Investigation of enhancer-blocking DNA insulators in *Arabidopsis thaliana*

by

Anh Tran

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Biology

Ottawa-Carleton Institute of Biology
University of Ottawa
Ottawa, Canada

Abstract

Currently research has focused on insulators from non-plant species such as the fruit fly, *Drosophila melanogaster*. The accumulated data suggests that many different insulator sequences exist in *D. melanogaster*, each one containing its own different primary binding protein, while sharing similar secondary binding proteins. Together, they produce chromatin loops separating enhancers and promoters into distinct domains preventing cross-talk between them. Is this the case in plants? To approach this question, we have investigated enhancer-blocking insulators in the model plant *Arabidopsis thaliana* using two unrelated approaches. Firstly, we have developed an assay for the direct selection of insulators in *Arabidopsis thaliana* using a random oligonucleotide library. This assay helped us to define four novel insulator sequences named InI-3, InII-12, InIII-50, and InIII-78. Secondly, we have used genetic analyses to characterize potential insulator sequences originally from three non-plant species: UASrpg from the fungus *Ashbya gossypii*, BEAD1c from human T-cell receptors, and gypsy from *D. melanogaster*, that have been reported to function in *A. thaliana.* Our findings suggest that non-plant insulators and their protein binding sites function in plants and support the model of multiple, functional, different insulator sequences as was found in *D. melanogaster*. They also argue for the conservation of insulator mechanisms across species.

Résumé

Présentement, la majorité des recherches scientifiques portant sur les isolateurs utilise des espèces animales telles que la mouche des fruits *Drosophila melanogaster*, et non des espèces végétales. Les données accumulées jusqu'à date suggèrent qu'il existe plusieurs différentes séquences d'isolateurs dans *D. melanogaster*, chacune ayant sa propre protéine primaire de liaison, mais partageant tout de même les mêmes protéines secondaires de liaison. Ensemble, elles produisent des boucles de chromatine qui séparent les éléments régulateurs, tels que les amplificateurs et les promoteurs, dans des domaines distincts, empêchant ainsi tout phénomène de diaphonie entre eux. Est-ce que ceci est aussi le cas pour les plantes? Pour répondre à cette question, nous avons examiné les isolateurs amplificateurs-bloqueurs dans la plante modèle *Arabidopsis thaliana* en utilisant deux approches différentes et non liées. Premièrement, nous avons développé une analyse pour la sélection directe d'isolateurs dans *Arabidopsis thaliana* par l'entremise d'une banque d'oligonucléotides aléatoires. Cette analyse nous a aidée à définir quatre nouvelles séquences d'isolateurs nommées InI-3, InII-12, InIII-50 et InIII-78. Deuxièmement, nous avons utilisé des analyses génétiques pour caractériser des séquences d'isolateurs potentielles, originairement de trois espèces non végétales: UASrpg du champignon *Ashbya gossypii*, BEAD1c des récepteurs de lymphocytes T humain, et gypsy de *D. melanogaster*, qui ont tous été démontré à fonctionner dans A. *thaliana*. Nos résultats suggèrent que les isolateurs non végétales et leurs sites de liaison de protéines peuvent fonctionner dans les plantes et supportent le modèle qu'il y a plusieurs séquences d'isolateurs qui sont fonctionnelles et différentes, comme dans le cas de *D. melanogaster*. Notre étude soutient aussi la conservation des mechanisms d'isolateurs au travers des espèces.

Acknowledgments

To my supervisor Dr. Douglas Johnson, I want to thank you for giving me the opportunity to work with you and learn from you first-hand in the lab. From day one you have motivated me to think one step further, pushing me to be a better student and scientist. You have always been available for discussions and I appreciate your dedication and knowledge to this project, and to helping me finish this thesis.

To my committee members Dr. Thérèse Ouellet and Dr. Owen Rowland, thank you for your expertise and valuable discussions during all my committee meetings. You have both made a big difference in this project.

To my (long distance) lab mate and friend Lara Rasooli, thank you for all your lab help and training in the beginning and always being there to listen to me talk about science and life.

Thank you to all my undergraduate students, Hassan Badreddine, Onkar Bhanushaki, Adina Popescu, Linda Dam, Shukria Ahmadi, Liam McCarthy, Ka Mien Lam, and Krishna Gelda, for all your help and dedication to this project and just making the lab a brighter place.

Thank you to our visiting scholars Dr. Maolong Hu and Cairong Yang for your discussions and company in the lab.

Thank you to Phil Pelletier for your constant support in the common lab, Michelle Brazeau for your assistance in the growth chambers, and to Huguette Allard for all your help with the growth chambers and for sharing your knowledge of plant care, you will always be remembered.

Thank you to Dr. Julian Starr for the training of Sequencher, Étienne Léveillé-Bourret for putting so much time into helping me prep and send out sequences, Dr. David Currie for your assistance with my statistical analysis, and Julie Bilodeau for translating my abstract.

Thank you to Dr. Nicolas Corradi and your students for allowing me to enter regularly into your lab to use the lab's equipment.

I want to thank my friend and Biograd buddy, Rob Lalonde for all the thought provoking science discussions regarding my project and prepping me for seminars.

To all the Biograds I have met over the years, and especially those who have become my close friends, I am grateful to have met such talented people.

Last, but not least, thank you to my all friends near and far, my significant other Alex Plachkov, and my loving family who believed in me when I had a hard time believing in myself.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5-FC	5-Fluorocytosine
5F-dUMP	5-Fluoro-2′-deoxyuridine-5′-monophosphate
5-FU	5-Fluorouracil
5'HS4	5'constitutive hypersensitive site
5'ApoB	5' Apolipoprotein B-100 Insulator
35S46	Cauliflower Mosaiic Virus 35S core promoter
*At*gypsy-like	*Arabidopsis thaliana* gypsy-like
AGIP	*AGAMOUS* second intron-derived promoter
Ars	Arylsulfatase
BEAD	Blocking element alpha/delta
BEAF	Boundary Element-Associated Factors
BTB/POZ Domain	Broad-Complex Tramtrack Bric a Brac/Pox virus and Zinc finger
CaMV35S	Cauliflower Mosaic Virus 35S enhancer
CHA1 UAS	Catabolic L-serine (L-threonine) Dehydratase Upstream Activating Site
ChIP-Chip	Chromatin Immunoprecipitation with DNA Microarray
ChIP-Seq	Chromatin Immunoprecipitation with DNA Sequencing
codA	Cytosine Deaminase
CP190	Centrosomal protein 190kD
CTCF	CCCTC-binding factor
dCTCF	*Drosophila* CCCTC-binding factor
dTopors	*Drosophila* Topoisomerase I binding protein
dTTP	Deoxythymidine Triphosphate
E2	Even-skipped stripe 2 enhancer
E3	Eve stripe 3 enhancer
EDTA	Ethylenediaminetetraacetic acid
Fab-7	Frontabdominal-7 Insulator
Fab-8	Frontabdominal-8 Insulator
GAF	GAGA Factor
GUS	β-glucuronidase
HML and HMR	*Saccharomyces cerevisiae* silent mating loci
hptII	Hygromycin Phosphotransferase II
iaaH	Indoleacetamide Hydrolase
MAR	Matrix Attachment Region
ME	Myoglianin and Eyeless Boundary Insulator
Mcp	Miscadestral Pigmentation Insulator
Mod(mdg4)	Modifier of mdg4
NaCl	Sodium Chloride
NR	Nitrate Reductase
PCR	Polymerase Chain Reaction
PE	Proximal Element
PIp	*PISTILLATA* promoter
Rap1	Repressor Activator Protein 1
RB7 MAR	*Nicotiana tabacum* Matrix Attachment Region
RO	Repeat Organizer Insulator
SAR	Scaffold Attachment Region
Scs	Special Chromatin Structure Insulator
SDS	Sodium dodecyl sulfate

SF1	*Scr–ftz* Intergenic Region Insulator
STARs	Subtelomeric Anti-silencing Regions Insulator
Su(Hw)	Suppressor of Hairy Wing protein
TBS	Transformation Booster Sequence
TCR	T-cell receptor
T-DNA	Transfer-Deoxyribonucleic Acid
TEF	Translation Elongation Factor
Tris-HCl	Trisaminomethane hydrochloric acid
TSS	Transcription Start Site
TS	Thymidylate Synthase
UASrpg	Upstream Activation Site for ribosomal protein genes
VRE	Ventral Repression Element
ZW5	Zeste-White 5 Insulator

Preface (1)

Within the eukaryotic genome is a large amount of chromatin packed into the nucleus of cells. Although this chromatin is organized into units of chromosomes, special regulatory elements exist to further organize the chromatin into structurally and functionally independent domains. These special regulatory elements are coined insulators and contain specific protein binding sites critical for function. It is believed that associated insulator proteins bind to other proteins (primary protein-secondary protein interactions) to create loops in the chromatin establishing high-order structural domains blocking unwanted interference of regulatory elements or chromatin, from different domains. The discovery of multiple insulators found in single species and numerous insulators found in multiple species, supports models based upon the conservation of mechanisms across species. A better understanding of these models requires us to consider that it is not the specific proteins that interact and bind one another, but the protein domains (motifs) found within these proteins.

1.0 Introduction

1.1. Gene expression: a brief overview

The expression of eukaryotic genes involves a complex process in which the activation (or repression) of transcription occurs via the transcription complex where it binds to regions of the gene containing multiple regulatory elements within the nucleus of a cell. In protein coding genes, normally but not exclusively, the complex initially binds to an area of the chromosome lying 5' of the transcription start site (TSS) and the potential transcribed region, called the promoter. Regions of the promoter are responsible for both the particular pattern of expression of the gene and general control of transcription. Other elements such as repressors and enhancers can influence gene expression. Enhancers are sequences that can be situated at large distances from the TSS and act in a position and orientation independent manner to stimulate the expression of a gene via communication with the promoter, sometimes disrupting normal gene expression patterns (reviewed by (Catarino et al., 2017). In higher eukaryotes, only a subset of the genome is expressed and actively transcribed at one time, while the majority of genes remain in an inactive state. The

active and inactive gene states are characterized by two different chromatin types known as euchromatin and heterochromatin, respectively. Euchromatin has a structure that contains more loosely packed DNA and proteins (histones) with higher gene expression due to increased acetylation of histones and access for the binding of transcription factors compared to heterochromatin which is more tightly packed and methylated. These structures organize the eukaryotic genome into distinct domains that contain individual genes and gene clusters that have distinct patterns of expression both during development and in differentiated cells (West et al., 2002). Elements which create demarcations between these chromatin types are necessary for gene regulation, recombination and repair (Bickmore, 2013; Ciabrelli and Cavalli, 2015; Cremer et al., 2015; Misteli, 2007). As well, the precise control of gene expression relies on the interaction between transcriptional machinery of the promoter and protein complexes at the enhancer. The ability of the enhancer to act over large distances in an orientation independent manner suggests that there must be elements and mechanisms which exist to confine enhancers and promoters to domains which prevent unnecessary interactions between one another, such as elements that regulate different genes. The specific cis-acting DNA sequences that are responsible for creating these types of boundaries are known as insulators and are the main subjects of this thesis.

1.2. Insulators and their functions

Insulators are cis-acting DNA sequences which contain specific protein binding sites able to create boundaries within chromatin and protect regulatory elements of genes from unwanted interactions, or misexpression. They ensure the accuracy of transcriptional activity by 3 basic functions (Figure 1.1A, B, C). First, they prevent gene activation by blocking the interference between an enhancer and distant promoter when placed between the two. Second, they prevent

gene silencing by blocking the spread of heterochromatin into adjacent euchromatin. And lastly, insulators are able to protect against position effects using one of the two previously mentioned functions when insulator sequences flank both ends of a gene, for example following transgene insertion and expression, or a translocation in the absence of transgenesis.

Insulators are not restricted to one function; studies have shown that a single insulator may possess two of the above functions such as the scs (special chromatin structure) insulator which acts as both an enhancer-blocker and a protector against position effects in *Drosophila melanogaster* assay studies (Kellum and Schedl, 1992). They may also possess two functions in different species such as the gypsy element acting as an enhancer-blocker and heterochromatin barrier element in *D. melanogaster* and *Saccharomyces cerevisiae* assays, respectively (Donze et al., 1999; Geyer et al., 1986).

Testing sequences for the different functions require different experimental systems. The system used in this thesis is the enhancer-blocking system in which insulators are tested by inserting them between enhancers and promoters and testing their ability to block interference, as described in full detail in the Materials and Methods sections of the thesis. Other experimental systems studies have tested transcriptional silencing of silencers for barrier blocking abilities (Bi and Broach, 1999) or elements for their ability to protect against position effects (Kellum and Schedl, 1991).

1.3. Identified insulators in non-plant species

Insulators have been discovered in a wide range of species from yeast to vertebrates (Table 1.1). The function of insulator sequences across species suggests that these important regulatory

elements are evolutionarily conserved. The most highly characterized insulators have been studied in *D. melanogaster,* and include the gypsy insulator (Bender et al., 1983; Geyer et al., 1986), the specialized chromatin structures scs and scs' (Udvardy et al., 1985), Fab-7 (Gyurkovics et al., 1990), Fab-8 (Barges et al., 2000), SF1 (Belozerov et al., 2003), 1A2 (Chetverina et al., 2013), ME (Sultana et al., 2011), and MCP (Gruzdeva et al., 2005; Karch et al., 1985).

