcript, as determined through RT-PCR. Although the long PAP-I mRNA was detected as a low abundance transcript in our *de novo* assembled transcriptome (Neller et al., 2016; our unpublished data), and visual inspection of raw reads from our most recent RNA-seq (Neller et al., 2019) showed that a small fraction of reads mapped to the PAP-I

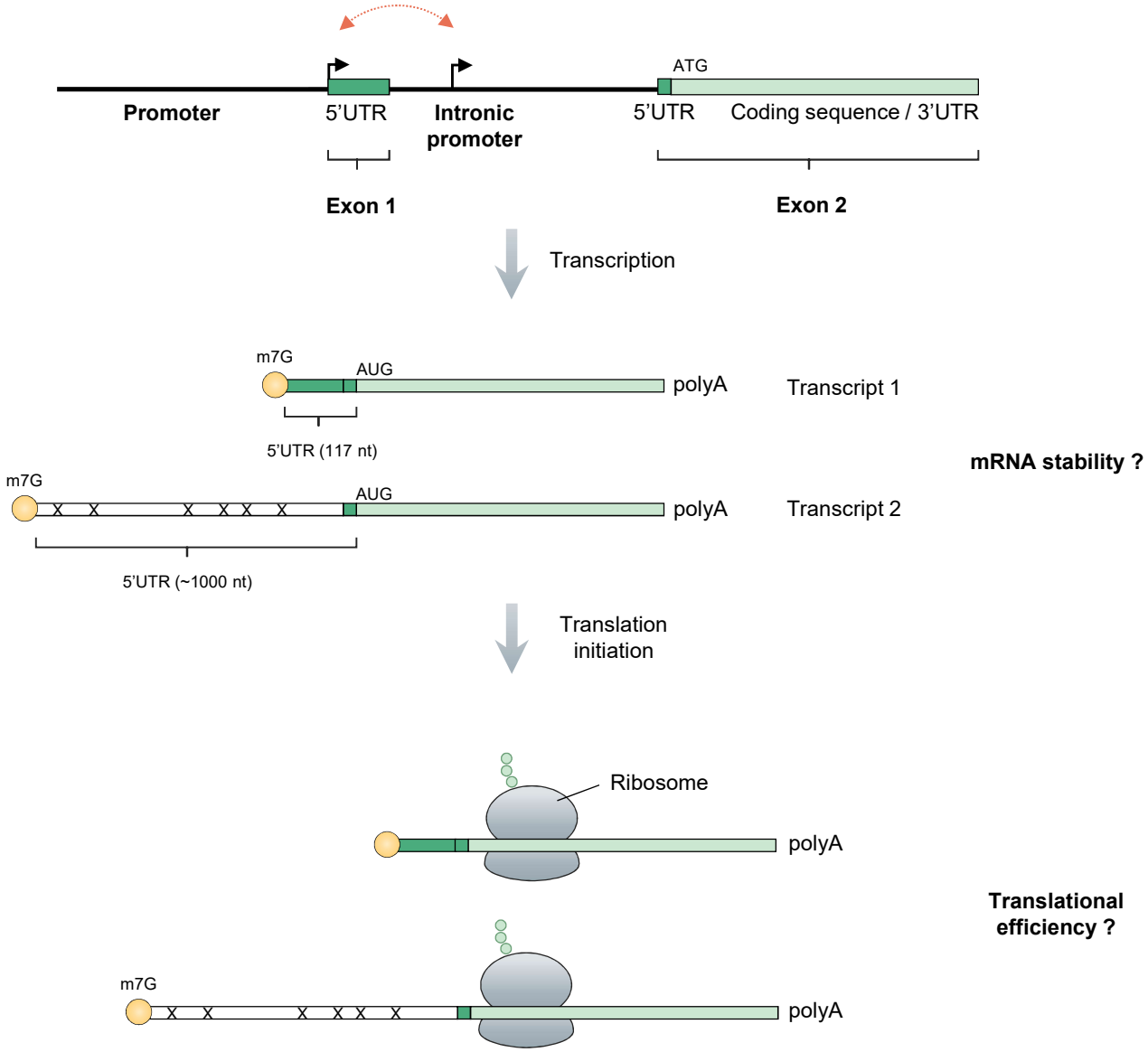


Figure 18. Model of PAP-I expression under two promoters. Transcription initiation from the main PAP-I promoter results in the production of Transcript 1, which has a 117-nt 5'UTR. The 1.6 Kb PAP-I intron is spliced. Transcription initiation from the intronic promoter results in the production of Transcript 2, which has a longer 5'UTR (at least 1 Kb) that is predicted to contain uORFs (shown as X's). In this case, the transcript lacks the 5' splice sequences required for splicing and a portion of the intron is retained. The two promoters do not work additively to increase mRNA levels, and the ability of the main promoter to drive expression may be affected by the presence of the intronic promoter (interaction indicated with dashed red double arrow). Both transcripts can give rise to functional protein in tobacco, although it is unknown how the two distinct 5'UTRs may affect the mRNA stability and translational efficiency of each transcript.

intron region (not shown), the long PAP-I mRNA was not annotated by MAKER as an alternative transcript. This is unsurprising, given that the identification of lesser-expressed transcript variants (such as those arising from alternative splicing or alternative TSSs) still poses a challenge for current bioinformatic programs, and often requires additional downstream procedures (Qin et al., 2018; Song et al., 2019). However, these results clearly indicate that the long PAP-I is a minor transcript variant in pokeweed, unlike in agroinfiltrated tobacco, where the two GFP-GUS mRNA populations are less different in terms of abundance. One possibility is that a distal enhancer in pokeweed may act to enhance transcription initiation at the main promoter over the intronic promoter. The ability of the intronic PAP-I promoter to negatively affect transcription from the main promoter appears to be directly tied to the strength of the main promoter, as simulated through the promoter truncations; therefore, further enhancement at the main PAP-I promoter may decrease the influence of the intronic promoter. In animals, previous studies have shown that enhancer-promoter specificity can be mediated by the presence or absence of different core promoter elements (Juven-Gershon et al., 2008; Zabidi et al., 2015). A recent study on *MYC*, a human proto-oncogene that is controlled by two tandem promoters, found that one promoter is preferentially activated by specific distal enhancers, and enhancer deletion preferentially downregulated the activity of this promoter (Bardales et al., 2018). Importantly, Bardales et al. (2018) also discovered that in the absence of the distal enhancer, the activities of the transiently expressed *MYC* promoters do not recapitulate the differential promoter usage observed in context of the endogenous *MYC* locus.

The presence of leader introns in all PAP genes as well as in other RIPs within the same order lends support to a functional role in the plant. Although the long PAP-I mRNA is not abundant in pokeweed leaf under normal growth conditions, the strength of each PAP-I promoter likely varies across different tissue types, developmental stages, and stimuli, and so the

dynamic interaction between the two PAP-I promoters may dictate the relative proportions of the long and short PAP-I transcripts in different contexts. Studies on plant leader introns have primarily focused on their enhancing properties through IME (Gallegos and Rose, 2015; Parra et al., 2011; Rose et al., 2011). From the perspective of alternative promoter usage, however, it is clear that the presence of multiple promoters does not always lead to a net increase in gene expression; rather, one promoter may be preferentially used over another (Alasoo et al., 2019). In both plants and animals, transcript variants stemming from multiple promoters have been shown to be differentially expressed under a variety of conditions (Bardales et al., 2018; Kurihara et al., 2018; Minegishi et al., 1998; Qi et al., 2007), and genome-wide analyses suggest that alternative promoter usage is widespread (Batut et al., 2013; Pal et al., 2014; Singer et al., 2008). If an alternative promoter is found downstream of the start codon, the protein produced from the alternative transcript may lack certain domains or localization signals, leading to a distinct functional activity (reviewed in Davuluri et al., 2008). In the case of PAP-I, both transcripts should encode identical proteins; however, sequence characteristics within their 5'UTRs, such as the presence of uORFs and secondary structure, may affect their mRNA decay rates or translational efficiencies, leading to differential protein expression. Transcripts with non-identical 5'UTRs can also be differentially targeted by noncoding RNAs (Srivastava et al., 2018).

4.3 CONCLUSIONS

It is well-established that RIPs can be induced by stresses (reviewed in Zhu et al., 2018), and several PAP isoforms have been shown to be responsive to JA (Neller et al., 2016). However, the majority of RIP studies have been conducted in heterologous systems using strong constitutive promoters, and little is known about how they are regulated endogenously, particularly at the transcriptional level. In this study, I identified the PAP promoters and discovered long leader introns in the PAP genes. Through a combination of *in silico* CRE

analysis and reporter gene assays in tobacco, a five-nucleotide G-BOX element in the PAP-I promoter was found to be essential for its JA-mediated upregulation. Furthermore, I showed that both the PAP-I promoter and intron can independently drive gene expression in tobacco, leading to the transcription of mRNA variants with distinct 5'UTRs. Preliminary results suggest that transcription from one promoter may negatively affect transcription from the other, although it is presently unclear how this inhibition may occur, and if distal elements may mediate this interaction.

Beyond learning about the endogenous regulation of PAP and RIPs, this research may also have relevance in applied biotechnology; while transgenic plants synthesizing PAP heterologously gained novel antiviral and antifungal properties, they also showed reduced growth and germination (Lodge et al., 1993; Wang et al., 1998; Zoubenko et al., 1997). These undesirable phenotypes were attributed to the expression of PAP, which had been placed under the control of a strong, constitutive viral promoter. Placing PAP under the control of its native promoter and intron may lead to a more controlled response that does not trade off pathogen resistance for partial cytotoxicity.

4.4 FUTURE DIRECTIONS

In Neller et al. (2019), we showed that the PAP isoforms have distinct expression profiles in response to four types of stresses (JA, SA, PEG, and wounding), and that most isoforms were highly responsive to JA. To gain more insight into their specific roles in pokeweed, this analysis can be extended with additional stress treatments, developmental stages, and tissue types. Even within a single stress treatment, different genes can be early or late responders (Hickman et al., 2017); therefore, performing a detailed RNA-seq or qRT-PCR time course of the JA response in pokeweed may help clarify their roles within the JA pathway. Moreover, the availability of complete coding sequences for all PAP genes means that each isoform can be expressed and

purified individually in *E. coli*, potentially leading to the generation of isoform-specific PAP antibodies. This will allow us to correlate PAP transcript levels with their respective protein abundances. These experiments can be complemented with the stable expression of PAP promoter-reporter genes in transgenic tobacco or *Arabidopsis*, which will allow us to determine how their promoters regulate their expression under a more diverse set of conditions.