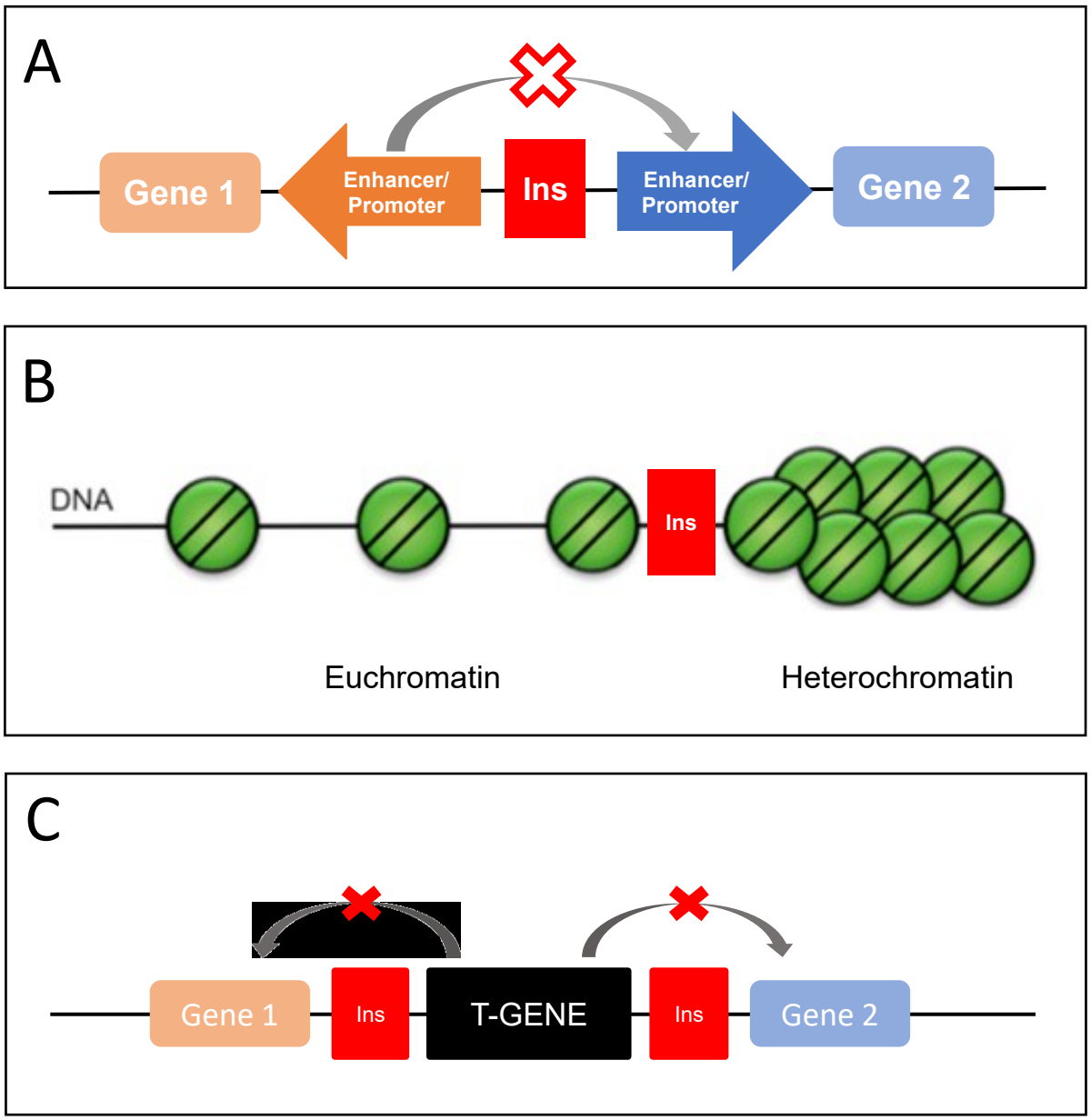


Figure 1.1: The 3 functional properties that define insulators. (A) Enhancer-blocking insulator sequences block enhancers from interacting with promoters when situated between the two elements. In this example, the Enhancer, which drives Gene 1, is blocked by the insulator (Ins) from interacting with the Promoter of Gene 2. (B) Barrier blocking insulators (Ins) prevent the linear spread of heterochromatin (condensed green spheres) from silencing euchromatin (gapped green spheres). (C) Insulator sequences within the T-DNA are able to protect against position effects using one of the two previously mentioned functions from (A) or (B) when flanked on both sides of the transgene (T-GENE).

To study the insulator mechanisms, work has been done mainly on the *D. melanogaster* species, a very useful model system due to the number of insulators discovered in the species and availability of molecular and genetic tools for their study. Initial work on the *D. melanogaster* insulators focused on the effects of copy number and position on enhancer-blocking strength in *D. melanogaster* embryos into which different transgenes had been introduced. Cai and Shen (2012) tested tandem gypsy sequences (arranged as direct repeats) within *D. melanogaster* blastoderm embryos testing an eve-lacZ reporter expression. Their system involved transgenics containing the zerknullt VRE (ventral repression element) enhancer and E2 (even-skipped stripe 2 enhancer) directing the eve-lacZ promoter. A single gypsy element has been previously shown to partially block the upstream VRE enhancer when inserted between the VRE and E2 enhancers (Cai and Levine, 1995). However, when they tested 2 tandem gypsy elements inserted between VRE and E2, instead of an enhanced blockage, the transgenic embryo exhibited a loss of enhancer-blocking activity compared to one. Yet, this affect only occurred when gypsy sequences were in tandem, since flanking VRE with two gypsy sequences more sufficiently blocked enhancer activity. They therefore concluded that it was the arrangement of multiple gypsy sequences and not the proximity that affects insulation. When tandem gypsy sequences are aligned, the two elements may preferentially interact with each other, excluding other interactions necessary to sequester the enhancer from the promoter and may even augment the enhancer-promoter interaction by looping out the intervening DNA (Cai and Shen, 2001).

Later studies set out to determine whether the abolishment of insulator activity by tandem arrangements (pairing) was a general characteristic among other chromatin insulators. Majumder and Cai (2003) tested the pairing of gypsy, scs and SF1 insulator combinations in a different assay which tested the effects of the insulators' ability to block a distal *twist* Proximal Element (PE)

enhancer and/or a proximal eve stripe 3 (E3) enhancer; both enhancers are active at different regions of embryos in the blastoderm stage. Their tests consisted of P-element constructs containing the two embryonic enhancers between divergently transcribed *eve-lacZ* and *white* reporter genes. They found that unlike homologous gypsy pairing (also tested in their experiments), heterologous combinations of gypsy and other insulators do not reduce enhancer-blocking activity, but instead increased enhancer-blocking activity compared to either single insulator alone, suggesting they can function independently or additively. This suggested that there are diverse mechanisms affecting insulator activity, and insulators do not interact obligatorily with each other but instead are selective upon interaction between themselves (when in tandem) and external sites.

In vertebrates, identified insulators include the human T-cell receptor α/δ locus BEAD element (Zhong & Krangel, 1997), the human 5'ApoB element (Antes et al., 2001), mice elements (Carlos Roman et al., 2011; Lunyak et al., 2007), the constitutive 5' HS4 site in the chicken β-globin LCR (Chung et al., 1993; Recillas-Targa et al., 2002) and the recently determined mammalian-wide interspersed repeats (MIRs) which function in humans (Wang et al., 2015). As discussed above, different research groups employ different experimental systems. In *S. cerevisiae*, there exist multiple insulators including UASrpg (Bi and Broach, 1999), STARs (Fourel et al., 1999), CHA1 UAS (Donze and Kamakaka, 2001), and HMR tRNA (Donze et al., 1999).

The discovery of multiple insulators in a single species, such as the insulators found in *D. melanogaster* and *S. cerevisae*, suggest the heterogeneity of insulators and leads to the hypothesis that all mechanisms may coalesce into a single ultimate pathway for their functionality despite the differences in protein binding sites between them, as will be described later. It has also been shown

that many of these insulators are functional in different species and may even act as both enhancer-blockers and chromatin barrier elements. The function of insulators across various organisms is evidence of its conservation and suggests these elements may utilize similar mechanisms even though the DNA sequences are different.

Table 1.1: Summary of insulators identified in non-plant systems. Represented is the organism of origin, insulator name, its specific binding protein, and references to its specific insulator function.

Family	Organism of Origin	Insulator Name	Binding Protein	Enhancer-Blocker	Chromatin Barrier	Protection against Position Effects
Drosophilidae	*Drosophila melanogaster*	gypsy	Su(Hw) (Suppressor of Hairy Wing) (Parkhurst et al., 1988)	(Cai and Levine, 1995; Geyer et al., 1986; Scott et al., 1999)	(Donze et al., 1999; Donze and Kamakaka, 2001)	(Markstein et al., 2008)
Drosophilidae	*Drosophila melanogaster*	scs (special chromatin sequence)	ZW5 (Zeste-White5) (Gaszner et al., 1999)	(Vazquez and Schedl, 1994; Zhong and Krangel, 1997)	(Cai and Levine, 1995; Kellum and Schedl, 1992)	(Kellum and Schedl, 1992)
		scs' (special chromatin sequence prime)	BEAF (Boundary Element-Associated Factors) (Zhao et al., 1995)			
Drosophilidae	*Drosophila melanogaster*	Fab-7 (Frontabdominal-7)	GAF (GAGA Factor) (Schweinsberg et al., 2004)	(Hagstrom et al., 1996; Zhou et al., 1996)		
Drosophilidae	*Drosophila melanogaster*	Fab-8 (Frontabdominal-8)	dCTCF *(Drosophila* CTCF) (Moon et al., 2005)	(Barges et al., 2000)		
Drosophilidae	*Drosophila melanogaster*	SF1 (*Scr–ftz* Intergenic Region)	GAF (Belozerov et al., 2003)	(Belozerov et al., 2003)		

Family	Organism of Origin	Insulator Name	Binding Protein	Enhancer-Blocker	Chromatin Barrier	Protection against Position Effects
Drosophilidae	*Drosophila melanogaster*	1A2	Su(Hw) (Golovnin et al., 2003)	(Chetverina et al., 2013)		
Drosophilidae	*Drosophila melanogaster*	ME (Myoglianin and Eyeless Boundary)	BEAF, GAF (Sultana et al., 2011)	(Sultana et al., 2011)		
Drosophilidae	*Drosophila melanogaster*	MCP (Miscadestral Pigmentation)	dCTCF (Holohan *et al.,* 2007)	(Kyrchanova et al., 2007)		
Phasianidae	*Gallus gallus*	5'cHS4 Chicken Globin (5' constitutive hypersensitive site)	CTCF (Bell et al., 1999)	(Chung et al., 1993; Recillas-Targa et al., 2002)	(Chung et al., 1993)	(Chung et al., 1993; Recillas-Targa et al., 2002)
Muridae	*Mus musculus*	B1 SINE (short interspersed nuclear element)	CTCF (Carlos Roman et al., 2011)	(Carlos Roman et al., 2011)		
Muridae	*Mus musculus*	B2 SINE (short interspersed nuclear element)	Unidentified	(Lunyak et al., 2007)		
Hominidae	*Homo sapiens*	BEAD1c (Blocking element alpha/delta)	CTCF (Bell et al., 1999)	(Zhong and Krangel, 1997)		
Hominidae	*Homo sapiens*	5'ApoB	CTCF	(Antes et al., 2001)		(Antes et al., 2001)
Hominidae	*Homo sapiens*	MIRs (Mammalian-wide interspersed repeats)		(Wang et al., 2015)		
Strongylocentrotidae	*Hemicentrotus pulcherrimus*	Ars (Arylsulfatase)	Unidentified	(Nagaya et al., 2001)		

	Xenopus laevis	RO (Repeat organizer)	CTCF (Bell et al., 1999)	(Robinett et al., 1997)		

Family	**Organism of Origin**	**Insulator Name**	**Binding Protein**	**Enhancer-Blocker**	**Chromatin Barrier**	**Protection against Position Effects**
Saccharomycetaceae	*Saccharomyces cerevisiae*	TEF2-UASrpg (Translation elongation Factor2-Upstream Activation Sequence)	3 Rap1 (Bi and Broach, 1999)		(Bi and Broach, 1999)	
Saccharomycetaceae	*Saccharomyces cerevisiae*	STARs (Subtelomeric anti-silencing regions)	Unidentified		(Fourel et al., 1999)	
Saccharomycetaceae	*Saccharomyces cerevisiae*	CHA1 UAS (Catabolic L-serine (L-threonine) Dehydratase Upstream Activating Site)	Unidentified		(Donze and Kamakaka, 2001)	
Saccharomycetaceae	*Saccharomyces cerevisiae*	HMR tRNA *Saccharomyces cerevisiae* silent mating loci transfer RNA)	TFIIIB, TFIIIC (Donze et al., 1999)		(Donze et al., 1999)	
Saccharomycetaceae	*Ashbya gossypii*	TEF-UASrpg (Translation elongation Factor-Upstream Activation Sequence)	2 Rap1 (Bi and Broach, 1999)		(Bi and Broach, 1999)	

1.4. Insulators in plants

To date, there have been few plant sequences tested for insulator function in plants (Table 1.2). Notable insulators include the transformation booster sequence (TBS) of *Petunia hybrida*, (Singer et al., 2011), and the At-gypsy-like element (Singer and Cox, 2013). TBS is a 2kb retrotransposon which contains a 516bp matrix attachment region (MAR) and is proposed to form the bases of chromatin loops (Mirkovitch et al., 1984). Currently, the majority of plant insulator studies have found various MARS/SARS elements to act as insulators (Allen et al., 1996; Allen et al., 1993; Breyne et al., 1992; Li et al., 2001; Mlynarova et al., 1994; Nagaya et al., 2001; Schoffl et al., 1993; Vain et al., 1999; Vandergeest et al., 1994; Xue et al., 2005).

In addition to plant specific insulators, several groups have demonstrated that insulators from non-plant species function in *A. thaliana*. An extensive study done by Gudynaite-Savitch et al. (2009) demonstrated that the insertion of one of several non-plant insulators between the 35S enhancer and the seed-specific napin promoter reduced the influence of 35S enhancer on napin expression and restricted the expression of the GUS reporter gene to seed tissue. The insulators which were tested in this study were BEAD-1 and its core sequence BEAD-1C (BEAD-A) of the human T-cell receptor α/δ locus (Zhong & Krangel, 1997), UASrpg from the filamentous fungus *A. gossypii* (Bi & Broach, 1999), Fab-7PRE, Fab-7bd, Mcp and Fab-8 from *D. melanogaster.* Of these elements, BEAD-1, BEAD-1c and UASrpg were able to significantly reduce non-specific interaction. In another study, the gypsy insulator was found to function as an insulator when two sequences flanked an *A. thaliana* transgene and protected it against positional effects (She *et al.,* 2010). As with previous (and subsequent) studies the exact nature of the insulator was not defined but assumed to be the same as the sequence in the non-plant species. In a further study, Singer *et. al.* unexpectedly found that a 1-kb bacteriophage lambda fragment cloned into *A. thaliana* acted

as an enhancer-blocking insulator in their *35S* enhancer and a flower-specific *AGAMOUS* second intron-derived promoter (*AGIP*) system. The study set out to originally test λ DNA fragments of varying sizes (1kb, 2kb and 4kb) as spacer sequences based on the assumption that the λ genome is free of DNA-binding motifs that might be recognized by plant regulatory factors. The result that the 1kb fragment functioned as an enhancer blocker and not the others, suggested that it is in fact possible for sequences from organisms thought to lack insulators to function as insulators in these assays and further suggests that larger, "random" sequences may contain insulators.