In the second portion of my project, I showed that a previously undiscovered PAP-I leader intron can act as an alternative promoter, leading to the transcription of an mRNA variant that contains a longer 5'UTR. RT-PCR results suggest that its 5'UTR is at least 1 Kb in length; however, 5'RACE or primer extension will need to be performed in order to precisely map its TSS and promoter region. When paired with the PAP-I promoter, the intronic promoter did not increase overall reporter expression in tobacco. While qRT-PCR results suggest that both the long and short GFP-GUS transcript variants were expressed, the ability of the PAP-I promoter to drive reporter expression may be negatively affected by the presence of the intronic promoter. It is important to note that qRT-PCR is only an indirect measure of transcription, as the amount of mRNA depends on both the rates of transcription and mRNA degradation (Bremer and Moyes, 2014; Das et al., 2017). Nuclear run-on assay and its high-throughput equivalent global run-on sequencing (GRO-seq) are used to specifically capture nascent RNAs, and can therefore provide a measure of transcription initiation independent of mRNA stability (Lopes et al., 2017). Stability of the long and short PAP-I mRNAs can in turn be assessed by expressing the two transcripts separately in tobacco protoplasts (under the control of the same constitutive promoter to ensure equal rates of transcription), treating cells with a transcription inhibitor such as cordycepin or actinomycin D, and quantifying PAP-I transcript abundance over time. Finally, the translational efficiencies of the two PAP-I transcripts can be measured *in vitro* by using equimolar amounts of RNA in cell-free translation assays. While both the long and short GFP-GUS transcripts could

give rise to functional reporter protein in tobacco, it is difficult to infer translational efficiency from those experiments because the two mRNAs may have also differed in abundance. For the mRNA stability and *in vitro* translation assays, a previously characterized PAP-I active site mutant (Hur et al., 1995) can be used to ensure that it does not depurinate ribosomes when translated. Collectively, these experiments will tease apart the interplay between transcription rates, mRNA stability, and translational efficiency, and shed light on precisely how the intronic promoter is contributing to PAP-I expression.

In pokeweed, the main promoter and the intronic promoter likely work in tandem to control PAP-I expression, and distal regulatory elements may contribute to the preferential usage of one promoter over the other. While the main PAP-I promoter is preferentially used in leaf tissue under normal growth, transcript variants originating from multiple promoters are often differentially expressed in different contexts. Therefore, the two sections of my project can be integrated by: 1) measuring the expression of both PAP-I transcripts in pokeweed under a variety of stress treatments, tissue types, and developmental stages; 2) determining how promoter usage changes under these conditions through stable tobacco transgenic lines that express PAP-I promoter-reporter gene constructs with the intron. Together, these experiments will lead to a more nuanced understanding of endogenous PAP regulation.

5. REFERENCES

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscisic Acid Signaling. *Plant Cell* 15, 63–78. doi:10.1105/tpc.006130.

Agarwal, M., Hao, Y., Kapoor, A., Dong, C. H., Fujii, H., Zheng, X., et al. (2006). A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *J. Biol. Chem.* 281, 37636–45. doi:10.1074/jbc.M605895200.

Akua, T., Berezin, I., and Shaul, O. (2010). The leader intron of AtMHX can elicit, in the absence of splicing, low-level intron-mediated enhancement that depends on the internal intron sequence. *BMC Plant Biol.* 10, 93. doi:10.1186/1471-2229-10-93.

Alasoo, K., Rodrigues, J., Danesh, J., Freitag, D. F., Paul, D. S., and Gaffney, D. J. (2019). Genetic effects on promoter usage are highly context-specific and contribute to complex traits. *Elife* 9, e41673. doi:10.7554/eLife.41673.

Anderson, J. P., Badruzsaufari, E., Schenk, P. M., Manners, J. M., Desmond, O. J., Ehlert, C., et al. (2004). Antagonistic Interaction between Abscisic Acid and Jasmonate-Ethylene Signaling Pathways Modulates Defense Gene Expression and Disease Resistance in Arabidopsis. *Plant Cell* 16, 3460–79. doi:10.1105/tpc.104.025833.

Aoyama, T., Dong, C.-H., Wu, Y., Carabelli, M., Sessa, G., Ruberti, I., et al. (2007). Ectopic Expression of the Arabidopsis Transcriptional Activator Athb-1 Alters Leaf Cell Fate in Tobacco. *Plant Cell* 7, 1773. doi:10.2307/3870186.

Barbieri, L., Aron, G. M., Irvin, J. D., and Stirpe, F. (1982). Purification and partial characterization of another form of the antiviral protein from the seeds of Phytolacca americana L. (pokeweed). *Biochem. J.* 203, 55–59.

Barbieri, L., Bolognesi, A., Cenini, P., Falasca, A. I., Minghetti, A., Garofano, L., et al. (1989). Ribosome-inactivating proteins from plant cells in culture. *Biochem. J.* 257, 801–807. doi:10.1042/bj2570801.

Barbosa, C., Peixeiro, I., and Romão, L. (2013). Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLoS Genet.* 9, e1003529. doi:10.1371/journal.pgen.1003529.

Bardales, J. A., Wieser, E., Kawaji, H., Murakawa, Y., and Darzacq, X. (2018). Selective activation of alternative MYC core promoters by Wnt-responsive enhancers. *Genes (Basel).*, E270. doi:10.3390/genes9060270.

Batut, P., Dobin, A., Plessy, C., Carninci, P., and Gingeras, T. R. (2013). High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res.* doi:10.1101/gr.139618.112.

Bechtold, U., and Field, B. (2018). Molecular mechanisms controlling plant growth during abiotic stress. *J. Exp. Bot.* 69, 2753–58. doi:10.1093/jxb/ery157.

Bolognesi, A., Barbieri, L., Abbondanza, A., Falasca, A. I., Carnicelli, D., Battelli, M. G., et al. (1990). Purification and properties of new ribosome-inactivating proteins with RNA N-glycosidase activity. *Biochim. Biophys. Acta* 1087, 293–302. doi:10.1016/0167-4781(90)90002-J.

Boter, M., Ruíz-Rivero, O., Abdeen, A., and Prat, S. (2004). Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. *Genes Dev.* doi:10.1101/gad.297704.

Bourdon, V., Harvey, A., and Lonsdale, D. M. (2001). Introns and their positions affect the translational activity of mRNA in plant cells. *EMBO Rep.* 2, 394–98. doi:10.1093/embo-reports/kve090.

Bremer, K., and Moyes, C. D. (2014). mRNA degradation: an underestimated factor in steady-state transcript levels of cytochrome c oxidase subunits? *J. Exp. Biol.* 217, 2212–20. doi:10.1242/jeb.100214.

Brown-Borg, H. M. (2007). Hormonal regulation of longevity in mammals. *Ageing Res. Rev.* 6, 28–45. doi:10.1016/j.arr.2007.02.005.

Buchel, A. S., Brederode, F. T., Bol, J. F., and Linthorst, H. J. M. (1999). Mutation of GT-1 binding sites in the Pr-1a promoter influences the level of inducible gene expression in vivo. *Plant Mol. Biol.* 40, 387–96. doi:10.1023/A:1006144505121.

Cercós, M., Gómez-Cadenas, A., and Ho, T. H. D. (1999). Hormonal regulation of a cysteine proteinase gene, EPB-1, in barley aleurone layers: Cis- and trans-acting elements involved in the co-ordinated gene expression regulated by gibberellins and abscisic acid. *Plant J.* 19, 107–118. doi:10.1046/j.1365-313X.1999.00499.x.

Chakravarthy, S., Tuori, R., D'Ascenzo, M., Fobert, P., Despres, C., and Martin, G. (2003). The Tomato Transcription Factor Pti4 Regulates Defense-Related Gene Expression via GCC Box and Non-GCC Box cis Elements. *Plant Cell* 15, 3033–50. doi:10.1105/tpc.017574.

Chan, A. P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., et al. (2010). Draft genome sequence of the oilseed species Ricinus communis. *Nat. Biotechnol.* 28, 951–6. doi:10.1038/nbt.1674.

Cheng, Y. T., Germain, H., Wiermer, M., Bi, D., Xu, F., García, A. V., et al. (2009). Nuclear Pore Complex Component MOS7/Nup88 Is Required for Innate Immunity and Nuclear Accumulation of Defense Regulators in Arabidopsis. *Plant Cell* 21, 2503–16. doi:10.1105/tpc.108.064519.

Chilton, M. D., Drummond, M. H., Merlo, D. J., Sciaky, D., Montoya, A. L., Gordon, M. P., et al.

(1977). Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell* 11, 263–71. doi:10.1016/0092-8674(77)90043-5.

Chow, C. N., Zheng, H. Q., Wu, N. Y., Chien, C. H., Huang, H. Da, Lee, T. Y., et al. (2016). PlantPAN 2.0: An update of Plant Promoter Analysis Navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* 44, D1154-60. doi:10.1093/nar/gkv1035.

Chuang, W. P., Herde, M., Ray, S., Castano-Duque, L., Howe, G. A., and Luthe, D. S. (2014). Caterpillar attack triggers accumulation of the toxic maize protein RIP2. *New Phytol.* 201, :928-39. doi:10.1111/nph.12581.

Chung, B. Y. W., Simons, C., Firth, A. E., Brown, C. M., and Hellens, R. P. (2006). Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. *BMC Genomics* 7, 120. doi:10.1186/1471-2164-7-120.

Conte, C., Dastugue, B., and Vaury, C. (2002). Promoter competition as a mechanism of transcriptional interference mediated by retrotransposons. *EMBO J.* 21, 3908–16. doi:10.1093/emboj/cdf367.

Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* 322, 1845–48. doi:10.1126/science.1162228.

Côté, C., and Rutledge, R. G. (2003). An improved MUG fluorescent assay for the determination of GUS activity within transgenic tissue of woody plants. *Plant Cell Rep.* 21, 619–24. doi:10.1007/s00299-002-0543-z.

Cui, Q., Yan, X., Gao, X., Zhang, D. M., He, H. B., and Jia, G. X. (2018). Analysis of WRKY transcription factors and characterization of two Botrytis cinerea-responsive LrWRKY genes from Lilium regale. *Plant Physiol. Biochem.* 127, 525–36. doi:10.1016/j.plaphy.2018.04.027.

Dai, W. D., Bonos, S., Guo, Z., Meyer, W. A., Day, P. R., and Belanger, F. C. (2003). Expression of pokeweed antiviral proteins in creeping bentgrass. *Plant Cell Rep.* 21, 497–502. doi:10.1007/s00299-002-0534-0.

Das, S., Sarkar, D., and Das, B. (2017). The interplay between transcription and mRNA degradation in Saccharomyces cerevisiae. *Microb. Cell* 4, 212–228. doi:10.15698/mic2017.07.580.

Daszkowska-Golec, A., and Szarejko, I. (2013). Open or Close the Gate – Stomata Action Under the Control of Phytohormones in Drought Stress Conditions. *Front. Plant Sci.* 4, 138. doi:10.3389/fpls.2013.00138.

Davuluri, R. V., Suzuki, Y., Sugano, S., Plass, C., and Huang, T. H. M. (2008). The functional

consequences of alternative promoter use in mammalian genomes. *Trends Genet.* 24, 167–177. doi:10.1016/j.tig.2008.01.008.

Di Maro, A., Citores, L., Russo, R., Iglesias, R., and Ferreras, J. M. (2014). Sequence comparison and phylogenetic analysis by the Maximum Likelihood method of ribosome-inactivating proteins from angiosperms. *Plant Mol. Biol.* 85, 575–88. doi:10.1007/s11103-014-0204-y.

Ding, Y., Sun, T., Ao, K., Peng, Y., Zhang, Y., Li, X., et al. (2018). Opposite Roles of Salicylic Acid Receptors NPR1 and NPR3/NPR4 in Transcriptional Regulation of Plant Immunity. *Cell* 173, 1454–1467. doi:10.1016/j.cell.2018.03.044.

Dohm, J. C., Minoche, A. E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., et al. (2014). The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). *Nature* 505, 546–9. doi:10.1038/nature12817.

Dong, J., Chen, C., and Chen, Z. (2003). Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. *Plant Mol. Biol.* 51, 21–37. doi:10.1023/A:1020780022549.

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

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al. (2003). OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* 33, 751–63. doi:10.1046/j.1365-313X.2003.01661.x.

Dunn, M. A., White, A. J., Vural, S., and Hughes, M. A. (1998). Identification of promoter elements in a low-temperature-responsive gene (blt4.9) from barley (Hordeum vulgare L.). *Plant Mol. Biol.* 38, 551–64. doi:10.1023/A:1006098132352.

Endo, Y., Tsurugi, K., and 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. *Biochem. Biophys. Res. Commun.* 150, 1032–1036. doi:10.1016/0006-291X(88)90733-4.

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., et al. (2008). Mediator-Dependent Recruitment of TFIIH Modules in Preinitiation Complex. *Mol. Cell* 31, 337–346. doi:10.1016/j.molcel.2008.06.021.

Eyboulet, F., Wydau-Dematteis, S., Eychenne, T., Alibert, O., Neil, H., Boschiero, C., et al. (2015). Mediator independently orchestrates multiple steps of preinitiation complex assembly in vivo. *Nucleic Acids Res.* 43, 9214–9231. doi:10.1093/nar/gkv782.

Ezcurra, I., Ellerström, M., Wycliffe, P., Stålberg, K., and Rask, L. (1999). Interaction between

composite elements in the napA promoter: Both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol. Biol.* doi:10.1023/A:1006206124512.

Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J.-M., Gimenez-Ibanez, S., Geerinck, J., et al. (2011). The *Arabidopsis* bHLH Transcription Factors MYC3 and MYC4 Are Targets of JAZ Repressors and Act Additively with MYC2 in the Activation of Jasmonate Responses. *Plant Cell* 23, 701–715. doi:10.1105/tpc.110.080788.

Fonseca, S., Chini, A., Hamberg, M., Adie, B., Porzel, A., Kramell, R., et al. (2009). (+)-7-iso-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. *Nat. Chem. Biol.* 5, 344–50. doi:10.1038/nchembio.161.

Furihata, T., Maruyama, K., Fujita, Y., Umezawa, T., Yoshida, R., Shinozaki, K., et al. (2006). Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. *Proc. Natl. Acad. Sci.* 103, 1988–1993. doi:10.1073/pnas.0505667103.

Gallegos, J. E., and Rose, A. B. (2015). The enduring mystery of intron-mediated enhancement. *Plant Sci.* 237, 8–15. doi:10.1016/j.plantsci.2015.04.017.

Garfinkel, D. J., Simpson, R. B., Ream, L. W., White, F. F., Gordon, M. P., and Nester, E. W. (1981). Genetic analysis of crown gall: Fine structure map of the T-DNA by site-directed mutagenesis. *Cell* 27, 143–153. doi:10.1016/0092-8674(81)90368-8.

Geertz, M., and Maerkl, S. J. (2010). Experimental strategies for studying transcription factor-DNA binding specificities. *Brief. Funct. Genomics* 9, 362–373. doi:10.1093/bfgp/elq023.

Gessner, S. L., and Irvin, J. D. (1980). Inhibition of elongation factor 2-dependent translocation by the pokeweed antiviral protein and ricin. *J. Biol. Chem.* 255, 3251–53.

Gu, Y. Q., Wildermuth, M. C., Chakravarthy, S., Loh, Y. T., Yang, C., He, X., et al. (2002). Tomato transcription factors pti4, pti5, and pti6 activate defense responses when expressed in Arabidopsis. *Plant Cell* 14, 817–31. doi:10.1105/tpc.000794.

Gubler, F., Kalla, R., Roberts, J. K., and Jacobsen, J. V. (1995). Gibberellin-Regulated Expression of a myb Gene in Barley Aleurone Cells: Evidence for Myb Transactivation of a High-pl alpha-Amylase Gene Promoter. *Plant Cell* 7, 1879–91. doi:10.1105/tpc.7.11.1879.

He, Y.-W., Guo, C.-X., Pan, Y.-F., Peng, C., and Weng, Z.-H. (2008). Inhibition of hepatitis B virus replication by pokeweed antiviral protein in vitro. *World J. Gastroenterol.* 14, 1592–1597. doi:10.3748/wjg.14.1592.

Hernandez-Garcia, C. M., and Finer, J. J. (2014). Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* 217–218, 109–19. doi:10.1016/j.plantsci.2013.12.007.

Herzel, L., Ottoz, D. S. M., Alpert, T., and Neugebauer, K. M. (2017). Splicing and transcription

touch base: Co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* 18, 637–650. doi:10.1038/nrm.2017.63.

Hetzel, J., Duttke, S. H., Benner, C., and Chory, J. (2016). Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci.* 113, 12316–21. doi:10.1073/pnas.1603217113.

Hickman, R., van Verk, M. C., Van Dijken, A. J. H., Pereira Mendes, M., Vroegop-Vos, I. A., Caarls, L., et al. (2017). Architecture and Dynamics of the Jasmonic Acid Gene Regulatory Network. *Plant Cell* 29, 2086–2105. doi:10.1105/tpc.16.00958.

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27, 297–300. doi:10.1093/nar/27.1.297.

Hobo, T., Asada, M., Kowyama, Y., and Hattori, T. (1999). ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. *Plant J.* 19, 679–89. doi:10.1046/j.1365-313X.1999.00565.x.

Hong, X., Scofield, D. G., and Lynch, M. (2006). Intron size, abundance, and distribution within untranslated regions of genes. *Mol. Biol. Evol.* 23, 2392–2404. doi:10.1093/molbev/msl111.

Honjo, E., Dong, D., Motoshima, H., and 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, 225–31. doi:10.1093/oxfordjournals.jbchem.a003092.

Hur, Y., Hwang, D. J., Zoubenko, O., Coetzer, C., Uckun, F. M., and Tumer, N. E. (1995). Isolation and characterization of pokeweed antiviral protein mutations in Saccharomyces cerevisiae: identification of residues important for toxicity. *Proc. Natl. Acad. Sci.* 92, 8448–52. doi:10.1073/pnas.92.18.8448.

Hwang, H.-H., Yu, M., and Lai, E.-M. (2017). Agrobacterium-mediated plant transformation: biology and applications. *Arab. B.* 15, e0186. doi:10.1199/tab.0186.

Iglesias, R., Pérez, Y., De Torre, C., Ferreras, J. M., Antolín, P., Jiménez, P., et al. (2005). Molecular characterization and systemic induction of single-chain ribosome-inactivating proteins (RIPs) in sugar beet (Beta vulgaris) leaves. *J. Exp. Bot.* 56, 1675–1684. doi:10.1093/jxb/eri164.

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.

Jabre, I., Reddy, A. S. N., Kalyna, M., Chaudhary, S., Khokhar, W., Byrne, L. J., et al. (2019). Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Res.* 47, 2716–2726. doi:10.1093/nar/gkz121.

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J. A., et al. (2017). The

genome of Chenopodium quinoa. *Nature* 542, 307–312. doi:10.1038/nature21370.

Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6, 3901–3907. doi:10.1002/j.1460-2075.1987.tb02730.x.

Jeong, J. Y., Yim, H. S., Ryu, J. Y., Lee, H. S., Lee, J. H., Seen, D. S., et al. (2012). One-step sequence-and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Appl. Environ. Microbiol.* 78, 5440–5443. doi:10.1128/AEM.00844-12.

Jeong, Y. M., Mun, J. H., Lee, I., Woo, J. C., Hong, C. B., and Kim, S. G. (2006). Distinct Roles of the First Introns on the Expression of Arabidopsis Profilin Gene Family Members. *Plant Physiol.* 140, 196–209. doi:10.1104/pp.105.071316.

Jiang, S. Y., Ramamoorthy, R., Bhalla, R., Luan, H. F., Venkatesh, P. N., Cai, M., et al. (2008). Genome-wide survey of the RIP domain family in Oryza sativa and their expression profiles under various abiotic and biotic stresses. *Plant Mol. Biol.* 67, 603–14. doi:10.1007/s11103-008-9342-4.

Juven-Gershon, T., Hsu, J. Y., and Kadonaga, J. T. (2008). Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. *Genes Dev.* 22, 2823–30. doi:10.1101/gad.1698108.

Kanzaki, H., Kagemori, T., Yamachika, Y., Nitoda, T., and Kawazu, K. (2005). Inhibition of Plant Transformation by Phytolaccoside B from Phytolacca americana Callus. *Biosci. Biotechnol. Biochem.* 63, 1657–59. doi:10.1271/bbb.63.1657.

Kaplan, B., Davydov, O., Knight, H., Galon, Y., Knight, M. R., Fluhr, R., et al. (2006). Rapid Transcriptome Changes Induced by Cytosolic Ca2+ Transients Reveal ABRE-Related Sequences as Ca2+-Responsive cis Elements in Arabidopsis. *Plant Cell* 18, 2733–48. doi:10.1105/tpc.106.042713.

Karran, R. A., and Hudak, K. A. (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. doi:10.1093/nar/gkn896.

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

Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., et al. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408, 796–815. doi:10.1038/35048692.