The ability of these non-plant insulators to function in plants suggest a conservation across species, and a possible conservation for the model that explains their function, which will be explored in later chapters of this thesis.

Table 1.2: Summary of insulators identified in plant systems. Represented is the organism of origin, insulator name, and references to its specific insulator function.

Family	Organism of Origin	Insulator Name	Enhancer-Blocker	Protection against Position Effects
Solanaceae	*Nicotiana tabacum*	RB7 SAR		(Allen et al., 1996)
Solanaceae	*Nicotiana tabacum*	Animal-analogous P1 SAR		(Breyne et al., 1992)
Fabaceae	*Phaseolus vulgaris*	Phaseolin 5' MARS (Matrix Attachment Region)		(Vandergeest et al., 1994)
Saccharomycetaceae	*Saccharomyces cerevisae*	ARS-1 SAR (Scaffold Attachment Region)		(Allen et al., 1993)
Strongylocentrotidae	*Hemicentrotus pulcherrimus*	Ars		(Nagaya et al., 2001)
Phasianidae	*Gallus gallus*	chicken lysozyme MAR		(Mlynarova et al., 1994; Nagaya et al., 2001)
Solanaceae	*Petunia hybrida*	TBS MAR	(Hily et al., 2009; Singer et al., 2011)	
Brassicaceae	*Arabidopsis thaliana*	*At*gypsy-like	(Singer and Cox, 2013)	
Drosophilidae	*Drosophila melanogaster*	gypsy		(Jiang et al., 2017; She et al., 2010a)
Saccharomycetaceae	*Ashbya gossypii*	*TEF*-UASrpg	(Gudynaite-Savitch et al., 2009)	
Hominidae	*Homo sapiens*	BEAD1c	(Gudynaite-Savitch et al., 2009)	
Brassicaceae	*Arabidopsis thaliana*	16bp NI29	(Gan and Xie, 2002)	
Siphoviridae	*Lambda (λ) phage*	1kb EXOB	(Singer et al., 2010)	

1.5. Critical protein binding sites

Evidence suggests that insulator function is mediated by trans-acting binding proteins which have been found to bind to specific sequences on the insulator. The first protein that was found critical for its insulator function was a zinc finger protein Su(Hw) that binds to the gypsy insulator in *D. melanogaster* (Parkhurst *et al.* 1988). Su(Hw) binds 12 copies of a 12 bp consensus sequence separated by AT-rich spacer sequences (Spana and Corces, 1990). Deletion analysis on the Su(Hw) binding region (containing 12 consensus sites) diminishes insulator function (Geyer et al., 1988; Peifer and Bender, 1988; Smith and Corces, 1992), and insertion of the Su(Hw) binding region alone is able to reproduce the function of the entire gypsy element (Holdridge and Dorsett, 1991; Geyer and Corces, 1992; Roseman et al., 1993). These results demonstrate that the Su(Hw) region is both necessary and sufficient for gypsy insulator function. It was later established that this protein does not function alone. The gypsy insulator bound to Su(Hw) interacts with two secondary proteins: Mod(mdg4) (Ghosh et al., 2001), and CP190 (Pai et al. 2004). Mod(mdg4) does not bind the gypsy insulator sequence directly, but instead interacts with Su(Hw) through its carboxy-terminal domain. In addition, Mod(mdg4) also contains a BTB domain in the N-terminal region to form homodimers to interact with Su(Hw) and the C-terminal region of the protein which is involved in interactions with the leucine zipper and adjacent regions of the Su(Hw) protein (Ghosh *et al.,* 2001). The centrosomal CP190 protein contains both a BTB/POZ domain and three C2H2 zinc fingers, thus capable of associating physically with both Su(Hw) and Mod(mdg4)2.2 and colocalizing with both proteins on polytene chromosomes (Pai *et al.* 2004).

Studies in other *D. melanogaster* insulators found a common pattern in which all primary insulator binding proteins, the 12-zinc finger Su(Hw) described above, 8-zinc finger ZW5 of scs (Gaszner *et al.,* 1999), 1-zinc finger BEAF-32 of scs' (Zhao *et al.,* 1995), and 2 zinc-finger GAGA

of SF1 (Belozerov *et al.,* 2003), are all associated with variants of Mod(mdg4) and with CP190 for function. Although these proteins recognized different DNA sequences, their similarities in that they contain zinc-finger motifs and share the BTB domain-containing protein CP190, suggest there is a conserved mechanism being used in this species. In the section describing insulator models, we will describe how these findings in *D. melanogaster* can help to define a general insulator model.

Characterization of vertebrate insulators have also revealed binding proteins. The most studied vertebrate binding protein, CCCTC-binding factor (CTCF), is a multi-functional protein found in numerous species. It is an 11-zinc finger protein that is highly conserved and ubiquitous (Filippova et al., 1996; Klenova et al., 1993). The binding sites associated with this protein are present at different genetic loci, however are all located between independently regulated genes (Bell, West, & Felsenfeld, 1999) and present in various chromatin states including intergenic, transcribed regions, promoters, and enhancers (Chen et al. 2012). CTCF is involved in both chromatin barrier and enhancer blocking insulator function. Studies have attempted to sequence the human genome for all possible CTCF binding sites to determine more potential insulators. Work by Kim *et al*. (2007) found 13,804 sites via genome wide mapping. Multiple studies have tested their insulator sequences for the CTCF site by bioinformatic sequence analysis on the insulator sequences they determined numerous sequence identities with CTCF (Singer et al., 2010). Although further studies must be done to determine whether they are truly functional as insulator proteins, this suggests the possibility of CTCF-like proteins in plants and a conservation between vertebrates and plants. It has also been determined that insulator sequences from *D. melanogaster* contain a homolog of the vertebrate CTCF insulator protein called *Drosophila* CTCF (dCTCF). dCTCF has 12 zinc fingers and shown to be necessary for Fab-8 (Mohan *et al.*, 2007; Gerasimova *et al.,* 2007),

Mcp and Fab-6 insulators of the bithorax complex (Holohan *et al.,* 2007). The protein also shares a similar characteristic to the previously mentioned *D. melanogaster* proteins in that it interacts with CP190 protein (Mohan *et al.*, 2007; Gerasimova *et al.,* 2007) which is supported by genome-wide mapping of dCTCF and CP190 sites (Bartkuhn *et al.,* 2009; Bushey *et al.,* 2009). This homolog is a strong indication of the further conservation of insulators and their mechanisms across species, as well as the potential for a CTCF homolog potentially in plants, which have not been thoroughly studied. The presence of zinc finger motifs in many of these proteins provides a mechanism for protein binding but at the same time makes it difficult to define which one(s) of the many zinc finger proteins are actually used as insulators.

The conservation of DNA binding proteins and/or protein motifs (for example zinc-fingers) among species has not been fully studied in plant insulators. However, the possibility that an insulator homolog in plants may exist was studied by Stinger and Cox (2012), in which a gypsy-like sequence from *A. thaliana* exhibited enhancer-blocking activity in transgenic plants. We have set out to characterize potential insulator protein binding sites in this thesis, from the functional sequences described in later chapters. This study examined specific genetic locations which are critical for insulator function in plants by characterizing the sequences via deletions, mutations and inversions of putative insulator sequences. These initial findings support the theory of insulator conservation among species and the notion that there are multiple protein interactions involved in their mechanisms which ultimately lead to the same pathway.

1.6. Insulator Models

There are numerous models that have been postulated to help understand the function and mechanisms of insulators. Models for enhancer-blocking activity of insulators are the focus of this

thesis. They involve either functional interactions between the insulator and enhancer elements or physical interactions between insulator elements (Valenzuela & Kamakaka, 2006).

Early evidence of the interaction between multiple insulators such as scs, gypsy and SF1 insulators through heterologous and homologous pairings showed that various combinations do not always reduce enhancer-blocking activity, suggesting insulators may function additively or independently (Majumder et al., 2009). These observations led to the "promoter-decoy model", a transcriptional model in which the insulator protein complex interacts with enhancer-bound proteins intercepting the enhancer signal into a non-productive interaction (Geyer, 1997). Another transcriptional model, the "facilitator model" or "physical barrier model", proposes that insulators may also interfere with the function of facilitator proteins by preventing the enhancer from communicating with the promoter (Dorsett, 1999).

The most widely accepted model today is one in which insulators act to create a physical organization of chromatin fibers into independent structural and functional domains separating enhancers and promoters in a way which blocks the two elements from communicating. This model is the "chromatin loop model", where primary and secondary insulator proteins are involved in creating loops in the chromatin (Figure 1.2). Secondary or anchoring proteins are those which bind to "chromatin anchor points" such as the nuclear matrix or nuclear envelope/lamina, providing steric and topological hindrance to enhancers and promoters (Matharu & Ahanger, 2015). These proteins interact with the primary insulator proteins which are bound to the insulator themselves, creating the loops. This model was first illustrated by the gypsy insulator of D. melanogaster. The gypsy insulator complex contains multiple proteins responsible for its function: Su(Hw) is a DNA binding protein that directly interacts with the gypsy element in a sequence specific manner and is responsible for recruiting Mod(mdg4) via protein-protein interactions. It

has been shows that Mod(mdg4) is required for insulator function (Gause, Morcillo, & Dorsett, 2001; Ghosh, Gerasimova, & Corces, 2001). The protein CP190 then interacts with both Su(Hw) and Mod(mdg4), forming a cluster of insulator bodies at multiple gypsy sites and forming a chromatin loop separating the enhancer and promoter into distinct loops. Mod(mdg4) then attaches the chromatin to the nuclear lamella through its interaction with a final protein, the dTopors, which has been shown to be associated indirectly with CP190, and directly in vivo with Mod(mdg4)2.2 assisting in nuclear organization (Capelson & Corces, 2005) (Figure 1.3).

The model previously described was also proposed for other insulators of D. melanogaster which have been found to contain different primary and secondary binding proteins interacting with CP190 (Figure 1.3) (Garudatta and Corces 2009), providing a basis for understanding D. melanogaster insulator mechanisms which have been discovered and possibly ones which remain to be discovered.

This is the simplest model to date and provides an experimental approach to defining insulation pathways. Although it only includes a few proteins or their variants, it is not certain that these are the only ones involved. Nonetheless, insulator studies have shown a variety of evidence to support all three models and it seems no single model is able to explain all insulator function suggesting there is no one single, all-inclusive model of insulator function currently (Yoon et al. 2007).

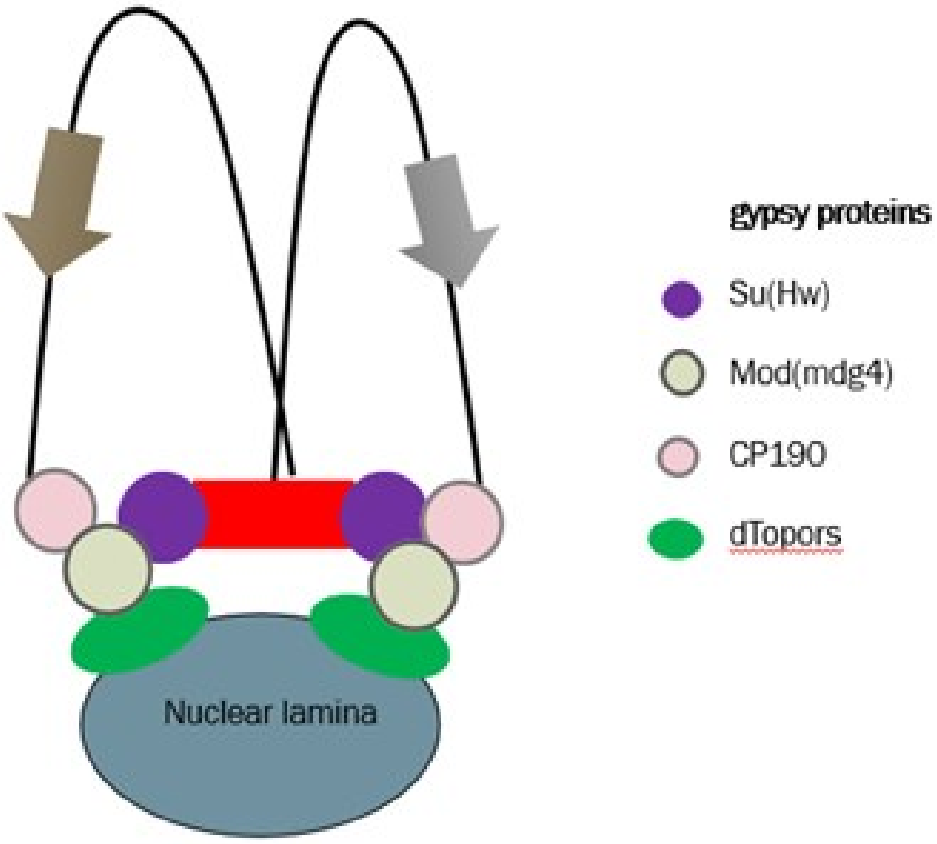


Figure 1.2: Model of gypsy insulator function. The formation of chromatin loops in the gypsy element via the interaction between the insulator protein complex consisting of Su(Hw), Mod(mdg4) and CP190 with dTopors which is attached to the nuclear lamina of the cell. This mechanism separates the enhancer (brown arrow) and promoter (grey arrow) into separate domains blocking interactions.

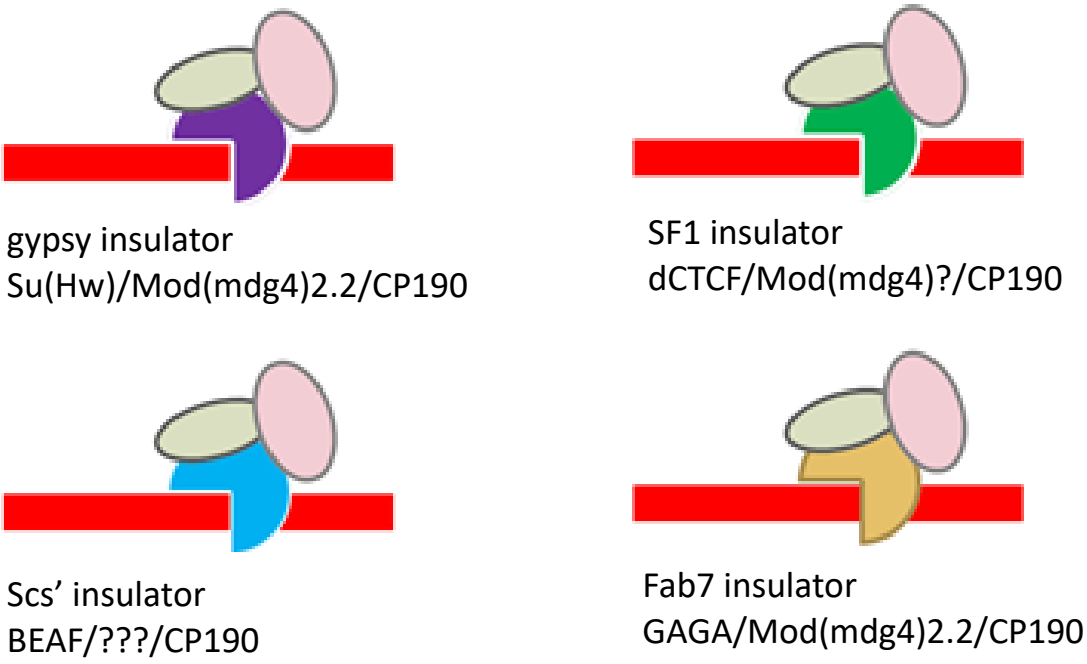


Figure 1.3: Drosophila insulators and their binding proteins. Four different D. melanogaster insulators (red) are shown with their corresponding primary DNA binding proteins (purple, green, blue, yellow) and secondary Mod(mdg4) protein isoform (grey) interacting with the common CP190 nuclear protein (pink). As seen in the diagram, the insulators gypsy and Fab7 contain the Mod(mdg4) isoform: Mod(mdg4)2.2. The SF1 lacks this specific isoform however contains a different variant. It is not known whether scs' contains a Mod(mdg4) variant.

1.7. Insulator Applications

The capacity of an insulator to block unwanted expressions within a gene is an important tool to consider in biotechnology. The insulator's powerful ability to separate regulatory elements such as enhancers and promoters from interacting, act as barriers between heterochromatin and euchromatin or prevent position effects, can be used when creating transgenics containing multiple transgene insertions. Both fundamental plant research and agricultural research could benefit from these sequences. The need for creating constructs containing multiple introduced genes will require elements that create chromatin boundaries to prevent unwanted expression, and the usage of insulators would be beneficial. However, in order to utilize sequences, a better understanding of their basic functions is required. To date, few studies have characterized plant insulators, or their associated proteins. This thesis describes a new experimental approach which tests sequences for insulator function and has proven to produce novel insulator sequences which function in *A. thaliana.* In addition, a detailed analysis of multiple non-plant insulators will be described.

Preface (2)

The next two chapters present the data, analyses, and interpretations for the two main projects of this thesis. Chapter 2 will present an experimental approach to select insulator sequences and discusses the discovery of novel insulator sequences from a random oligonucleotide library. Chapter 3 will present the findings for the testing and manipulations of previously identified insulators from non-plant species. We have chosen to write these chapters in the form of potential papers and include results which belong to the author, unless otherwise indicated in the 'Statements of Contributions'. Experiments could not have been achieved without the help of numerous dedicated undergraduate students over the years: Hassan Badreddine, Onkar Bhanushaki, Adina Popescu, Linda Dam, Shukria Ahmadi, Ka Mien Lam, Krishna Gelda, and Liam McCarthy.

Publication Plan:

Paper 1
Running Title: Direct selection of DNA insulators in Arabidopsis thaliana.
Authors: Anh Tran, Loreta Gudynaite-Savitch, Lara Rasooli, Batool Gandorah, Douglas A. Johnson.

Paper 2
Running title: Genetic analysis of non-plant DNA insulators in Arabidopsis thaliana.
Authors: Anh Tran, Douglas A. Johnson

2.0 An assay for the direct selection of functional insulators in *Arabidopsis thaliana*: validation using a random oligonucleotide library

2.1. Introduction and previous work

Few plant insulators have been subjected to detailed analyses. Based upon studies of insulators in Drosophila melanogaster (reviewed in the Introduction) we hypothesize that plants may contain several different sequences with insulator function. Therefore, it is necessary to identify more of these sequences for a more thorough understanding of their functions and mechanisms within plants, as well as a better comprehension of their interaction with elements that regulate plant gene expression. We know from the work of Singer et al. (2010) who unexpectedly discovered a 1-kb bacteriophage lambda sequence to have insulator function, that non-plant DNA sequences can also fortuitously possess insulator function in Arabidopsis thaliana. This observation led us to search for novel DNA sequences that can act as functional insulators in plants. This study also investigated the potential binding sites responsible for insulator function in our novel insulator sequences to study and develop further the models of plant insulator function.

To select for insulator sequences de novo, a three-step selection method was developed. Candidate sequences were cloned sequentially into three selection/screening vectors (pC1, pB31 and pL1). Insulator function in this assay was defined as the ability of these sequences to block enhancer-promoter interactions within each of the three vectors. Constructs were transformed into A. thaliana and transgenic samples (leaves, flowers and siliques) were tested for GUS expression and scored. The details of the methodology for the production of the random oligonucleotide library have been described in the thesis of Batool Gandorah, and only a summary is described in

the Materials and Methods section below. To validate the methodology, we generated a random oligonucleotide library as a potential source of novel insulator sequences.

2.2. Hypothesis and Objectives

The hypothesis for this chapter is that plants have multiple insulator types each capable of blocking unwanted enhancer-promoter interactions. To test this hypothesis, we identified functional novel plant insulators by selecting and screening random oligonucleotide sequences in *A. thaliana.*

2.3. Materials and Methods

2.3.1. Transformation Vectors

All vectors were constructed by Dr. L. Gudynaite-Savitch using the pCAMBIA series of plasmids . These vectors include pC1, pB31, and pL1 (Figure 2.1A, B, C).

pC1 (pCAM1300-35S46-codA) is a negative selection vector (Figure 2.1A). The vector contains the conditional negative selective marker gene cytosine deaminase (codA) (Perera et al., 1993) expressed by the core CaMV35S promoter (35S46) (Covey et al., 1981). The level of expression is increased by CaMV35S enhancers ("35S double enhancer") which is responsible for driving the hygromycin resistance gene (hptII) used for selection in plants. The codA gene is a substrate-dependent conditional negative selectable marker. The gene encodes a cytosine deaminase that catalyzes the deamination of cytosine into uracil (Andersen et al., 1989; Dubeau et al., 2009; Stougaard, 1993) and converts the non-toxic 5-Fluorocytosine (5-FC) present in selective agar into the highly toxic metabolite 5-Fluorouracil (5-FU). The product produced in planta from this reaction, 5F-dUMP, irreversibly inhibits thymidylate synthase (TS), and as a result the cells are

deprived of deoxythymidine triphosphate (dTTP) necessary for DNA synthesis, which leads to cell death of both eukaryotic and prokaryotic organisms (Mullen et al., 1992) (Appendix, Figure A2.1). The presence of a functional insulator located between CaMV35S enhancer and 35S46 will block enhancer influence on codA and plants will survive selection. DNA fragments were inserted upstream from the 35S46 to test their ability to block the influence of the CaMV35S enhancer and thus expression from the 35S46 promoter.

pB31 (pCAMBIA1300-35S46-GUS) is our screening vector in which codA from pC1 is replaced by β-glucuronidase (GUS) reporter gene (Jefferson et al., 1987), leading to blue staining in various tissues (Figure 2.1B). GUS expression relies on the CaMV35S enhancer influencing the 35S46 core promoter. The presence of an insulator between these two elements will block this interaction, eliminating the production of GUS resulting in no blue stain, while the absence of a functional insulator will result in the staining of flowers, leaves and siliques, reflecting the constitutive expression of the 35S46 core promoter influenced by the CaMV35S enhancer.

The final vector, pL1 (pCAMBIA 1391 napin-GUS) uses a similar strategy to pB31 except that GUS expression is now controlled by the napin seed-specific promoter (Ericson et al., 1991) instead of the 35S46 core promoter (Figure 2.1C). The napin promoter is also influenced by the CaMV35S enhancer, which drives non-specific GUS expression, showing blue staining in all tissues. However, the presence of an insulator cloned between the two elements drives specific GUS expression directing staining only in seeds. In addition to seed specificity, this vector is also able to test our sequences for repressor activity, which pB31 is unable to do. In the presence of a functional repressor, insertion of our sequence between the napin promoter and CaMV35S enhancer would block all gene expression, resulting in a lack of both specific and non-specific staining in transgenic tissues (Gudynaite-Savitch et al. 2007).

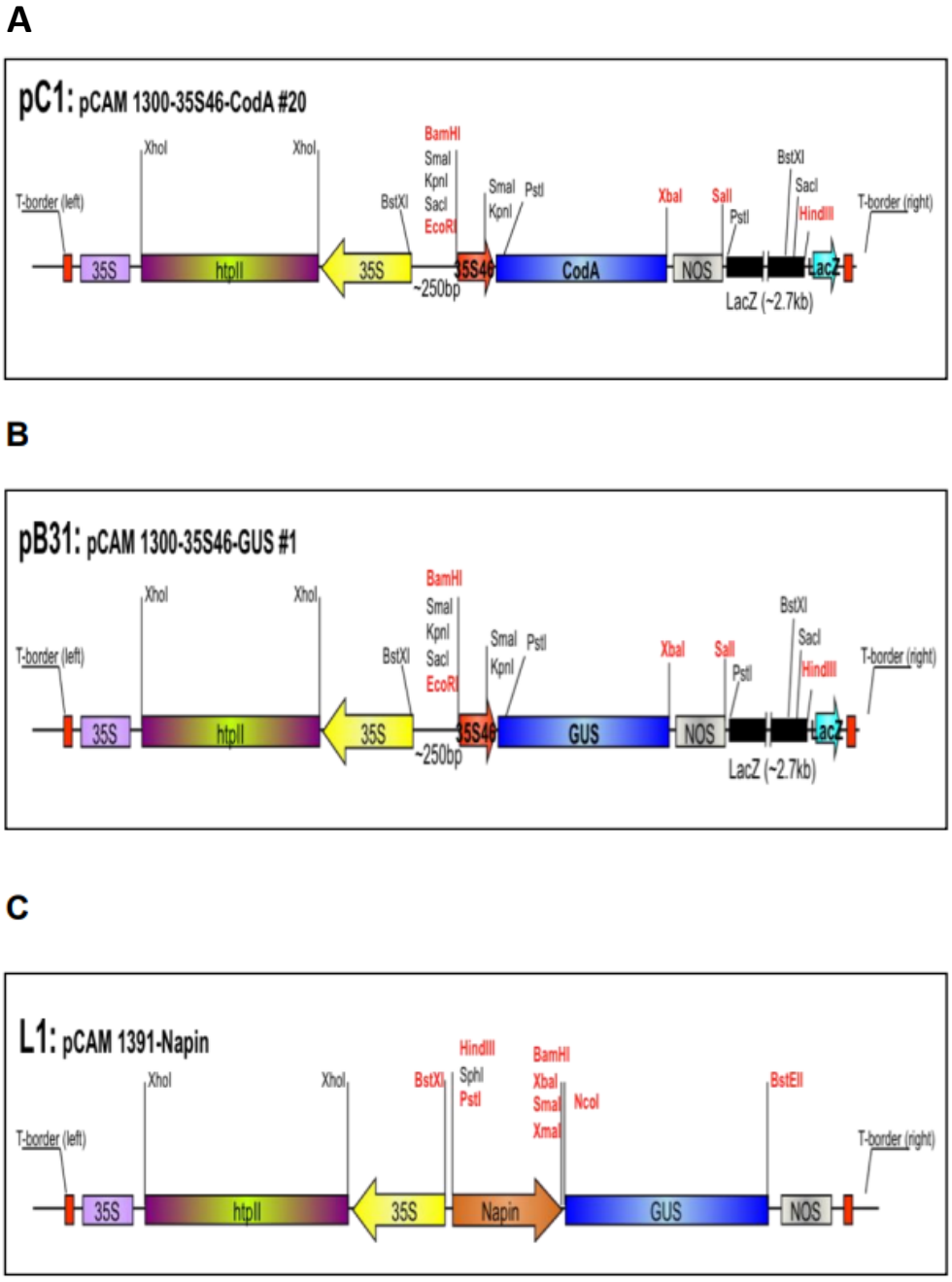


Figure 2.1: Summary of the 3 plasmid vectors used for selection and screening of potential insulator sequences. (A) Plasmid vector pC1(pCAM1300-35S46-CodA); used for the isolation of potential insulator sequences by negative selection using the CodA gene (blue box). (B) Plasmid vector pB31 (pCAM1300-35S46-GUS); used to confirm the presence of a potential insulator sequence using the GUS gene for screening. (C) Plasmid vector pL1; used to confirm the presence of a potential insulator sequence using the GUS gene for screening and to test that insulator function is not promoter-dependent. All 3 vectors were constructed by Dr. L. Gudynaite-Savitch. Coding: the red boxes are T-DNA borders; the selectable marker gene consists of hygromycin phosphotransferase (hptII, green-purple box) with the CaMV 35S terminator at the 3' end (purple box) expressed from the CaMV 35S promoter (yellow box) at the 5' end; the marker gene consists of GUS (blue box) with the Nos terminator at the 5' end (grey box) expressed from the 35S46 promoter (red arrow) or napin seed-specific promoter (orange arrow). Figures used with permission from author, Dr. L. Gudynaite-Savitch.

2.3.2. Generation of a random oligonucleotide library

Potential insulator sequences were initially constructed by the creation of a random oligonucleotide library by Dr. L. Gudynaite-Savitch. The oligomer: 5'-AGTGGATCCGAGACAAGC-(BamHI)(N124)(EcoRI)-CCTCCTCCTGAATTCTGC 3' was synthesized at McMaster University. From 5' to 3', the sequence contained an 18b PCR primer binding sequence with a BamHI restriction site, a 124b random oligonucleotide fragment and an 18bp PCR primer binding sequence with an EcoRI restriction site. Prior to cloning, this was converted to ds DNA via PCR using the CLO-Lib-F and CLO-Lib-R primers (Table 2.1) Sequences were digested with BamHI and EcoRI and cloned into pC1 which had also been digested with BamHI and EcoRI. Prior to transformation of *A. tumefaciens*, the library was amplified by transformation into electrocompetent *E. coli* DH10 cells (Invitrogen).