Kazan, K., and Manners, J. M. (2013). MYC2: The master in action. *Mol. Plant* 6, 686–703. doi:10.1093/mp/sss128.

Keane, P. A., and Seoighe, C. (2016). Intron Length Coevolution across Mammalian Genomes.

Mol. Biol. Evol. 33, 2682–91. doi:10.1093/molbev/msw151.

Kikkert, J. R., Vidal, J. R., and Reisch, B. I. (2005). Stable transformation of plant cells by particle bombardment/biolistics. *Methods Mol. Biol.* 286, 61–78.

Kim, H. J., Kim, Y. K., Park, J. Y., and Kim, J. (2002). Light signalling mediated by phytochrome plays an important role in cold-induced gene expression through the C-repeat/dehydration responsive element (C/DRE) in Arabidopsis thaliana. *Plant J.* 29, 693–704. doi:10.1046/j.1365-313X.2002.01249.x.

Kim, M. J., Kim, H., Shin, J. S., Chung, C. H., Ohlrogge, J. B., and Suh, M. C. (2006). Seed-specific expression of sesame microsomal oleic acid desaturase is controlled by combinatorial properties between negative cis-regulatory elements in the SeFAD2 promoter and enhancers in the 5′-UTR intron. *Mol. Genet. Genomics* 276, 351–368. doi:10.1007/s00438-006-0148-2.

Kim, S. I., Veena, and Gelvin, S. B. (2007). Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. *Plant J.* 51, 779–91. doi:10.1111/j.1365-313X.2007.03183.x.

Kim, S. Y., Chung, H. J., and Thomas, T. L. (1997). Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the Dc3 promoter using a modified yeast one-hybrid system. *Plant J.* 11, 1237–51. doi:10.1046/j.1365-313X.1997.11061237.x.

Kim, Y. J., Björklund, S., Li, Y., Sayre, M. H., and Kornberg, R. D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77, 599–608. doi:10.1016/0092-8674(94)90221-6.

Kizis, D., and Pagès, M. (2002). Maize DRE-binding proteins DBF1 and DBF2 are involved in rab17 regulation through the drought-responsive element in an ABA-dependent pathway. *Plant J.* 30, 679–89. doi:10.1046/j.1365-313X.2002.01325.x.

Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., et al. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* 34, 192–198. doi:10.1038/nbt.3450.

Krenek, P., Samajova, O., Luptovciak, I., Doskocilova, A., Komis, G., and Samaj, J. (2015). Transient plant transformation mediated by Agrobacterium tumefaciens: Principles, methods and applications. *Biotechnol. Adv.* 33, 1024–42. doi:10.1016/j.biotechadv.2015.03.012.

Krivdova, G., and Hudak, K. A. (2015). Pokeweed antiviral protein restores levels of cellular APOBEC3G during HIV-1 infection by depurinating Vif mRNA. *Antiviral Res.* 122, 51–54. doi:10.1016/j.antiviral.2015.08.007.

Kulaeva, O. I., Nizovtseva, E. V., Polikanov, Y. S., Ulianov, S. V., and Studitsky, V. M. (2012).

Distant Activation of Transcription: Mechanisms of Enhancer Action. *Mol. Cell. Biol.* 32, 4892–7. doi:10.1128/mcb.01127-12.

Kumari, S., and Ware, D. (2013). Genome-wide computational prediction and analysis of core promoter elements across plant monocots and dicots. *PLoS One* 8. doi:10.1371/journal.pone.0079011.

Kurihara, Y., Makita, Y., Kawashima, M., Fujita, T., Iwasaki, S., and Matsui, M. (2018). Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in Arabidopsis. *Proc. Natl. Acad. Sci.* 115, 7831–36. doi:10.1073/pnas.1804971115.

Kurinov, I. V., and Uckun, F. M. (2003). High resolution X-ray structure of potent anti-HIV pokeweed antiviral protein-III. *Biochem. Pharmacol.* 65, 1709–1717. doi:10.1016/S0006-2952(03)00144-8.

Kutach, A. K., and Kadonaga, J. T. (2002). The Downstream Promoter Element DPE Appears To Be as Widely Used as the TATA Box in Drosophila Core Promoters. *Mol. Cell. Biol.* 20, 4754–64. doi:10.1128/mcb.20.13.4754-4764.2000.

Lackman, P., Gonzalez-Guzman, M., Tilleman, S., Carqueijeiro, I., Perez, A. C., Moses, T., et al. (2011). Jasmonate signaling involves the abscisic acid receptor PYL4 to regulate metabolic reprogramming in Arabidopsis and tobacco. *Proc. Natl. Acad. Sci.* 108, 5891–96. doi:10.1073/pnas.1103010108.

Laloi, C., Mestres-Ortega, D., Marco, Y., Meyer, Y., and Reichheld, J. (2004). The Arabidopsis Cytosolic Thioredoxin h5 Gene Induction by Oxidative Stress and Its W-Box-Mediated Response to Pathogen Elicitor. *Plant Physiol.* 134, 1006–1016. doi:10.1104/pp.103.035782.

Laxa, M. (2017). Intron-Mediated Enhancement: A Tool for Heterologous Gene Expression in Plants? *Front. Plant Sci.* 7:1977. doi:10.3389/fpls.2016.01977.

Lee, L.-Y., and Gelvin, S. B. (2008). T-DNA Binary Vectors and Systems. *Plant Physiol.* 146, 325–332. doi:10.1104/pp.107.113001.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262.

Lodge, J. K., Kaniewski, W. K., and Tumer, N. E. (1993). Broad-spectrum virus resistance in transgenic plants expressing pokeweed antiviral protein. *Proc. Natl. Acad. Sci. U. S. A.* 90, 7089–7093. doi:10.1073/pnas.90.15.7089.

Lopes, R., Agami, R., and Korkmaz, G. (2017). GRO-seq, a tool for identification of transcripts regulating gene expression. *Methods Mol. Biol.* 1543, 45–55. doi:10.1007/978-1-4939-6716-2_3.

López-Ochoa, L., Acevedo-Hernández, G., Martínez-Hernández, A., Argüello-Astorga, G., and Herrera-Estrella, L. (2007). Structural relationships between diverse cis-acting elements are critical for the functional properties of a rbcS minimal light regulatory unit. *J. Exp. Bot.* 58, 4397–4406. doi:10.1093/jxb/erm307.

Lorenzo, O., Chico, J. M., Sánchez-Serrano, J. J., and Solano, R. (2004). JASMONATE-INSENSITIVE1 Encodes a MYC Transcription Factor Essential to Discriminate between Different Jasmonate-Regulated Defense Responses in Arabidopsis. *Plant Cell* 16, 1938–50. doi:10.1105/tpc.022319.

Lorenzo, O., Piqueras, R., Sánchez-Serrano, J. J., and Solano, R. (2003). ETHYLENE RESPONSE FACTOR1 Integrates Signals from Ethylene and Jasmonate Pathways in Plant Defense. *Plant Cell* 15, 165–78. doi:10.1105/tpc.007468.

Loss-Morais, G., Turchetto-Zolet, A. C., Etges, M., Cagliari, A., Körbes, A. P., Maraschin, F. dos S., et al. (2013). Analysis of castor bean ribosome-inactivating proteins and their gene expression during seed development. *Genet. Mol. Biol.* 36, 74–86. doi:10.1590/S1415-47572013005000005.

Luo, H., Song, F., Goodman, R. M., and Zheng, Z. (2005). Up-regulation of OsBIHD1, a rice gene encoding BELL homeodomain transcriptional factor, in disease resistance responses. *Plant Biol.* 7, 459–68. doi:10.1055/s-2005-865851.

Manners, J. M., Penninckx, I. A. M. A., Vermaere, K., Kazan, K., Brown, R. L., Morgan, A., et al. (1998). The promoter of the plant defensin gene PDF1.2 from Arabidopsis is systemically activated by fungal pathogens and responds to methyl jasmonate but not to salicylic acid. *Plant Mol. Biol.* 38, 1071–80. doi:10.1023/A:1006070413843.

Mansouri, S., Choudhary, G., Sarzala, P. M., Ratner, L., and Hudak, K. A. (2009). Suppression of human T-cell leukemia virus I gene expression by pokeweed antiviral protein. *J. Biol. Chem.* 284, 31453–31462. doi:10.1074/jbc.M109.046235.

Marcotte, W. R., Russell, S. H., and Quatrano, R. S. (1989). Abscisic Acid-Responsive Sequences from the Em Gene of Wheat. *Plant Cell* 1, 969–76. doi:10.2307/3868997.

Maruyama-Nakashita, A., Nakamura, Y., Watanabe-Takahashi, A., Inoue, E., Yamaya, T., and Takahashi, H. (2005). Identification of a novel cis-acting element conferring sulfur deficiency response in Arabidopsis roots. *Plant J.* 42, 305–14. doi:10.1111/j.1365-313X.2005.02363.x.

Mena, M., Cejudo, F., Isabel-Lamoneda, I., and Carbonero, P. (2002). A Role for the DOF Transcription Factor BPBF in the Regulation of Gibberellin-Responsive Genes in Barley Aleurone. *Plant Physiol.* 130, 111–9. doi:10.1104/pp.005561.

Minegishi, N., Ohta, J., Suwabe, N., Nakauchi, H., Ishihara, H., Hayashi, N., et al. (1998). Alternative promoters regulate transcription of the mouse GATA-2 gene. *J. Biol. Chem.* 273,

3625–34. doi:10.1074/jbc.273.6.3625.

Mohanty, B., Krishnan, S. P. T., Swarup, S., and Bajic, V. B. (2005). Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. *Ann. Bot.* 96, 669–81. doi:10.1093/aob/mci219.

Molina, C., and Grotewold, E. (2005). Genome wide analysis of Arabidopsis core promoters. *BMC Genomics* 6, 25. doi:10.1186/1471-2164-6-25.

Morello, L., Bardini, M., Sala, F., and Breviario, D. (2002). A long leader intron of the *Ostub16* rice β-tubulin gene is required for high-level gene expression and can autonomously promote transcription both *in vivo* and *in vitro*. *Plant J.* 29, 33–44. doi:10.1046/j.0960-7412.2001.01192.x.

Neller, K. C. M., Diaz, C. A., Platts, A. E., and Hudak, K. A. (2019). *De novo* Assembly of the Pokeweed Genome Provides Insight Into Pokeweed Antiviral Protein (PAP) Gene Expression. *Front. Plant. Sci.* 10:1002. doi: 10.3389/fpls.2019.01002.

Neller, K. C. M., Klenov, A., and Hudak, K. A. (2016). The Pokeweed Leaf mRNA Transcriptome and Its Regulation by Jasmonic Acid. *Front. Plant Sci.* 7:283. doi:10.3389/fpls.2016.00283.

Ng, D. W. K., Abeysinghe, J. K., and Kamali, M. (2018). Regulating the regulators: The control of transcription factors in plant defense signaling. *Int. J. Mol. Sci.* 19, 3737. doi:10.3390/ijms19123737.

Niedz, R. P., Sussman, M. R., and Satterlee, J. S. (1995). Green fluorescent protein: an in vivo reporter of plant gene expression. *Plant Cell Rep.* 14, 403–406. doi:10.1007/BF00234043.

Nishiuchi, T., Shinshi, H., and Suzuki, K. (2004). Rapid and transient activation of transcription of the ERF3 gene by wounding in tobacco leaves: Possible involvement of NtWRKYs and autorepression. *J. Biol. Chem.* 279, 55355–61. doi:10.1074/jbc.M409674200.

Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y., and Yamaguchi, S. (2003). Gibberellin Biosynthesis and Response during Arabidopsis Seed Germination. *Plant Cell* 15, 1591–604. doi:10.1105/tpc.011650.

Ohta, S., Mita, S., Hattori, T., and Nakamura, K. (1990). Construction and Expression in Tobacco of a β-Glucuronidase (GUS) Reporter Gene Containing an Intron Within the Coding Sequence. *Plant Cell Physiol.* 31, 805–13. doi:10.1093/oxfordjournals.pcp.a077982.

Ow, D. W., Wood, K. V., Deluca, M., De Wet, J. R., Helinski, D. R., and Howell, S. H. (1986). Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science* 234, 856–859. doi:10.1126/science.234.4778.856.

Păcurar, D. I., Thordal-Christensen, H., Păcurar, M. L., Pamfil, D., Botez, M. L., and Bellini, C. (2011). Agrobacterium tumefaciens: From crown gall tumors to genetic transformation. *Physiol. Mol. Plant Pathol.* 76, 76–81. doi:10.1016/j.pmpp.2011.06.004.

Pal, S., Gupta, R., and Davuluri, R. V. (2014). Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq. *Methods Mol. Biol.* 39, 190–201. doi:10.1007/978-1-4939-992-6_1.

Palmer, A. C., Barry Egan, J., and Shearwin, K. E. (2011). Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. *Transcription* 2, 9–14. doi:10.4161/trns.2.1.13511.

Pandey, P., Irulappan, V., Bagavathiannan, M. V., and Senthil-Kumar, M. (2017). Impact of Combined Abiotic and Biotic Stresses on Plant Growth and Avenues for Crop Improvement by Exploiting Physio-morphological Traits. *Front. Plant Sci.* 8:537. doi:10.3389/fpls.2017.00537.

Park, H., Kim, M., Kang, Y., Jeon, J., Yoo, J., Kim, M., et al. (2004). Pathogen- and NaCl-Induced Expression of the SCaM-4 Promoter Is Mediated in Part by a GT-1 Box That Interacts with a GT-1-Like Transcription Factor. *Plant Physiol.* 135, 2150–61. doi:10.1104/pp.104.041442.

Park, S., CB, L., JC, L., and JM, V. (2002). Isolation and characterization of a novel ribosome-inactivating protein from root cultures of pokeweed and its mechanism of secretion from roots. *Plant Physiol.* 130, 164–78. doi:10.1104/pp.000794.

Parra, G., Bradnam, K., Rose, A. B., and Korf, I. (2011). Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nucleic Acids Res.* 39, 5328–5337. doi:10.1093/nar/gkr043.

Peccarelli, M., and Kebaara, B. W. (2014). Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. *Eukaryot. Cell* 13, 1126–35. doi:10.1128/EC.00090-14.

Peng, K., Luo, C., You, W., Lian, C., Li, X., and Shen, Z. (2008). Manganese uptake and interactions with cadmium in the hyperaccumulator-Phytolacca Americana L. *J. Hazard. Mater.* 154, 674–681. doi:10.1016/j.jhazmat.2007.10.080.

Pieterse, C. M. J., Van der Does, D., Zamioudis, C., Leon-Reyes, A., and Van Wees, S. C. M. (2011). Hormonal Modulation of Plant Immunity. *Annu. Rev. Cell Dev. Biol.* 28, 489–521. doi:10.1146/annurev-cellbio-092910-154055.

Poyet, J.-L., and Hoeveler, A. (1997). Presence of an Intron in a Gene of PAP II, the Ribosome-inactivating Protein from Summer Leaves of Phytolacca americana. *Ann. Bot.* 80, 685–688. doi:10.1006/anbo.1997.0478.

Qi, X. T., Zhang, Y. X., and Chai, T. Y. (2007). The bean PvSR2 gene produces two transcripts by alternative promoter usage. *Biochem. Biophys. Res. Commun.* 356, 273–278. doi:10.1016/j.bbrc.2007.02.124.

Qin, X., Shao, C., Hou, P., Gao, J., Lei, N., Jiang, L., et al. (2010). Different functions and expression profiles of curcin and curcin-L in Jatropha curcas L. *Z. Naturforsch. C.* 65, 355–

62.

Qin, X., Zheng, X., Shao, C., Gao, J., Jiang, L., Zhu, X., et al. (2009). Stress-induced curcin-L promoter in leaves of Jatropha curcas L. and characterization in transgenic tobacco. *Planta* 230, 387–395. doi:10.1007/s00425-009-0956-9.

Qin, Z., Stoilov, P., Zhang, X., and Xing, Y. (2018). SEASTAR: systematic evaluation of alternative transcription start sites in RNA. *Nucleic Acids Res.* 46, e45. doi:10.1093/nar/gky053.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. URL: http://www.R-project.org/.

Rajamohan, F., Venkatachalam, T. K., Irvin, J. D., and Uckun, F. M. (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, 453–58. doi:10.1006/bbrc.1999.0922.

Ready, M. P., Brown, D. T., and Robertus, J. D. (1986). Extracellular localization of pokeweed antiviral protein. *Proc. Natl. Acad. Sci.* 83, 5053–5056. doi:10.1073/pnas.83.14.5053.

Reinbothe, S., Reinbothe, C., Lehmann, J., Becker, W., Apel, K., and Parthier, B. (2006). JIP60, a methyl jasmonate-induced ribosome-inactivating protein involved in plant stress reactions. *Proc. Natl. Acad. Sci.* 91, 7012–7016. doi:10.1073/pnas.91.15.7012.

Reynolds, N., O'Shaughnessy, A., and Hendrich, B. (2013). Transcriptional repressors: multifaceted regulators of gene expression. *Development* 140, 505–512. doi:10.1242/dev.083105.

Rieping, M., and Schöffl, F. (1992). Synergistic effect of upstream sequences, CCAAT box elements, and HSE sequences for enhanced expression of chimaeric heat shock genes in transgenic tobacco. *Mol. Gen. Genet.* 231, 226–32. doi:10.1007/BF00279795.

Rochon, A., Boyle, P., Wignes, T., Fobert, P. R., and Després, C. (2006). The Coactivator Function of Arabidopsis NPR1 Requires the Core of Its BTB/POZ Domain and the Oxidation of C-Terminal Cysteines. *Plant Cell* 18, 3670–85. doi:10.1105/tpc.106.046953.

Rodriguez, L., Gonzalez-Guzman, M., Diaz, M., Rodrigues, A., Izquierdo-Garcia, A. C., Peirats-Llobet, M., et al. (2014). C2-Domain Abscisic Acid-Related Proteins Mediate the Interaction of PYR/PYL/RCAR Abscisic Acid Receptors with the Plasma Membrane and Regulate Abscisic Acid Sensitivity in Arabidopsis. *Plant Cell* 26, 4802–20. doi:10.1105/tpc.114.129973.

Rose, A. B., Emami, S., Bradnam, K., and Korf, I. (2011). Evidence for a DNA-Based Mechanism of Intron-Mediated Enhancement. *Front. Plant Sci.* 2:98. doi:10.3389/fpls.2011.00098.

Rouster, J., Leah, R., Mundy, J., and Cameron-Mills, V. (1997). Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley

grain. *Plant J.* 11, 513–523. doi:10.1046/j.1365-313X.1997.11030513.x.

Santner, A., Calderon-Villalobos, L. I. A., and Estelle, M. (2009). Plant hormones are versatile chemical regulators of plant growth. *Nat. Chem. Biol.* 5, 301–307. doi:10.1038/nchembio.165.

Satoh, R., Fujita, Y., Nakashima, K., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004). A novel subgroup of bZIP proteins functions as transcriptional activators in hypoosmolarity-responsive expression of the ProDH gene in Arabidopsis. *Plant Cell Physiol.* 45, 309–17. doi:10.1093/pcp/pch036.

Schambach, A., Mueller, D., Galla, M., Verstegen, M. M. A., Wagemaker, G., Loew, R., et al. (2006). Overcoming promoter competition in packaging cells improves production of self-inactivating retroviral vectors. *Gene Ther.* 13, 1524–53. doi:10.1038/sj.gt.3302807.

Schluttenhofer, C., Pattanaik, S., Patra, B., and Yuan, L. (2014). Analyses of Catharanthus roseus and Arabidopsis thaliana WRKY transcription factors reveal involvement in jasmonate signaling. *BMC Genomics* 15, 502. doi:10.1186/1471-2164-15-502.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326, 1112–15. doi:10.1126/science.1178534.

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. *J. Virol.* 73, 7823–7829.

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

Shaul, O. (2017). How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* 91, 145–155. doi:10.1016/j.biocel.2017.06.016.

Simpson, S. D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. *Plant J.* 33, 259–70. doi:10.1046/j.1365-313X.2003.01624.x.

Singer, G. A. C., Wu, J., Yan, P., Plass, C., Huang, T. H. M., and Davuluri, R. V. (2008). Genome-wide analysis of alternative promoters of human genes using a custom promoter tiling array. *BMC Genomics* 9, 349. doi:10.1186/1471-2164-9-349.

Söderman, E., Mattsson, J., and Engström, P. (1996). The Arabidopsis homeobox gene ATHB-7 is induced by water deficit and by abscisic acid. *Plant J.* 10, 375–381. doi:10.1046/j.1365-313X.1996.10020375.x.

Song, Q. A., Catlin, N. S., Brad Barbazuk, W., and Li, S. (2019). Computational analysis of

alternative splicing in plant genomes. *Gene* 685, 186–195. doi:10.1016/j.gene.2018.10.026.

Song, S. K., Choi, Y., Moon, Y. H., Kim, S. G., Choi, Y. D., and Lee, J. S. (2000). Systemic induction of a Phytolacca insularis antiviral protein gene by mechanical wounding, jasmonic acid, and abscisic acid. *Plant Mol. Biol.* 43, 439–50.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* 19, 262–274. doi:10.1038/nrm.2017.115.