This library was the source of all sequences cloned into the pC1 vector for negative selection. Inserts were generated by Polymerase Chain Reaction (PCR) and cloned from pC1 into pB31. Potential insulator sequences were then cloned from pB31 to pL1 using a variety of methods as described in the following section.

2.3.3. Molecular techniques for selection and screening of putative insulator sequences

All (PCR) experiments underwent the following cycle parameters (1) 95°C denaturation for 5 minutes, (2) 94°C for 30 seconds, (3) 55°C annealing for 30 seconds, (4) 72°C extension for 30 seconds. Steps 2-4 were repeated 31 times and a final extension step at 72°C for 5 minutes was performed to ensure proper extension of DNA. PCR products were then run on 1.25% agarose gel

at 90 Volts for approximately 45 minutes to identify and/or characterize the PCR products. All PCR primers are summarized in Table 2.1.

Various cloning techniques were used. Restriction enzyme and ligation for fragment cloning and cloning into pGEM®-T Easy (Promega Corporation) were done following the manufacturer's recommendation. The Takara Bio USA, Inc In-fusion® HD cloning technique was also used following manufacturer's instructions and online primer design protocols[1].

Plasmid constructs with DNA fragment of interest were transformed into DH5α E. coli cells (Life Technologies Inc. Burlington, Canada) according to the manufacturer's protocol. E. coli transformants were selected on 200 μg/mL of ampicillin (pGEM-T Easy, and other vectors used for cloning intermediates) or 50 μg/mL kanamycin (for pC1, pB31 and pL1 vectors). Plasmid constructs with DNA fragment of interest were also transformed into GV3101 A. tumefaciens by electroporation (Wang et al. 2006). Transformants were selected on agar media containing 50 μg/mL kanamycin and 20 μg/mL rifampicin.

Plasmid preparation for sequencing of clones used GeneElute Plasmid Mini-Prep Kit (Sigma-Aldrich) and the EZ-10 Plasmid DNA Minipreps Kit (Bio Basic Inc) following the manufacturers instructions to isolate DNA for cloning and sequencing.

[1] http://www.clontech.com/US/Products/Cloning_and_Competent_Cells/Cloning_Resources/Online_In-Fusion_Tools

Table 2.1 List of PCR primers, target of amplification and sequence.

Primer Name	Target of Amplification	Sequence
CLO-Libr-Fbpx	Oligomer fragments	5'-CGTTCTAGACTGCAGAGGATCCG AGACAAGC-3' (31mer)
CLO-Libr-Resx	Oligomer fragments	5'-AGTTCTAGAGTCGACGAATTCAGGA GGAGG-3' (30mer)
M13For	Cloning vectors	5'-CGCCAGGGTTTTCCCAGTCACGAC-3' (24mer)
M13Rev	Cloning vectors	5'-AGCGGATAACAATTTCACACAGGA-3' (24mer)
SALK_049131_RP2	Genomic *A. thaliana* DNA	5'-GTCTCTACCGTACGCGCTTC-3' (20mer)
SALK_049131_LP2	Genomic *A.thaliana* DNA	5'-GGTTTGCATTTGACCTTTCG-3' (20mer)
1300LacZ-For	pB31 vector insert	5'-CACTCATTAGGCACCCCAGG-3' (20mer)
GUS5'-Rev	pB31 vector insert	5'-GTGGCTAGCTTGTTTGCCTC-3' (20mer)
1381F	pL1 vector insert	5'-CAGATAGCTGGGCAATGGAATC-3' (22mer)
pL1F	pL1 vector insert	5'-ACCATGTTGGGCCCGGCG-3' (18mer)
NapinSeqR	pL1 vector insert	5'-ATCCTCGAAACTCTACTCC-3' (19mer)
CLO_InfusionFor	In sequence Infusion	5'- CCGGCGCGCCAAGCTTAAGCTTGAAT TCAGGAGGAG-3'
CLO_InfusionRev	In sequence Infusion	5'- TTTCGATAATTCCGCTGCAGCTGCAG GGATCCGAGACA-3'

BEAD-INVFor	BEAD1c PCR	5'-CTGCTGCAGCAGTAATACGGGTA GCTGG-3’(28)
BEAD-INVRev	BEAD1c PCR	5'-TGCAAGCTTCTGTGCTTGGGTATGG TCAG-3’(29)
BEAD-F2	BEAD1c PCR	5' -CTGAAGCTTCAGTAATACGGGTAGCTG G 3'- (28)
BEAD-R2	BEAD1c PCR	5' -TGCCTGCAGCTGTGCTTGGGTATGGTC AG- 3' (29)
BEADA_delinfF	BEADA infusion	5'CCGGCGCGCCAAGCTTTTCAGTAATACG GGTAGCTG-3' (39mer)
BEADA_delinfR	BEADA infusion	5'-TTTCGATAATTCCGCTGCAGCTGCAGCT GTGCTTGGGT-3' (39mer)
CTC_infF	UAS_ΔCTC sequence for infusion	5’-CCGGCGCGCCAAGCTTAAGCTTCGAAA AATTACGGC-3’
CTC_infR	UAS_ ΔCTC sequence for infusion	5’-TTTCGATAATTCCGCTGCAGCTGCAGTT TCTCTACAGGG-3’

2.3.4. Plant material and growth conditions

Seed sterilization was performed on all wild type and transgenic seeds prior to planting in soil or plating on agarose. The desired number of seeds was transferred into a 15 mL conical tube. Approximately 10 mL of tap water was added to the tube which was agitated for 30 minutes using an automatic shaker to hydrate the seeds. The seeds were left to settle and the water was removed, 10mL of 70% alcohol was added and the seeds slowly mixed using the automatic shaker for 5 minutes. The solution was removed and 10ml of freshly made 10% bleach, and 0.1% SDS was added to the tube and was mixed using the automatic shaker for 5 minutes. The solution was removed and 10mL of double-distilled water was added to the tube, gently mixed, and decanted 4 times to remove residual sterilizing solutions.

To plant on soil 10 mL of 0.1% agarose is added to the tube and mixed, followed by the application of seeds using a Pasteur pipette. Application of seeds on selective plates involved the addition of a thin layer of seeds mixed with 0.1% agarose. Plates were dried in fume hoods for approximately 3 hours or until agarose had completely dried on the surface, to lessen contamination.

A. thaliana wild-type seeds (Columbia) were initially grown on soil (Premier, Pro-Mix PGX) that was rehydrated with water and a small amount (5 g/L) of MiracleGro. The soil was compacted into individual pots and covered with a plastic casing for 2 days. The seeds were then grown under 100 µmol/m2/s photosynthetically active radiation (PAR) day/8h at 22 oC. The plants were watered every 2-3 days rotating between Hoaglands nutrient solution (Gudynaite-Savitch et al. 2007), a mixture of fertilizer and tap water, and regular tap water, until mature enough for transformation.

Mature A. thaliana wild type plants underwent Agrobacterium-mediated transformation using the floral dip method (Clough & Bent, 1998). Seeds were then collected and grown for selection on Murashige and Skoog medium (Murashige and Skoog, 1962) containing 30 µg/ml hygromycin-B for selection of plants carrying the T-DNA insert and 250 µg/ml timentin to inhibit bacterial growth. Seeds were grown on selective media until shoots and roots could be seen and then transferred to a second selective plate for approximately 3 weeks in a growth cabinet until visible roots and leaves formed. Transgenic plants were selected and then transferred to soil for further growth. Once plants were fully mature, approximately 3-4 weeks, flowers, leaves and siliques were collected into a 1.5 ml microtube. Plants were kept for another week in growth chambers for a second collection of siliques. Each sample contained a minimum of 3 tissues - flowers, leaves, siliques – in some cases with stems attached. For siliques, a second set of tissues were collected a week later to eliminate the possibility of immature siliques. In addition, two cauline leaves were collected and stored at -20 oC for DNA analysis. GUS histochemical staining was carried out as described by (Malik et al., 2002). The intensity of GUS staining was determined visually using the following scale: 0 (none) 1 (weak), 2 (medium), 3 (high) (Gudynaite-Savitch et al. 2007; Figure 2.2A). For pB31constructs, we included samples that contained non-specific staining and samples which contained no staining. For pL1 constructs, specific staining, non-specific staining, and samples which contained no staining, were calculated. Samples with specific staining included all samples which had only staining in seeds. Non-specific staining included all samples which contained staining in at least one tissue of flower, leaf or silique. Examples of specific and non-specific staining, as seen in pL1 transgenics, are shown in Figure 2.2B.

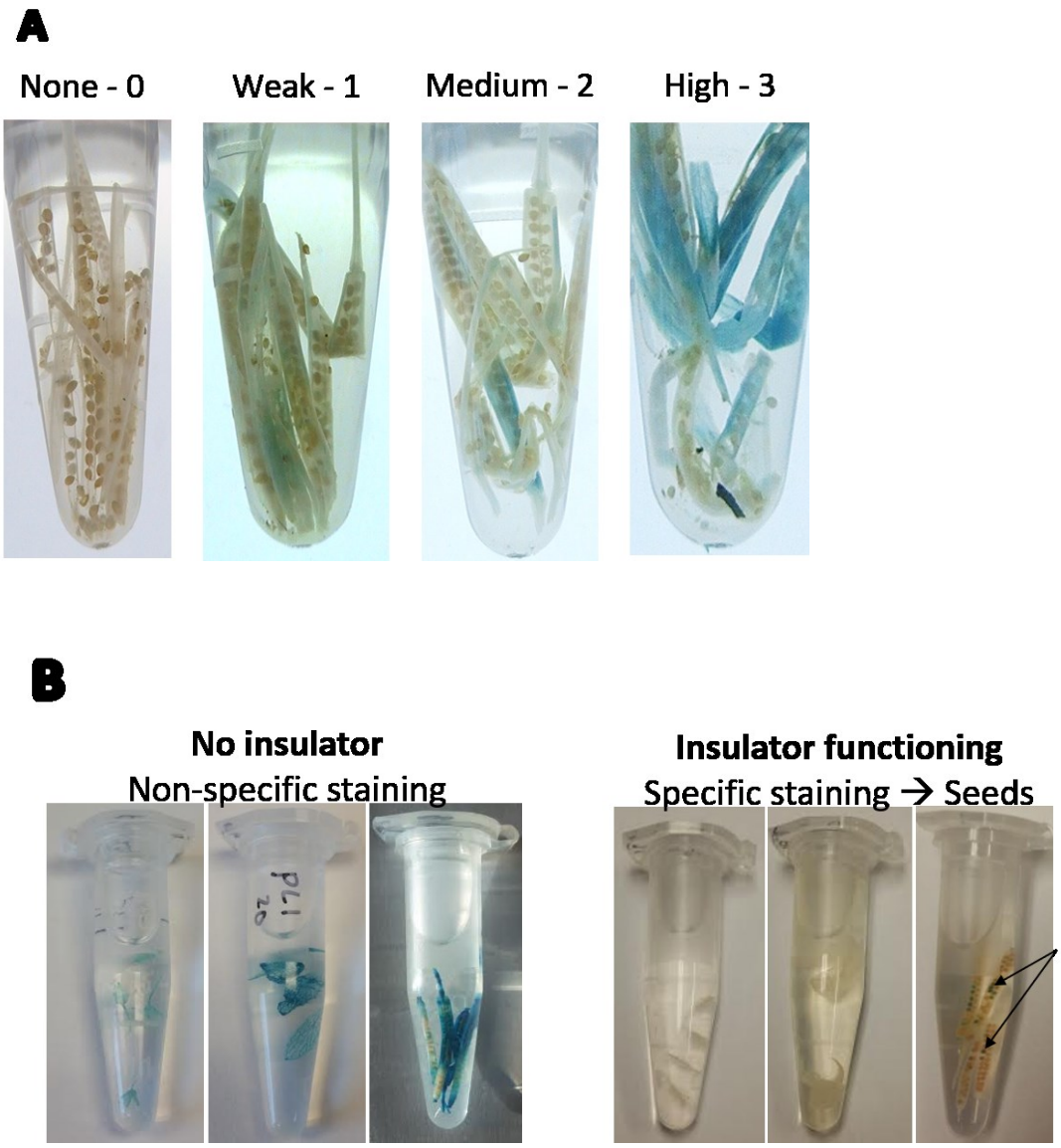


Figure 2.2: Transgenic plant tissues expressing GUS expression. (A) Scale used to assess the intensity of GUS staining, from 0 to 3. 0 indicates no staining in any tissue sample, 1 indicates weak blue staining in at least one tissue sample, 2 indicates the staining intensity is medium in at least one tissue sample, and 3 indicates staining intensity is strong in at least one tissue sample. **(B)** Examples of non-specific staining in transgenic tissues (no insulator function) and specific staining in transgenic seeds only, driven by napin promoter in pL1 vector.

2.3.5. Plant molecular techniques for PCR testing and insert confirmation

Transgenic A. thaliana genomic DNA for PCR and/cloning was extracted using a small-scale procedure (McKinney et al., 1995). The following modifications were used: plant tissue was ground with a sterile, plastic pestle and 0.5 mL of DNA extraction buffer (100 mM Tris-HCl pH 8.0, 250 mM NaCl, 10 mM EDTA, 0.5% SDS) was added to the tube and underwent more grinding. The tube was placed on ice to homogenize for 2 minutes. The samples were centrifuged for 2 minutes at 13000 rpm in a microfuge. To recover the DNA, 0.3 mL of supernatant was transferred to tubes containing 0.3 mL of isopropanol, the tubes were mixed thoroughly by vortexing and incubated at room temperature for 5 minutes, and then were centrifuged for 5 minutes at 13000 rpm. The supernatant was removed and the pellet was washed with 0.75 mL of ice cold 70% ethanol. Following centrifugation at 13000 rpm for 5 minutes, the supernatant was removed and the pellets were dried using the airflow in the fume hood. The dried pellets were resuspended in 0.1 mL of TE mix (10 mM Tris, 1 mM EDTA, pH8.0) and were mixed by vortexing. The mix was spun using the microfuge for 2 minutes and 0.04 mL of the supernatant was transferred into a sterile, labeled microfuge tube. These DNA preparations were stored in a 4°C unit for storage and use.