Spoel, S. H., Koornneef, A., Claessens, S. M. C., Korzelius, J. P., Van Pelt, J. A., Mueller, M. J., et al. (2003). NPR1 Modulates Cross-Talk between Salicylate- and Jasmonate-Dependent Defense Pathways through a Novel Function in the Cytosol. *Plant Cell* 15, 760–770. doi:10.1105/tpc.009159.

Srivastava, A. K., Lu, Y., Zinta, G., Lang, Z., and Zhu, J. K. (2018). UTR-Dependent Control of Gene Expression in Plants. *Trends Plant Sci.* 23, 248–259. doi:10.1016/j.tplants.2017.11.003.

Sutoh, K., and Yamauchi, D. (2003). Two cis-acting elements necessary and sufficient for gibberellin-upregulated proteinase expression in rice seeds. *Plant J.* 34, 635–45. doi:10.1046/j.1365-313X.2003.01753.x.

Suttipanta, N., Pattanaik, S., Kulshrestha, M., Patra, B., Singh, S. K., and Yuan, L. (2011). The Transcription Factor CrWRKY1 Positively Regulates the Terpenoid Indole Alkaloid Biosynthesis in Catharanthus roseus. *Plant Physiol.* 157, 2081–93. doi:10.1104/pp.111.181834.

Tada, Y., Spoel, S. H., Pajerowska-Mukhtar, K., Mou, Z., Song, J., Wang, C., et al. (2008). Plant immunity requires conformational charges of NPR1 via S-nitrosylation and thioredoxins. *Science* 321, 952–56. doi:10.1126/science.1156970.

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., et al. (2007). JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. *Nature* 448, 661–665. doi:10.1038/nature05960.

Umezawa, T., Sugiyama, N., Mizoguchi, M., Hayashi, S., Myouga, F., Yamaguchi-Shinozaki, K., et al. (2009). Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. *Proc. Natl. Acad. Sci.* 106, 17588–93. doi:10.1073/pnas.0907095106.

Urasaki, N., Takagi, H., Natsume, S., Uemura, A., Taniai, N., Miyagi, N., et al. (2017). Draft genome sequence of bitter gourd (Momordica charantia), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Res.* 24, 51–58. doi:10.1093/dnares/dsw047.

Verma, V., Ravindran, P., and Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* doi:10.1186/s12870-016-0771-y.

Vernimmen, D., and Bickmore, W. A. (2015). The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends Genet.* 31, 696–708. doi:10.1016/j.tig.2015.10.004.

Vernimmen, D., Gobbi, M. De, Sloane-Stanley, J. A., Wood, W. G., and Higgs, D. R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* 26, 2041–51. doi:10.1038/sj.emboj.7601654.

Vierstra, J., Wang, H., John, S., Sandstrom, R., and Stamatoyannopoulos, J. A. (2014). Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nat. Methods* 11, 66–72. doi:10.1038/nmeth.2713.

Vockley, C. M., Ippolito, A. M. D., Mcdowell, I. C., Song, L., Crawford, G. E., Reddy, T. E., et al. (2016). Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. *Cell* 166, 1269–1281.e19. doi:10.1016/j.cell.2016.07.049.

Wang, P., Zoubenko, O., and Tumer, N. E. (1998). Reduced toxicity and broad spectrum resistance to viral and fungal infection in transgenic plants expressing pokeweed antiviral protein II. *Plant Mol. Biol.* 38, 957–64. doi:10.1023/A:1006084925016.

Wang, X. Q., and Rothnagel, J. A. (2004). 5′-Untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. *Nucleic Acids Res.* 32, 1382–91. doi:10.1093/nar/gkh305.

Wang, Z., Zang, C., Cui, K., Schones, D. E., Barski, A., Peng, W., et al. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* 138, 1019–1031. doi:10.1016/j.cell.2009.06.049.

Wasternack, C., and Hause, B. (2013). Jasmonates: Biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Ann. Bot.* 111, 1021–58. doi:10.1093/aob/mct067.

Weiss, D., and Ori, N. (2007). Mechanisms of Cross Talk between Gibberellin and Other Hormones. *Plant Physiol.* 144, 1240–46. doi:10.1104/pp.107.100370.

Wickham, H. (2011). ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3, 180–185. doi:10.1002/wics.147.

Wirén, M., Silverstein, R. A., Sinha, I., Walfridsson, J., Lee, H. M., Laurenson, P., et al. (2005). Genome-wide analysis of nucleosome density histone acetylation and HDAC function in fission yeast. *EMBO J.* 24, 2906–2918. doi:10.1038/sj.emboj.7600758.

Wu, J. Y., Xiao, J. F., Wang, L. P., Zhong, J., Yin, H. Y., Wu, S. X., et al. (2013). Systematic analysis of intron size and abundance parameters in diverse lineages. *Sci. China Life Sci.* 56, 968–74. doi:10.1007/s11427-013-4540-y.

Wu, P., Zhou, C., Cheng, S., Wu, Z., Lu, W., Han, J., et al. (2015). Integrated genome sequence and linkage map of physic nut (Jatropha curcas L.), a biodiesel plant. *Plant J.* 81, 810–821. doi:10.1111/tpj.12761.

Xu, L., Liu, F., Lechner, E., Genschik, P., Crosby, W. L., Ma, H., et al. (2002). The SCF COI1 Ubiquitin-Ligase Complexes Are Required for Jasmonate Response in Arabidopsis. *Plant Cell* 14, 1919–35. doi:10.1105/tpc.003368.

Yamamoto, S., Nakano, T., Suzuki, K., and Shinshi, H. (2004). Elicitor-induced activation of transcription via W box-related cis-acting elements from a basic chitinase gene by WRKY transcription factors in tobacco. *Biochim. Biophys. Acta* 1679, 279–87. doi:10.1016/j.bbaexp.2004.07.005.

Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., and Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: Structure, function, and evolution. *Trends Plant Sci.* 18, 267–276. doi:10.1016/j.tplants.2012.09.001.

Yasuda, M., Ishikawa, A., Jikumaru, Y., Seki, M., Umezawa, T., Asami, T., et al. (2008). Antagonistic Interaction between Systemic Acquired Resistance and the Abscisic Acid–Mediated Abiotic Stress Response in Arabidopsis. *Plant Cell* 20, 1678–92. doi:10.1105/tpc.107.054296.

Yu, D., Chen, C., Chen, Z., Jiang, C.-J., Ono, K., Toki, S., et al. (2001). Evidence for an Important Role of WRKY DNA Binding Proteins in the Regulation of NPR1 Gene Expression. *Plant Cell* 13, 1527–1540. doi:10.1105/TPC.010115.

Yuan, Y., Zhong, S., Li, Q., Zhu, Z., Lou, Y., Wang, L., et al. (2007). Functional analysis of rice NPR1-like genes reveals that OsNPR1/NH1 is the rice orthologue conferring disease resistance with enhanced herbivore susceptibility. *Plant Biotechnol. J.* 5, 313–24. doi:10.1111/j.1467-7652.2007.00243.x.

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., et al. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518, 556–559. doi:10.1038/nature13994.

Zander, M., La Camera, S., Lamotte, O., Métraux, J. P., and Gatz, C. (2010). Arabidopsis thaliana class-II TGA transcription factors are essential activators of jasmonic acid/ethylene-induced defense responses. *Plant J.* 61, 200–210. doi:10.1111/j.1365-313X.2009.04044.x.

Zhang, Y., Fan, W., Kinkema, M., Li, X., and Dong, X. (2002). Interaction of NPR1 with basic leucine zipper protein transcription factors that bind sequences required for salicylic acid induction of the PR-1 gene. *Proc. Natl. Acad. Sci.* 1999, 6523–28. doi:10.1073/pnas.96.11.6523.

Zhu, F., Zhou, Y.-K., Ji, Z.-L., and Chen, X.-R. (2018). The Plant Ribosome-Inactivating Proteins Play Important Roles in Defense against Pathogens and Insect Pest Attacks. *Front. Plant Sci.* 9:146. doi:10.3389/fpls.2018.00146.

Zhu, J., Oger, P. M., Schrammeijer, B., Hooykaas, P. J. J., Farrand, S. K., and Winans, S. C. (2000). The bases of crown gall tumorigenesis. *J. Bacteriol.*, 3885–95. doi:10.1128/JB.182.14.3885-

3895.2000.

Zhu, Z., An, F., Feng, Y., Li, P., Xue, L., A, M., et al. (2011). Derepression of ethylene-stabilized transcription factors (EIN3/EIL1) mediates jasmonate and ethylene signaling synergy in Arabidopsis. *Proc. Natl. Acad. Sci.* 108, 12539–44. doi:10.1073/pnas.1103959108.

Zoubenko, O., Uckun, F., Hur, Y., Chet, I., and Tumer, N. (1997). Plant resistance to fungal infection induced by nontoxic pokeweed antiviral protein mutants. *Nat. Biotechnol.* 15, 992–996. doi:10.1038/nbt1097-992.

APPENDIX A: CONTRIBUTION TO PUBLISHED MANUSCRIPT

De novo assembly of the pokeweed genome provides insight into pokeweed antiviral protein (PAP) gene expression

Kira C. M. Neller, **Camille A. Diaz**, Adrian E. Platts, and Katalin A. Hudak

Frontiers in Plant Science (2019), 10:1002. doi: 10.3389/fpls.2019.01002

Contributions:

KCMN and KAH designed the study. KCMN performed bioinformatic analyses including genome assembly, annotation, and RNA-Seq analysis, and drafted the manuscript. KAH and KCMN edited the manuscript.

CAD performed all cloning, gene model and RNA-seq validations (**Figure 3B; Supplementary Figure 4**), measurement of gene expression from reporter constructs (**Figure 3C; Figure 6**), and identification of promoter elements (**Table 5**). CAD drafted methods and results sections corresponding to these figures.

AEP provided advice on bioinformatic analyses and data interpretation.

APPENDIX B: LIST OF PRIMERS

Table 2. List of primer sequences used for PCR amplification and cloning of PAP promoter fragments. SLIC/Gibson assembly sequences are in lowercase.

Primer names	Primer sequences (5'—3')	Purpose
pLK-1-FOR	ATTGGAAGTGGATAAGCATGCAAGCTTGGC	pHSG299 vector PCR
pLK-1-REV	ATTGGATTGGAAGTACTATAGTGAGTCGTATTAGAATTCACTG	
PAP-I-prom-SLIC-FOR	ctcactatagtacttccaatccaatGACATCACTCACAGGATCAGGATTGGCTG	Cloning of the promoter and intron into pHSG299
PAP-I-prom-SLIC-REV	gcttgcatgcttatccacttccaatCTTCCCTTCCTACTAGCTAGTTCCAATACTTTCGC	
pCambia-prom-FOR	ATGGCTACTACTAAGCATTTGGCTCTTGCCATCCTTG	pCambia 0305.2 GFP-GUS vector PCR
pCambia-prom-REV	GATATTTTTGGAGTAGACAAGTGTGTCGTGCTCCAC	
P1-FOR	cacacttgtctactccaaaaatatcGACATCACTCACAGGATCAGGATTGGCTG	PAP-I promoter (1262) forward primer. Paired with 5-UTR-SLIC-REV
P2-FOR	cacacttgtctactccaaaaatatcGAGTGTTAGCAAATCAAGAAGGAGCAAGCAC	PAP-I promoter truncation (1124) forward primer. Paired with 5-UTR-SLIC-REV
P3-FOR	cacacttgtctactccaaaaatatcAATGGGTAGTCGCCCCTTCTAGGACATTATC	PAP-I promoter truncation (711) forward primer. Paired with 5-UTR-SLIC-REV
P4-FOR	cacacttgtctactccaaaaatatcACTAGACTTCAGTAGTCTGTAGTCAACCCTATTTGG	PAP-I promoter truncation (584) forward primer. Paired with 5-UTR-SLIC-REV
P5-FOR	cacacttgtctactccaaaaatatcCGGAGGTCTCTCTTGTATACAGCCTCTCTATCC	PAP-I promoter truncation (432) forward primer. Paired with 5-UTR-SLIC-REV
P6-FOR	cacacttgtctactccaaaaatatcCCCTACATGATTTACTCGAAAAATCATACGTTCTAATTGGTTGG	PAP-I promoter truncation (296) forward primer. Paired with 5-UTR-SLIC-REV
P7-FOR	cacacttgtctactccaaaaatatcGTTACGCATGCTAGGCGCCACAATTTTACAAAC	PAP-I promoter truncation (102) forward primer. Paired with 5-UTR-SLIC-REV
Int-FOR	cacacttgtctactccaaaaatatcGCAACCACATGAGATCCTGACGCAAAACC	PAP-I intron forward primer. Paired with 5-UTR-SLIC-REV to amplify 1.4 Kb of the intron
5-UTR-no-int-REV	CTTCCCTTCCTACTAGCTAGTTCCAATACTTTCGCCCTGTAGTTAACTCACAACTTTCTTTTTTTCTTACAACG	PAP-I 5'UTR reverse primer (intronless)
5-UTR-SLIC-REV	gagccaaatgcttagtagtagccatCTTCCCTTCCTACTAGCTAGTTCCAATACTTTCGC	PAP-I 5'UTR reverse primer

Table 2, continued.

Primer names	Primer sequences (5'—3')	Purpose
Int-transl-FOR	GTTGTAAGTGCTCTTCCTTTCATATTACTCTTTTTTGTATGATTC	PAP-I intron forward and reverse primers. Amplifies the full 1.6 Kb intron
Int-transl-REV	CTATCAATGAAAAACAAGGTAAGAACCTGGATTACATGTACGATG	
Int-transl-Gib-FOR	cacacttgtctactccaaaaatatcGTTGTAAGTGCTCTTCCTTTCATATTACTC	PAP-I intron forward primer with Gibson assembly sequence. Used to translocate the 1.6 Kb intron upstream of the PAP-I promoter (1262 and 102)
Int-transl-Gib-1-REV	catctatcaacactatgtcatccgcCTATCAATGAAAAACAAGGTAAGAACCTGG	PAP-I intron reverse primer with Gibson assembly sequence. Used to translocate the 1.6 Kb intron upstream of the PAP-I promoter (1262)
Int-transl-Gib-2-REV	aattgtggcgcctagcatgcgtaacCTATCAATGAAAAACAAGGTAAGAACCTGG	PAP-I intron reverse primer with Gibson assembly sequence. Used to translocate the 1.6 Kb intron upstream of the PAP-I promoter (102)
P1-transl-FOR	GACATCACTCACAGGATCAGGATTGGCTG	Paired with pCambia-prom-REV to generate a linear vector. Used to make the translocated intron-1262::GFP-GUS construct
P7-transl-FOR	GTTACGCATGCTAGGCGCCACAATTTTACAAAC	Paired with pCambia-prom-REV to generate a linear vector. Used to make the translocated intron-102::GFP-GUS construct
pCambia-1-no-cat-FOR	CCTCAGATCTCCAGAGCCACCGCCA	Forward and reverse primers for the removal of the catalase intron in GUS
pCambia-1-no-cat-REV	CAACCAGGCACCGACGCCGTGGAAT	
pCambia-2-no-cat-FOR	tggcggtggctctggagatctgaggAACCGACGAACTAGTCTGTACCCGA	
pCambia-2-no-cat-REV	attccacggcgtcggtgcctggttgTTCTTGATTTTCCATGCCGCCTCCTTTAG	
pCambia-GBOX-mut-1-FOR	aaccaaccaagaacatgaaaCCCTTGAAAGAAACAAAAAATGTAATGAAG	Forward and reverse primers for the deletion of the JA-associated T/GBOXATPIN2 element in the PAP-I promoter
pCambia-GBOX-mut-1-REV	taaacgctcttttctcttagGTTTACCCGCCAATATATCCTGTCAAAC	
pCambia-GBOX-mut-2-FOR	CTAAGAGAAAAGAGCGTTTATTAGAATAACGG	
pCambia-GBOX-mut-2-REV	TTTCATGTTCTTGGTTGGTTATTTTACTATACATATAC	
nPAP-prom-FOR	GTTTTCCCTTGAAGTCACTTGGTTTCCTTGAC	Novel PAP promoter forward primer
nPAP-prom-REV	AAACCTGGTATAATAGCCGAAGCTTCGTGG	Novel PAP promoter reverse primer

Table 2, continued.

Primer names	Primer sequences (5'—3')	Purpose
nPAP-UTR-REV	ATTTCACTAGTTCTAGATTTGTTTCACCTGAAACCTGGTATAATAGCCGAAGCTTCGTGG	Novel PAP 5'UTR (intronless)
nPAP-prom-Gib-FOR	cacacttgtctactccaaaaatatcGTTTTCCCTTGAAGTCACTTGGTTTCCTTGAC	Novel PAP promoter forward primer with Gibson sequence
nPAP-UTR-Gib-REV	gagccaaatgcttagtagtagccatATTTCACTAGTTCTAGATTTGTTTCACCTGAAACC	Novel PAP 5'UTR primer with Gibson sequence
PAP-II-A-prom-FOR	CACTATGGCTTGGAAGGCACTTCAAGAG	PAP-II_A promoter forward primer
PAP-II-A-prom-REV	TCCCTGTTGCGCAGCTGAAGCTTTTGTAATTG	PAP-II_A promoter reverse primer
PAP-II-A-prom-Gib-FOR	cacacttgtctactccaaaaatatcCACTATGGCTTGGAAGGCACTTCAAGAG	PAP-II_A promoter forward primer with Gibson sequence
PAP-II-A-prom-Gib-REV	gagccaaatgcttagtagtagccatTCCCTGTTGCGCAGCTGAAGCTTTTGTAATTG	PAP-II_A promoter reverse primer with Gibson sequence
PAP-II-B-prom-FOR	CGGGACGTCTCCGGAATTAGTCAATTGC	PAP-II_B promoter forward primer
PAP-II-B-prom-REV	GCAACCTGTTACGTAGCAGATGAGGC	PAP-II_B promoter reverse primer
PAP-II-B-prom-Gib-FOR	cacacttgtctactccaaaaatatcCGGGACGTCTCCGGAATTAGTCAATTGC	PAP-II_B promoter forward primer with Gibson sequence
PAP-II-B-prom-Gib-REV	gagccaaatgcttagtagtagccatGCAACCTGTTACGTAGCAGATGAGGC	PAP-II_B promoter reverse primer with Gibson sequence
PAP-alpha-prom-FOR	GGTAATGGGTCATGCTTGTTATAGTGTTGTAGAGTAC	PAP-α promoter forward primer
PAP-alpha-prom-REV	CATATTTGCAACCTGTTACGTAGCAGATGAGGC	PAP-α promoter reverse primer
PAP-alpha-UTR-REV	CTTCCCTTGTTGCTGCTTTCAACACTTTCACCTGTTACGTAGCAGATGAGGCTTTTTATGAG	PAP-α 5'UTR reverse primer (intronless)
PAP-alpha-prom-Gib-FOR	cacacttgtctactccaaaaatatcGGTAATGGGTCATGCTTGTTATAGTGTTGTAGAGTAC	PAP-α promoter forward primer with Gibson sequence
PAP-alpha-UTR-Gib-REV	gagccaaatgcttagtagtagccatCTTCCCTTGTTGCTGCTTTCAACACTTTC	PAP-α 5'UTR reverse primer with Gibson sequence
PAP-S2-A-prom-FOR	GAACAGCAGAGCATCTAGACTGAACCC	PAP-$S2_A$ promoter forward primer
PAP-S2-A-prom-REV	GCAACCTGTTGTACGTAGCAGCTGATGC	PAP-$S2_A$ promoter reverse primer
PAP-S2-A-UTR-REV	CTTCCCTTTGCCGCTTCTTTCAATACTTTCACCCTGTTGTACGTAGCAGCTGATGC	PAP-$S2_A$ 5'UTR reverse primer (intronless)
PAP-S2-A-prom-Gib-FOR	cacacttgtctactccaaaaatatcGCAACCTGTTGTACGTAGCAGCTGATGC	PAP-$S2_A$ promoter forward primer with Gibson sequence
PAP-S2-A-UTR-Gib-REV	gagccaaatgcttagtagtagccatCTTCCCTTTGCCGCTTCTTTCAATAC	PAP-$S2_A$ 5'UTR reverse primer with Gibson sequence

Table 3. List of primers used for Sanger sequencing of promoter-reporter gene constructs.