Extracted DNA from each transgenic A. thaliana sample was tested with two PCR primer sets. The first PCR run used SALK_049131_RP2 and SALK_049131_LP2 primers to amplify the Atchrom-1 region 11,078,981-11,078,219 with expected size of 762 bp; the presence of this sequence confirmed successful DNA extraction and quality of A. thaliana DNA. An example of SALK amplification is represented in. The second PCR run used vector specific primers to confirm successful cloning of the sequence of interest with expected size of 600 bp or 900 bp, depending on size of the insert. Examples of SALK amplification and vector specific amplifications (LacZF

and GUS-R primers for amplification of pB31 clones; pL1For and NapinSeqRev primers for amplification of pL1 clones) are shown in Figure 2.3, 2.4, and 2.5. The rest of the samples can be found in the Appendix for Chapter 2. All PCR products were analyzed by gel electrophoresis, using 1.25% agarose gels. All primers used in this section are summarized in Table 2.1. Only plants having PCR products of bands of expected size were included in further analyses.

To verify the transgenic DNA sequences, PCR products were generated from each transgenic. Initially, a number of PCR products (n ≥3) were cloned into plasmid vectors and sequenced at PBI (Plant Biotechnology Institute, Saskatoon, Saskatchewan). Latterly, PCR products derived from some pB31 transformants and from all pL1 plant transformants, were sequenced directly at McGill University and Genome Quebec Innovation Centre according to their guidelines and requirements for sequencing.

2.3.6. Bioinformatics and sequence analysis

DNA sequences were analyzed in both DNAMAN Software (http://www.lynnon.com/) and Sequencher DNA Sequence Analysis Software. Both programs were used to confirm that the correct input sequences were cloned, to analyze restriction sites, and to examine chromatographs for DNA quality and potential incorrect base pairs.

To determine similarities between our novel plant insulators and other non-plant insulator sequences which have been tested to function in plants, Meme Suite 4.12.0 CentriMo was used. Our novel insulator sequences were entered into the program along with identified and characterized insulator protein DNA consensus sequences from non-plant species: Su(Hw), Zw5, BEAF-32, GAGA Factor, dCTCF, Rap1 (Table 2.2), and local motif alignments were achieved. CentriMo is a program which identifies the region of maximum central enrichment in a set of

ChIP-seq peak regions and displays the positional distributions of predicted sites (Bailey et al., 2009). Program parameters were kept at default values (match score threshold ≥ 5 and E-value ≤ 10) to avoid reducing statistical power or receiving less significant motif enrichments. Only most significant alignments are reported if the sequence contained multiple or overlapping alignments. Searches were also done using sequences that did not act as insulators in our assay to control for possible false positives.

Potential motifs and regulatory elements within the novel insulator sequences were analyzed in PlantCARE, New PLACE, JASPAR, and WEEDER databases. PlantCARE is a plant cis-acting regulatory element database which allows sequences in FASTA format to be submitted into the Search For Care option which presents an output containing potential binding motif, organism, location, consensus sequence and function (Lescot et al., 2002). The second database used, New PLACE, is a database of nucleotide sequence motifs found in plant cis-acting regulatory DNA elements. The novel insulator sequences were inputted into the databases submission box in FASTA format and an output of factor name, location, signal sequence, and New PLACE ID where motif description, PubMed ID and citations can be found. (Higo et al., 1999). Searches for potential transcription factor binding motif sequences within our potential insulator sequences were done in JASPER CORE (7th edition 2018) database online. JASPAR is an open-access database of curated, non-redundant transcription factor (TF) binding profiles across multiple species in six taxonomic groups, and include Plantae, Vertebrata, Insecta, Fungi, Nematoda, and Urochordata. Searches were done within the JASPAR Core full collection (total of 1564 profiles) with a 100% relative profile score threshold (Khan et al., 2017). The final online database, WEEDER2.0 was used to determine any novel binding motifs on the insulator sequences. WEEDER2.0 was no longer available online and therefore downloaded version was used. Sequences were entered into

the program and potential binding motifs (transcription factor binding sites) from co-regulated genes of *A. thaliana* were given.

Table 2.2: Insulator protein binding consensus sequences found in non-plant species. Shown are the insulator protein name, its consensus or core binding sequence on the insulator which the protein binds to, and the reference.

Insulator Protein Name	Consensus Sequence/core binding site	Reference
Su(Hw)	YRYTGCATAYYY	(Adryan et al., 2007; Dorsett, 1990; Spana et al., 1988)
Zw5	TCGCTGCGAATGACAAAACGGGCTGAGCA	(Gaszner et al., 1999)
BEAF	CGATA	(Zhao et al., 1995)
GAGA Factor	GAGAG , GAGAA	(Belozerov et al., 2003; Schweinsberg et al., 2004)
Rap1	ACACCCRYACAYM	(Graham and Chambers, 1994)
CTCF	CCCTC	(Bell et al., 1999)

2.4. Results

2.4.1. Statement of contributions

All pCAMBIA-derived plasmid vectors used in this study were previously constructed by Dr. L. Gudynaite-Savitch. The 154 bp random DNA oligonucleotide library for initial cloning into pC1 was synthesized at McMaster University. Thousands of seeds containing these sequences were screened resulting in 100 plants able to survive selection, a process initiated by Dr. Gudynaite-Savitch and Tatiana Semiz, who ultimately identified 60 candidates through PCR and sequencing analysis. Interestingly, the sequences that were cloned into pC1 produced both ~450 bp and the original 154 bp sequences. Mapping all of the EcoRI and BamHI restriction sites and subsequent sequencing of most inserts indicated that the ~450 bp inserts had been most likely generated by ligation of three different ~154 bp sequences joined through their EcoRI and BamHI sequences. See Batool Gandorah, 2012 for an in-depth explanation. These recovered inserts (putative insulators) were classified into four groups (I, II, III, and IV) based upon the selection trial completed. As resources were limited, an initial nineteen out of the sixty were chosen based on plant growth/health on selective media and cloning of PCR-generated inserts into pB31 was initiated by Dr. Gudynaite-Savitch. These nineteen sequences included: InI-2, InI-3, InII-3, InII-7, InII-10, InII-12, InIII-4, InIII-17, InIII-22, InIII-27, InIII-52, InIII-53, InIII-55, InIII-57, InIII-58, InIII-63, InIII-74, InIII-78 and InIII-80, collectively tested by Dr. Gudynaite-Savitch, Batool Gandorah[2] and Lara Rasooli[3]. In addition to these sequences, Dr. D.A. Johnson had cloned two

[2] https://ruor.uottawa.ca/bitstream/10393/23226/1/Gandorah_Batool_2012_thesis.pdf
[3] https://ruor.uottawa.ca/bitstream/10393/31329/1/Rasooli_Lara_2014_thesis.pdf

more sequences from the original 60 sequences (after pC1 selection) into pB31: InI-6 and InIII-50 both of which were strong candidates for insulator activity. The 4 most promising candidates were then cloned by Dr. D.A. Johnson into pL1 vector and included InI-3, InII-12, InIII-50, and a deletion of InIII-78 (InIII-78_5').

The work of this thesis included repeating the transformation and the testing in *A. thaliana* of InII-10, InII-12, and InIII-78 (all in pB31), and the transformation and testing in *A. thaliana* of InI-3, InII-12, InIII-50, and InIII-78_5' (all in pL1). Results for all 21 sequences, are included in the results section of this thesis (Table 2.3 A, B, C and Table 2.4).

2.4.2. Defining insulator function

The assay described in this thesis determined insulator function based on the ability of the tested sequence to block the interaction/influence of the CaMV 35S enhancer in the vector on the 35S46 core promoter (pC1 and pB31) or the napin promoter (pL1). As described in section 2.3.1, potential insulators initially survive negative selection in pC1 vector. The inserts were then re-cloned into pB31 and pL1 and transformed into *A. thaliana* to be tested and scored.

Growth in the presence of 5FU may arise for other reasons e.g. mutation of *codA* during transformation or the cloning of a site for binding a strong repressor, as well as the presence of an insulator. Therefore, potential insulator sequences were cloned into a second vector. Sequences tested in pB31 will show no staining when a functional insulator is present, as the sequence is able to block 35S enhancer influence on core 35S promoter. Staining in any of the tissue types (flowers, leaves, siliques) was interpreted as lack of insulator activity; however, the cloning of a site for binding a strong repressor would present the same pattern. Scoring involves assigning an intensity value for each tissue type for each sample from 0 to 3 (Figure 2.2 A) and adding up the number of

samples for each tissue type that stained. GUS expression is scored where any transgenic sample that contains no staining is added together as “none” and any sample containing staining in at least one of the tissue types as “stained”. For example, in the Appendix Chapter 2 Table A2.1 for construct InIII-52pB31, samples which contained staining for each tissue was totaled separately (flowers: 3 samples; leaves: 1 sample; silique test 1: 0 samples, and silique test 2: 0 samples), and the values for GUS expression were tallied up: 3/15 samples contained at least one stained tissue and designated ‘stained’, and 12/15 were identified as ‘none’, for no staining present in any tissue tested.

Sequences tested in pL1 are expected to produce either specific staining or non-specific staining. Samples demonstrating insulator activity will show specific staining, meaning the napin promoter will direct GUS expression specifically in seeds and no other tissues tested as the insulator will block the constitutive influence on 35S enhancer on the napin promoter. The absence of an insulator will produce staining in any of the other tissue types. Similar to pB31 samples, each tissue type for each sample will be assigned a staining intensity value. Samples containing staining in only seeds are added up as ‘specific’ GUS expression, and samples containing staining in at least one of the 3 tissue types (flowers, leaves, siliques) are totaled as ‘non-specific’ GUS expression. In addition to these two outcomes we have observed samples which contained neither specific or non-specific staining, which we have allocated as ‘none’. In the example of InI-3pL1 (Appendix Chapter 2 Table A2.2), samples which contained staining were totaled for each tissue type (flowers: 0 samples; leaves: 0 samples; siliques test 1:2 samples; seeds test 1: 15 samples; siliques test 2: 1 sample; seeds test 2: 18 samples). The total number of samples which contained specific GUS expression in seeds were summed (16/18), and the total number of samples which contained non-specific GUS expression, so staining in at least one of the tissue types, totalled

(2/18). And finally, 2 samples contained no staining in any tissue. This value was not added into the total number of samples. In the case of this latter result, we have postulated this to be the result of the activity of a repressor binding to the sequences (see Discussion).

Scoring was stringent - if only a small region of any tissues sampled contained GUS staining, it was scored as positive for staining. For example, see Figure 2.2 where samples Weak-1 and Medium-2 are scored as positive (stained) even though few of the siliques stained blue. All samples were scored independently by at least 2 individuals to reduce bias. At least 5 *A. thaliana* transformants were tested for each construct, unless samples clearly showed a lack of enhancer blocking function. All transgenic samples were analyzed by PCR with both SALK primers for DNA quality and vector specific primers for presence of insert, as well, PCR products were sent for PCR sequencing to confirm correct sequence.

2.4.3. Identification and analysis of potential insulator candidates

From the initial screening of oligonucleotide library, 60 potential candidates were obtained of which 21 were tested in pB31 (Table 2.3A, B, and C) From these, 9 sequences contained no insulator activity in pB31 (Table 2.3A) and 7 sequences showed promising insulator function in pB31 based on criteria tested (Table 2.3B). Sampling for additional sequences was incomplete; however, they are included in the thesis for completeness (Table 2.3C). Although the sample sizes are too small for confident assignment as potential insulators or not, the limited data suggests that InI-2, InIII-63 and InIII-74 are not promising candidates while InI-2 and InIII-80 may be worth further analysis.

Based on plant health, sample size obtained, and staining results, 4 of these sequences were cloned into pL1 and found to possess insulator function: InI-3, InII-12, InIII-50, and InIII-78_5'

(a deletion of III-78) (Table 2.4). Thus, of the 21 sequences tested, 4 are strong candidate insulators - a frequency of 19%. Due to time constraints and a lack of resources, InI-6, InII-10, and InIII-52 were not cloned into pL1. We can conclude from this limited sampling that all of the potential candidates identified in pB31 tested positively for insulator activity when cloned into pL1.

DNA samples were extracted from all transgenic plants with potential insulator inserts. For each sample, PCR amplification was carried out with SALKF and SALKR primers, and vector specific primers (GUS5'Rev and 1300LacZ-For for pB31 and pL1F and NapinSeqR for pL1) separately. Sample electrophoresis gels for various clones with SALK primers, pB31 vector primers, and pL1 vector primers, are shown in Figure 2.3, 2.4, and 2.5, respectively. For samples amplified with SALK primers, an expected band size of 762bp would indicate successful A. thaliana DNA extraction and amplification of the A. thaliana specific Atchrom-1 region. For samples in pB31 which were amplified with GUS5'Rev and 1300LacZ-For, an expected band size of ~600bp (cloning of short sequences approximately 150bp) or ~900bp (cloning of long sequences approximately 450bp) would indicate successful cloning into appropriate pB31 vector. The amplification of pB31 vector control alone produces a band size of 450bp. Samples in pL1 which were amplified with pL1F and NapinSeqRev primers, with an expected band size of 320bp (cloning of short sequences approximately 150bp) and 720bp (cloning of long sequences approximately 450bp), would indicate successful cloning into the pL1 vector. The amplification of the pL1 vector control produces a band size of 170bp. Only plants which produce the correct band size were included in the final analysis. A sample that did not produce a band was repeated to confirm its status.

Within the data obtained, we observed results which showed sequences that did not display 100% in one direction (100% staining or 100% no staining). For example, sequences which were

deemed insulators for a group of plants arising from transformants arising from a single clone, produced samples which were stained as well, and vice versa. In pB31, all 14 sequences which we have designated as having no insulator function did not show 100% stained samples, as we would have expected. Similarly, some sequences which we considered potential insulators, did not show 100% of samples not stained. Instead, InII-12pB31 and InIII-52pB31 (Table 2.3B) both showed some samples with stained tissue, although less than 30% of each sequence showed stained samples. Similar results were obtained from testing in pL1. Expected insulator function would display specific staining in all samples, however InI-3pL1 and II-12pL1 contained non-specific stained samples, and interestingly InIII-50pL1 and InIII-78_5'pL1 produced samples with no staining in either seeds nor other tissues (Table 2.4).

Our results are comparable to previously published data. Although other insulator papers vary in the specific assay system (different enhancers, promoters, etc.), the results produce similar outcomes in that they rarely yield 100% insulator function in plants (Gudynaite-Savitch et al., 2009; Hily et al., 2009; Singer et al., 2010) or other species even when testing insulators from the same species (Cai and Levine, 1995; Hagstrom et al., 1996).

Table 2.3: The β-glucoronidase (GUS) staining results for all 21 potential Arabidopsis thaliana insulator sequences tested in pB31 vector, including a positive control. The columns indicate the sequence number, construct name, size, number of total transformations, the number of individual transformants tested for flower, leaf and silique staining, and a calculation of the nonspecific expression of the GUS transgenes vs number of samples with no staining. Samples scored as stained, are a result of at least 1 of the 4 tissues which contained staining.

A. Results for sequences tested that possessed no insulator activity and control pB31 vector containing no insert.

					GUS staining			GUS expression	
Seq No.	**Construct Name**	**Size (bp)**	**No. trans-formations**	**No. plants tested**	**Flowers**	**Leaves**	**Siliques**	**Stained**	**None**
	pB31[‡]		2	62	11	39	38	46/62	16/62
1	InII-3*	151	1	17	9	6	9	13/17	4/17
2	InII-7*	150	2	30	11	14	8	19/30	11/30
3	InIII-4*	440	1	26	16	4	2	17/26	9/26
4	InIII-17*	443	1	36	25	15	2	26/30	4/30
5	InIII-22*	425	2	36	26	11	3	29/36	7/36
6	InIII-27*	128	1	13	13	4	2	13/13	0/13
7	InIII-53*	152	1	15	13	14	14	14/15	1/15
8	InIII-55*	432	2	23	22	5	2	22/23	1/23
9	InIII-57*	154	1	16	2	8	9	11/16	5/16

Results include data contributions of Batool Gandorah (*)

B. Results for sequences possessing insulator activity.

Seq No.	Construct Name	Size (bp)	No. trans-formations	No. plants tested	GUS staining			GUS expression	
					Flowers	Leaves	Siliques	Stained	None
10	InI-3‡	438	1	51	0	0	0	0/51	51/51
11	InI-6	438	1	24	0	0	0	0/24	24/24
12	InII-10	154	1	20	0	0	0	0/20	20/20
13	InII-12‡	427	2	37	0	0	3	3/37	34/37
14	InIII-50	443	1	11	0	0	0	0/11	11/12
15	InIII-52*	446	1	15	3	1	0	3/15	12/15
16	InIII-78*	433	1	20	0	0	0	0/20	20/20

Results include data contributions of Lara Rasooli (‡)

C. Results for sequences potentially possessing insulator activity (insufficient sample sizes).

Seq No.	Construct Name	Size (bp)	No. trans-formations	No. plants tested	GUS staining			GUS expression	
					Flowers	Leaves	Siliques	Stained	None
17	InI-2*	446	2	6	3	0	0	3/6	3/6
18	InIII-58*	150	1	3	0	0	0	0/3	3/3
19	InIII-63*	154	1	7	2	3	5	5/7	2/7
20	InIII-74*	153	1	6	3	4	4	4/6	2/6
21	InIII-80*	451	1	9	1	0	0	1/9	8/9

Results include data contributions of Batool Gandorah (*)

Table 2.4: The β-glucoronidase (GUS) staining results for four novel *Arabidopsis thaliana* insulator sequences in pL1 vector, including a positive control. The columns indicate the construct name and size, the number of individual transformants tested for leaf, flower, silique and seed staining, and a calculation of the specific and nonspecific expression of the GUS transgenes.

					GUS staining				GUS expression		
Seq No.	**Construct Name**	**Size (bp)**	**No. Transfor-mations**	**No. plants tested**	**Flowers**	**Leaves**	**Siliques**	**Seeds**	**Specific**	**Non-specific**	**None**
	pL1		4	53	41	49	42	41	2/51	49/51	2
1	InI-3	450	1	20	0	0	2	18	16/18	2/18	2
2	InII-12	439	1	27	0	1	0	22	22/23	1/23	4
3	InIII-50	455	1	12	0	0	0	12	12/12	0/12	0
4	InIII-78_5’	153	1	12	0	0	0	11	11/11	0/11	1

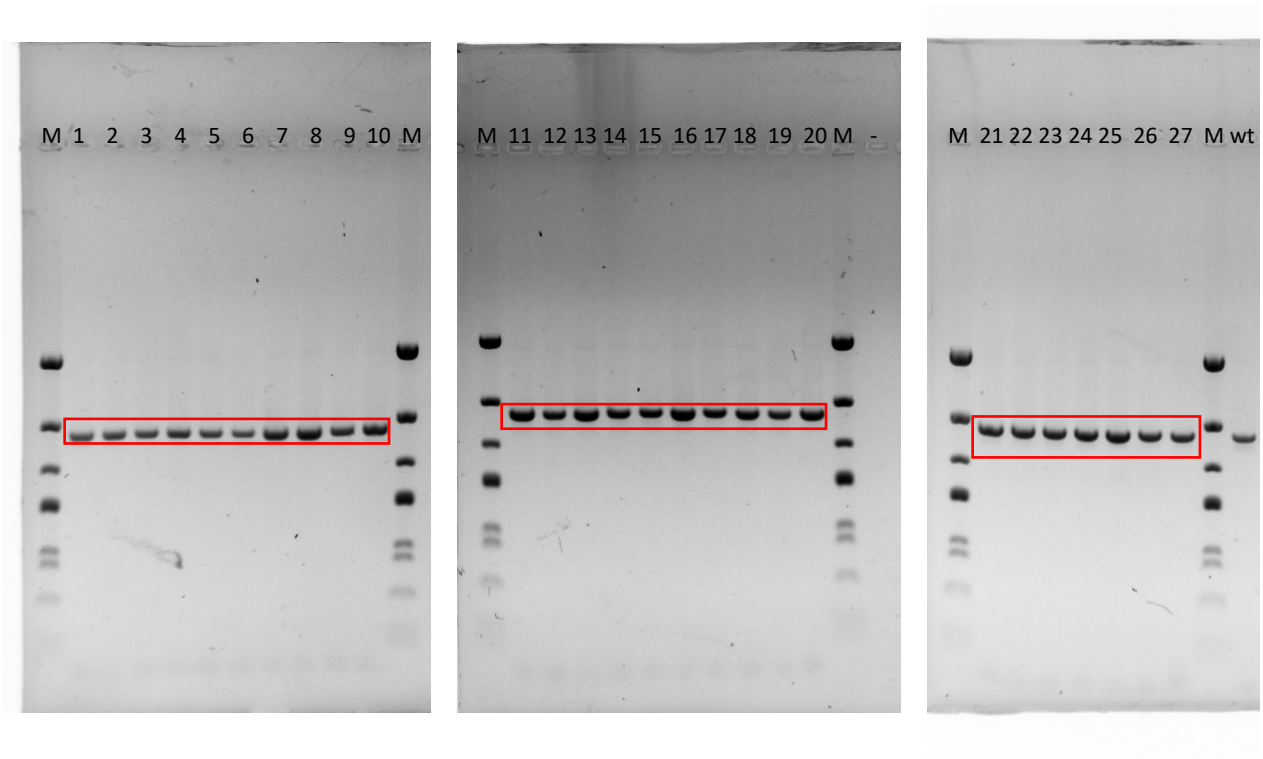


Figure 2.3: Example of agarose gel electrophoresis showing PCR amplification of clone InI-6 in pB31 with SALK primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *A. thaliana* DNA samples containing candidate InI-6 insulator are shown. Amplification of DNA used SALK_049131_RP2 and SALK_049131_LP2 primers to produce a band with an expected size of 762bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449,434), 298, 267, 174, 102, 80
1-27: Transgenic plant DNA containing InI-6pB31 inserts
wt: *A. thaliana* wild type DNA
- : Negative control containing water in place of DNA

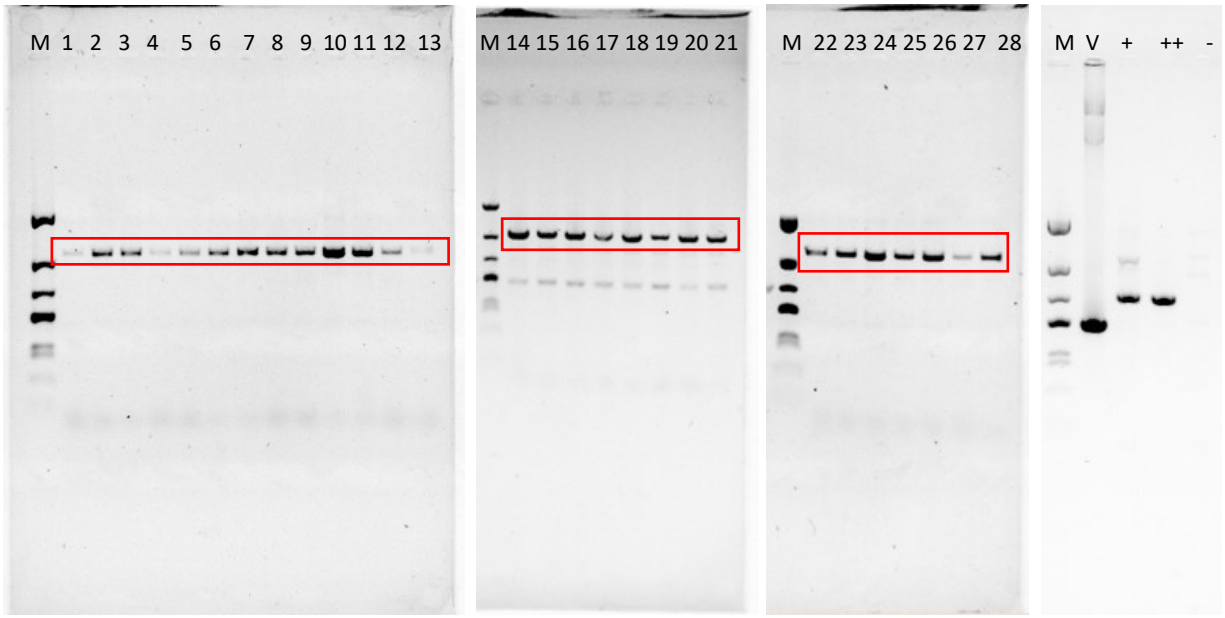


Figure 2.4: Example of agarose gel electrophoresis showing PCR amplification of InI-6 in pB31 with 1381F and GUS5'Rev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *A. thaliana* DNA samples containing candidate insulator InI-6pB31 are shown. Amplification of DNA used 1300LacZF and GUS5'Rev primers with an expected band size of 888bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-28: Transgenic plant DNA containing InI-6pB31 inserts (888bp)
V : pB31 control vector (no insert, 450bp)
+ : II-10 in pB31 sequenced control (604bp)
++ : I-3Δ in pB31 sequenced control (587bp)
- : Negative control containing water in place of DNA

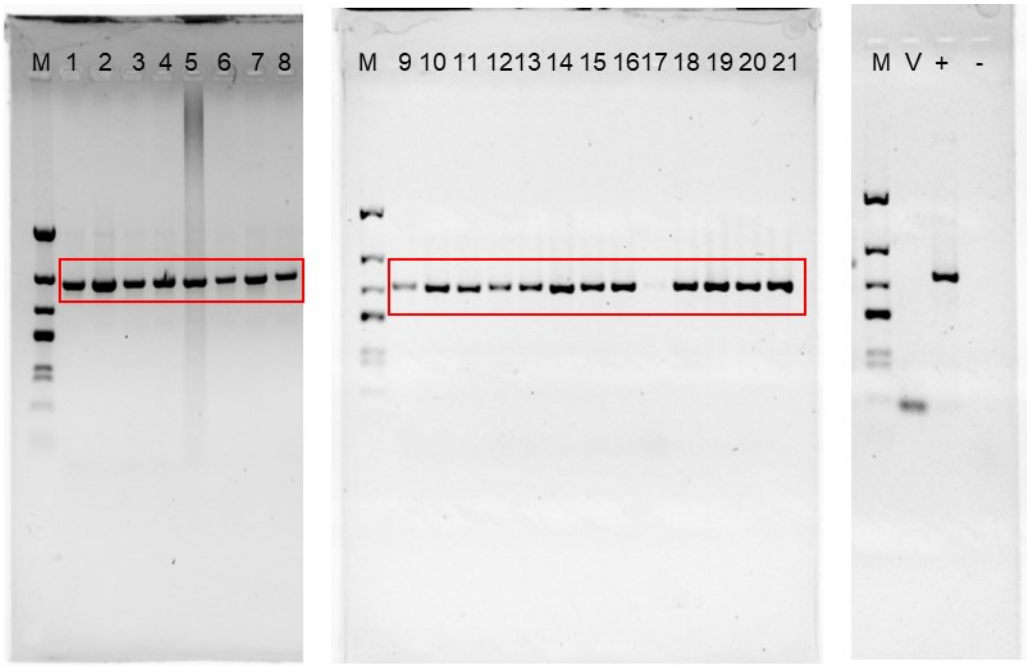


Figure 2.5: Example of agarose gel electrophoresis showing PCR amplification of clone InI-3 in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *A. thaliana* DNA samples containing candidate InI-3pL1 insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 620bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-21: Transgenic plant DNA containing InI-3pL1 inserts (620bp)
V : pL1 vector (no insert, 170bp)
+ : BEAD1c control (718bp)
- : Negative control containing water in place of DNA

2.4.4. Bioinformatics and sequence analysis of candidate insulator sequences

The analysis and interpretation of biological data through computational approaches, known as bioinformatics, is a rapidly emerging field. With increasingly large numbers of biological datasets organized into searchable online databases available, we have set out to identify important factors potentially interacting with our novel insulator sequences. The goal was to perform a preliminary search within multiple databases to identify known protein binding sites, transcription factors, repressors, or other regulatory elements, as well as novel motifs, that theoretically bind to our insulator sequences and may be required for their function. We have therefore used multiple databases to collect relevant data to infer potential binding sites.