Primer names	Primer sequences (5'—3')	Purpose
pCambia-seq-1-FOR	GAGCACGACACACTTGTCTACTCCAAAAATATC	Sequencing of PAP-I promoter and intron
pCambia-seq-2-FOR	GGATAGGAGTCTGTGGAGAGTTAGAATTAGG	
pCambia-seq-3-FOR	ATGGTGTGCTGTGCCTGTATAAAAGGG	
pCambia-seq-4-FOR	GGTCCTCCTGATGTAAACACTAGCCATG	
pCambia-seq-5-FOR	CAACAAGGTTGTGGTTATAAGAGTATTCAGTTG	
GRP-REV	CTAAGGAGGACAAGGATGGCAAGAGCC	
pCambia-seq-6-FOR	CGACGGCAACTACAAGACCCG	Sequencing of GFP-GUS reporter gene
pCambia-seq-7-FOR	TGGAAGAGAAGTGGTACGAAAGCAAGC	
pCambia-seq-8-FOR	GTCTATGAAGAGCCGTTCGGCGTG	
pCambia-seq-9-FOR	GCTTTCACGACATTGATCCAGTGATGTTC	

Table 4. List of primer sequences used for qRT-PCR of GFP-GUS transcripts in agroinfiltrated tobacco leaves.

Primer names	Primer sequences (5'—3')	qPCR final concentration (µM)	Purpose
GUS-RT-REV	GAGGATATTGAAATCCATCACATTGCTCGC	—	GFP-GUS RT
GUS-2-FOR	CACATGGTCCTGCTGGAGTTCG	0.6	GFP-GUS qPCR (GUS CDS)
GUS-2-REV	CAGACGCCATTGAGGTCGAAGACG	0.6	
PAP-I-1-35-FOR	ATAACTGCATGTTCTCATAAAAAAGCCTCAGC	0.3	GFP-GUS qPCR (5'UTR)
PAP-I-112-83-REV	CTTCCTACTAGCTAGTTCCAATACTTTCGC	0.3	
L25-FOR	CCGAGTCTGCAATGAAGAAGATTGAGG	0.3	L25 qPCR (reference gene)
L25-REV	CTGAATGTCATACATCTTCTTCACAGCATCC	0.3	L25 qPCR; L25 RT (reference gene)
Ntubc-FOR	CTCGCCAGCTAATTCAGAAGCAGC	0.6	Ubc qPCR (reference gene)
Ntubc-REV	GAAGTCAGTCTGCTGTCCAGCTC	0.6	Ubc qPCR; Ubc RT (reference gene)

Table 5. List of primer sequences used for qRT-PCR of PAP transcripts in pokeweed leaves.

Primer names	Primer sequences (5'—3')	qPCR final concentration (μM)	Purpose
PAP-I-full REV	GATATGATTTGAATCACTCGAATTCACC AAGG	—	PAP-I RT
nPAP full REV	GCTTTGAGGAGCATGTGATTATAGAATG GTG	—	Novel PAP RT
PAP-II-720-690 REV	CCTTGATGCCTCATTAACCATTTGAACG	—	PAP-II RT
PAP-alpha-893-864 REV	CATCCACTCTCAGCACTATCCACTTGCT ACCGTTTG	—	PAP-α RT
PAP-S1-full-REV	GGCATTTTGTTAAGTTGCTTGGCAAGTC CC	—	PAP-S1 RT
BDX-RT	CTTCACTCTCAGGATCCGTTC	—	BDX RT (reference gene)
EF1G-RT	CTTCAAAAGGCTCCTGGTC	—	EF1G RT (reference gene)
PAP-I-1-35-FOR	ATAACTGCATGTTCTCATAAAAAAGCCT CAGC	0.3	PAP-I qPCR
PAP-I-112-83-REV	CTTCCTACTAGCTAGTTCCAATACTTTC GC		
nPAP-117-145-FOR	CGCCACGAAGCTTCGGCTATTATACCAG	0.3	Novel PAP qPCR
nPAP-214-185-REV	CAAACCCATATTGCACATACAAGTGACA CC		
PAP-II-535-566-FOR	GAAAAGAGTTACAAAGGGATGGAATCAA AGG	0.3	PAP-II qPCR
PAP-II-644-618-REV	CGTTGCATCCTTGCCGTAGATTTTACC		
PAP-alpha-20-47-FOR	CACATTGCATGTTCTCATAAAAAGCCTC	0.3	PAP-α qPCR
PAP-alpha-162-136-REV	CAAGTTGAAGGTGGTTTAAGAATGAGCC		
PAP-S1-8-37-FOR	CAGAGTTATCCATCACATTGCATGCATG	0.3	PAP-S1 qPCR
PAP-S1-91-63-REV	CCGCTTCTTTCAACACTTTCACCTGTTA C		
BDX-FOR	GGAGCACACTACACCTTCGCTC	0.3	BDX qPCR (reference gene)
BDX-REV	CCTGCAACCCTGATACAATGGAAG		
EF1G-FOR	ACCAACTTCCGTGAAGTAGCAATTAAAG	0.3	EF1G qPCR (reference gene)
EF1G-REV	CTTCCCAAATGCGTACTTGCGAG		

Table 6. List of primer sequences used for gene model validations of PAP genes through PCR and RT-PCR. SLIC/Gibson assembly sequences are in lowercase.

Primer names	Primer sequences (5'—3')	Purpose
PAP-I-FOR	ATAACTGCATGTTCTCATAAAAAAGCCTCAGCTG	PAP-I PCR (gene model validation)
PAP-I-SLIC-FOR	ctcactatagtacttccaatccaatATAACTGCATGTTCTCATAAAAAAGCCTCAGCTG	
PAP-I-REV	GTTTATGATCAGAATCCTTCAAATAGATCACCAAG	
PAP-I-SLIC-REV	gcttgcatgcttatccacttccaatGTTTATGATCAGAATCCTTCAAATAGATCACCAAG	
PAP-IIa-FOR	CTCCGGTTATATATATGGCTATGCACTGCAG	PAP-II_A PCR (gene model validation)
PAP-IIa-SLIC-FOR	ctcactatagtacttccaatccaatCTCCGGTTATATATATGGCTATGCACTGCAG	
PAP-IIb-FOR	GCAGTAAATCTTAACGTACGTAGAGTCCCCTACAAAG	PAP-II_B PCR (gene model validation)
PAP-IIb-SLIC-FOR	ctcactatagtacttccaatccaatGCAGTAAATCTTAACGTACGTAGAGTCCCCTACAAAG	
PAP-II-REV	GATATGATTTGAATCACTCGAATTCACCAAGG	PAP-II PCR (gene model validation)
PAP-II-SLIC REV	gcttgcatgcttatccacttccaatGATATGATTTGAATCACTCGAATTCACCAAGG	
PAP-alpha-FOR	GTTATCCATCACATTGCATGTTCTCATAAAAAGCCTC	PAP-α PCR (gene model validation)
PAP-alpha-SLIC-FOR	ctcactatagtacttccaatccaatGTTATCCATCACATTGCATGTTCTCATAAAAAGCCTC	
PAP-alpha-REV	CATCCACTCTCAGCACTATCCACTTGCTACCGTTTG	
PAP-alpha-SLIC-REV	gcttgcatgcttatccacttccaatCATCCACTCTCAGCACTATCCACTTGCTACCGTTTG	
PAP-S1-FOR	CACAGAGTTATCCATCACATTGCATGCATG	PAP-$S1_A$ PCR (gene model validation)
PAP-S1-SLIC-FOR	ctcactatagtacttccaatccaatCACAGAGTTATCCATCACATTGCATGCATG	
PAP-S1-REV	GGCATTTTGTTAAGTTGCTTGGCAAGTCCC	
PAP-S1-SLIC-REV	gcttgcatgcttatccacttccaatGGCATTTTGTTAAGTTGCTTGGCAAGTCCC	
PAP-S2a-FOR	CTGCAACGCAGAGTTCTCCATCACATCAC	PAP-$S2_A$ PCR (gene model validation)
PAP-S2a-SLIC-FOR	ctcactatagtacttccaatccaatCCAACGCAGAGTTATCCATCACATTGTATG	
PAP-S2b-FOR	CCAACGCAGAGTTATCCATCACATTGTATG	PAP-$S2_B$ PCR (gene model validation)
PAP-S2b-SLIC-FOR	ctcactatagtacttccaatccaatCCAACGCAGAGTTATCCATCACATTGTATG	
PAP-S2-REV	GTCTGACAGGTTCCATTAACGTACTTAAGGAGTGCC	PAP-S2 PCR (gene model validation)
PAP-S2-SLIC-REV	gcttgcatgcttatccacttccaatGTCTGACAGGTTCCATTAACGTACTTAAGGAGTGCC	

Table 7. List of primer sequences used to validate the existence of the PAP-I alternative transcript.

Primer names	Primer sequences (5'—3')	Purpose
Int-1-FOR	CGCATACATCGTACATGTAATCCAGGTTC	PAP-I intron forward primers
Int-2-FOR	GACTCGGTTACAAACCTAATTGGTCTAAGG	
Int-3-FOR	GTATTACCAAGGCTCGACACGACTC	
Int-4-FOR	GGCTGGTCCTCCTGATGTAAACAC	
Int-5-FOR	CTAGCCATGTTAGTTGCTACGTTCTACC	
Int-6-FOR	GGTAGCAACCACATGAGATCCTGACG	
PAP-I-1-35-FOR	ATAACTGCATGTTCTCATAAAAAAGCCTCAGC	PAP-I 5'UTR forward primer
PAP-I-full-REV	GATATGATTTGAATCACTCGAATTCACCAAGG	PAP-I reverse primer

APPENDIX C: RECIPES

Yeast Extract Peptone (YEP) (1 L)		**Lysogeny Broth (LB) (1 L)**	
5 g	NaCl	10 g	NaCl
10 g	Yeast extract	10 g	Tryptone
10 g	Peptone	5 g	Yeast extract
15 g	Agar (for plates)	15 g	Agar (for plates)
	dH_2O to 1 L		dH_2O to 1 L

Super Optimal Broth (SOB) (1 L)		**Super Broth (SB) / Witch's Broth (1 L)**	
0.5 g	NaCl	5 g	NaCl
20 g	Tryptone	35 g	Tryptone
5 g	Yeast extract	20 g	Yeast extract
20 mL	(w/v) glucose (for SOC Media)	5 mL	1 M NaOH
	dH_2O to 1 L	4 mL	100% glycerol
			dH_2O to 1 L

Isothermal Start Mix (4 mL)		**2X Gibson Master Mix** (1295 µL)	
3 mL	1 M Tris-HCl, pH 8.0	450 µL	405 µL Isothermal Start Mix
150 µL	1 M $MgCl_2$	25 µL	1 M DTT
1.5 g	PEG-8000	20 µL	25 mM dNTPs
		50 µL	NAD^+ (New England Biolabs #B9007S)
		1 µL	T5 exonuclease (New England Biolabs #M0363S)
		31.25 µL	Phusion High-Fidelity DNA Polymerase (New England Biolabs #M0530S)
		250 µL	Taq Ligase (New England Biolabs #M0208S)
		467.75 µL	dH_2O