2.4.4.1. Presence of functional insulator binding motifs in novel sequences

The presence of functional insulator protein binding motifs from non-plant sequences were searched in MEME Suite 4.12.0 CentriMo. CentriMo is a Local Motif Enrichment analysis program which identifies user-provided motifs that show a significant preference for particular locations in inputted sequences. CentriMo uses the binomial test to compute the significance of the number of sequences where the best match occurs in a given region. It then reports the location and significance of the best region on the forward and reverse sequence for each motif (Bailey et al., 2009). Novel insulator sequences were entered along with known protein binding consensus sequences from other non-plant species (Su(Hw), Zw5, BEAF-32, GAGA Factor, dCTCF, Rap1) (See Table 2.2). In addition, sequences which were observed not to have insulator activity (Table 2.3A), were also entered to act as a control. The results from this analysis provided us with multiple alignments to both our novel insulator sequences, and those which contained no insulator activity. Although alignments were found (See Appendix Figures A2.11-A2.18), there were no significant

results ($p>0.05$). Similar to previous theses from our lab (Batool Gandorah (2012) and Lara Rasooli (2014)), identification of the CCCTC (CTCF) site was found in insulator sequences, however they also found the same site in sequences with no insulator activity.

2.4.4.2. Identification of potential regulatory factors

In the search for potential plant regulatory binding elements, numerous amounts of potential sites were found on all 4 novel insulators (InI-3, InII-12, InIII-50 and InIII-78) through PlantCARE, New PLACE, JASPAR Core, and WEEDER2.0 motif discovery tool (See Appendix Tables A2.3-A2.18), and include, but not limited to, elements in enhancer and promoter regions, zinc finger factors, MYB transcription factors, and homeodomain factors. Determining these potential binding motifs will allow us to create associations between our novel insulators and possible gene regions, specifically promoters, as it has been suggested that there is a high similarity that exists between insulators and promoters (Reviewed by (Raab and Kamakaka, 2010)).

The JASPAR CORE 2018 database comprises of transcription factor binding profiles across multiple species, including plants and non-plants. The search within JASPAR Core determined that multiple amounts of non-plant binding motifs potentially exist on the 5 novel insulators from *Drosophila melanogaster, Saccharomyces cerevisiae* and *Homo sapiens.* The most recurring motifs among the 6 insulator sequences in this search were C_2H_2 zinc finger factor, C6 zinc cluster factor, and homeo domain factor.

2.4.4.3. Identification of repressor sites

We have explored the possibility of repressor sites present on our putative insulator sequences due to numerous samples showing repressor-like activity such as InI-3pL1, InII-12pL1 and InIII-

78_5'pL1. These sequences showed samples with no staining present, evidence of possible repressor binding sites within the sequence. The bioinformatics search found multiple possible repressor motifs (Appendix Table A19). The ERF-associated amphiphilic repressor, WBOXNTERF3, was found on InI-3, InII-12, and InIII-50. The repressor ARFAT was found on all 6 insulators, RAV1AT was found on InI-3, InIII-50, and III-78, MYB2AT was found on InI-3, MYB2CONSENSUSAT was found on InI-3, InIII-50 and InIII-78. The WRKY7OS was found in InI-3, InII-12, InIII-50, and InIII-78.

The bioinformatics analysis provided us with a huge amount of output data, some not significant such as the results of the CentriMo search. It is clear that the Bioinformatics search is a preliminary study and further experiments for protein analysis must be done to determine true binding sites on our novel insulators.

2.5. Discussion

The discovery of novel plant insulators is critical to understanding the underlying mechanisms and models of insulation in plants. Their ability to block unwanted gene expression and act as chromatin barrier elements plays an important role in gene regulation. To date, few plant insulators have been identified, and therefore our goal was to discover and study novel insulator sequences to broaden our understanding of these regulatory elements and their possible conservation across species. As a first step we have devised an experimental approach, the first of its kind, for the selection of novel insulators functional in all plants. Our approach is based upon the theory that even random sequences may possess insulator activity if they contain the correct consensus sequences for insulator binding proteins. To determine whether random sequences are potential insulators, we devised a system in which candidate sequences from a random oligonucleotide

library were screened sequentially through 3 vectors to test their ability to block enhancer-promoter interactions: pC1, pB31, and pL1. Ultimately, upon successful testing and development of this experimental system we can apply such assay to natural plant insulator DNA.

2.5.1. Vector rationale

In the initial stages of assay development, multiple factors were considered. An identified insulator would need to significantly block enhancer-promoter interactions implying a strong DNA-protein interaction, possess the capacity to function in multiple tissues and retain function through one round of negative selection and two rounds of subcloning followed by screening. In addition, the cloned DNA should not contain additional sequences that may interfere with insulator identification or function such as a repressor. Insulators are classified based on GUS expression in pB31 and pL1 vectors. In pB31, sequences which are insulators will block the influence of the 35S enhancer on the 35S46 core promoter and no staining will be present in tissues. Sequences in pL1 with insulator function will block the influence of the 35S enhancer on the napin promoter so that staining is directed only in seeds (specific staining).

The first vector, pC1, is a negative selection vector which contains the negative selective marker gene cytosine deaminase (*codA*). *codA* converts the harmless 5-Fluorocytosine (5-FC) present in selective agar plates into the highly toxic metabolite 5-Fluorouracil (5-FU) and eventually 5-F-dUMP which irreversibly inhibits thymidylate synthase (TS) blocking DNA synthesis and ultimately leading to cell death. The *codA* gene is expressed by CaMV35S core promoter and CaMV35S enhancer. The presence of a functional insulator between the enhancer and core promoter will block the influence on *codA* and plants will survive selection. We have chosen the *codA* selection system for our assay because of the high sensitivity level and obvious

effects of *codA* in *A. thaliana. A. thaliana* plants that do not express *codA* are completely insensitive to 5FC while those that express *codA* are quite sensitive to toxic 5-FC (Perera et al., 1993). *codA* selection is also effective in other plants including lotus (Stougaard, 1993), tobacco (Schlaman and Hooykaas, 1997), rice (Dai et al., 2001), and soybean (Shao et al., 2015), implying that this marker is likely suitable for negative selection in many plant species, as no homologs of *codA* have yet been found in plants (Shao et al., 2015). This system is also superior to other negative selection systems, such as the negative markers, *iaaH* (indoleacetamide hydrolase) which produces significant background of non-specific hydrolase activity in most plants interfering with selection activity (Perera et al., 1993), and NR (nitrate reductase) which is sensitive to growth conditions (Cheng et al., 1991).

The negative selection system may allow growth for other reasons as well. During transformation or plant growth, mutations that inactivate c*odA* could give a phenotype that would mimic the function of an insulator. The possible insertion of the transgene into a chromosomal region that allows selection on hygromycin but not the expression of *codA* would also mimic insulator function as could the cloning of a site for binding a strong repressor. Therefore, further cloning into additional vectors was necessary.

pB31 is the screening vector which contains a β-glucuronidase (GUS) reporter gene, in place of *codA* gene, expressed by CaMV35S enhancer and the 35S46 core promoter. The presence of an insulator between these two elements will block interactions, resulting in no blue staining in tested plant tissues. This vector allows for the simultaneous screening of insulator function in flowers, leaves, siliques, seeds and stems.

The final pL1 screening vector contains a GUS reporter gene controlled by the napin seed specific promoter and the constitutive CaMV35S enhancer. The presence of an insulator between these two regulatory elements will block the influence of the CaMV35S enhancer and staining will be directed to seeds only. This vector allows us to assess the likelihood of a strong repressor competing with the enhancer. As we saw with multiple sequences cloned into pL1, some sequences produced samples with neither specific staining or non-specific staining. This is a result to expect when a repressor is present as the repressor reduces the level of transcription from both the napin promoter and the 35S46 core promoter (Gaszner and Felsenfeld, 2006).

2.5.2. Confirmation of insulator activity in potential insulator sequences

Sequences tested in the pB31 vector were inserted between a CaMV35S enhancer and 35S46 core promoter. The expression of GUS in this system (blue staining in tissues) would signify the absence of an insulator as there would be influence of the enhancer on the promoter. The presence of an insulator would result in zero staining in tissues, as the sequence would block the interactions between the enhancer and promoter. Our findings indicate that 7 sequences substantially blocked the interactions between the CaMV35S enhancer and 35S46 Core promoter in pB31and include InI-3, InI-6, II-10, InII-12, InIII-50, III-52 and InIII-78 (Table 2.3B). Within these results, sequences InII-12 and III-52 showed some staining (3/37 and 3/15 samples respectively). Although only ~8% of samples contained staining for II-12 and ~20% for III-52, it is important to note that typical results in literature produce similar outcomes in that rarely do experiments produce the expected 100% insulator activity (Singer et al*.,* 2011; Cai et al*.,* 2001; Muravyova *et al.,* 2001). The differences in staining that result from the introduction of one sequence can be attributed to the fact that *Agrobacterium*-mediated transformation causes T-DNA insertion to be integrated randomly in any chromosome or locus in plants (Forsbach et al., 2003; Kim et al., 2007), and

therefore the exact same DNA sequence can produce varying effects depending on where in the genome it is inserted. Further investigations could be done to determine whether differences in staining is due to strength of insulators, insertion sites, or influenced by various test systems. For example, identification of T-DNA insertion sites for samples tested could determine whether samples are inserted in similar locations of the genome (Huang et al., 2007; Majumder and Cai, 2003). In future studies, the usage of the CRISPR-Cas9 system would allow us to modify the *A. thaliana* gene to integrate our insulator construct in the same precise locations to test for insulator function.

The ability of sequences InI-3, InII-12, InIII-50 and InIII-78 to block enhancer influence in our pB31 system led us to clone these sequences into vector, pL1 (Table 2.4). The capacity of the sequences to produce insulator function is marked by staining strictly in seeds driven by the napin seed specific promoter. Values determined for sequences in the pL1 vector showed similar activity percentages. InIII-50 and III-78_5' produced 100% specific staining, a strong indication of insulator function. InI-3 and II-12 however showed some non-specific staining in siliques and leaves respectively. This result is not unexpected. As noted above, most published insulator assay results have not achieved 100% function even with varying vectors and vector elements such as differing promoters (Gudynaite-Savitch et al., 2009; Hily et al., 2009; Singer et al., 2010), possibly due to the many factors not fully understood about the interactions between primary and secondary insulator binding proteins. Characterizing these sequences, by creating deletions, could potentially identify sequences within the cloned DNA necessary for insulator function.

2.5.3. Deletion analysis of InIII-78

In III-78 contains a 433bp DNA that functions exceptionally well as an insulator in the pB31 vector (100% enhancer blocker activity). During cloning of InIII-78 into pL1, a deletion occurred that allowed the recovery of a 153bp fragment with insulator activity. InIII-78_5' has significant insulator activity, with 11/11 samples possessing specific GUS expression in seeds, and 1 sample showing no staining. Although 9 more samples would need to be completed for a full result comparable to other sequences, 100% insulator activity in samples currently tested is a strong indication of insulator action. The identification of insulator activity in InIII-78_5' indicates that relatively small random sequences may be identified by our assay system. Deletion analysis is a technique frequently employed in insulator studies to identify and localize regions of activity (Hagstrom *et al.,* 1998; Singer *et al.,* 2011).

2.5.4. Insulators and specificity for tissues and other elements

The sequences we have discovered to be novel insulators were tested in flowers, leaves, siliques, and seeds specifically. However, the ideal "universal" plant insulator would be one able to function in all plant tissues in all higher plants. Plant insulator studies have used various assays to test for insulator function, utilizing different promoters which target different plant tissue and testing several types of tissues. Compared to most studies, our assay tests the most tissue types for enhancer-blocking insulator function. Both Hily *et al.* (2009) and Singer *et al.* (2010) used the flower-specific *AGAMOUS* second intron-derived promoter (*AGIP*) to test insulator function in transgenic *A. thaliana* and analyzed only leaf and floral tissue. Correspondingly, Singer *et al.* (2011) tested the *Arabidopsis thaliana* gypsy-like (*At*gypsy-like) sequence within leaf and floral tissues only, although instead used the petal- and stamen-specific *PISTILLATA* promoter (*PIp).* It

is important to test insulators in multiple tissue types as the response in different cell-types in each plant tissue could vary, creating a different effect. Studies have not shown that insulators are cell-type specific, however the likelihood of insulators being functional in one tissue and not another of the same transgenic plant or sample is a possibility that needs to be considered when testing insulator function. Our system attempts to test sequences in multiple tissues (roots, flowers, leaves, siliques, seeds) to determine whether sequences are tissue specific. The novel insulators we have identified through our assay does not seem to be tissue specific, an advantage for the production of transgenic plants.

Although previous studies have used different vector systems than ours (pC1/pB31/pL1 system) they typically tested fewer tissue types (Singer *et al.,* 2011). We have used deletion analysis to localize key sequences required for insulator function in plants for candidate sequences in other species described in the next chapter.

2.5.5. Bioinformatics investigation and implications

The field of bioinformatics sets out to answer fundamental biological questions through the application and usage of computational approaches. It is increasingly popular for biological studies to utilize software and online databases to search for potential patterns or functional elements present within their experimental analyses. We have chosen to use this method as well to study our novel insulator sequences and search for potential binding sites for insulator protein associations, transcription factors, repressors and other regulatory elements which may be important to the function of the sequences. Our search involved discovering known and novel regulatory element motifs from both plant and non-plant species. The databases used for this thesis contained published and experimentally defined data from multiple sources. The significance in

using multiple sources allowed us to cover a wide range of binding sites found by multiple search algorithms, as each site discovered binding sites using different models.

From our CentriMo analysis, multiple known protein binding sites were computationally determined to be found in the novel insulator sequences. This is important to our study because it suggests a link between insulator function in plants and other species. However, it is clear through results from our analyses that further studies must be done in addition to these database search to find true insulator binding proteins. The results we have obtained can not directly describe any protein binding sites on our sequences as the program showed no significant protein binding sites for the consensus sequences we had entered into the program. The analysis we have done here is a preliminary search into potentially narrowing experimental procedures and aiding in experimental designs for efficiency.

Staining results in the pL1 vector for the novel insulators produced samples that showed neither specific nor non-specific staining. In sequences InI-3, InII-12 and InIII-78_5', few samples showed zero staining in any tissue. This result suggests a possibility of repressor action. The bioinformatics search produced promising results of this possibility. 6 repressors were found to contain binding sites on the novel insulators, WBOXNTERF3, ARFAT, RAV1AAT, MYB2AT, MYB2CONSENSUSAT, and WRKY71OS. If these sequences are in fact repressors binding to our sequence, a possible model which describes why some of our samples show no staining could be a strong repressor interference.

The abundance in data that is given from bioinformatics analyses may pose problems in drawing conclusions, as many results contain false positives. The results produced from our search is an initial step to studying the novel insulator sequences we have discovered.

2.6. Conclusions

The rigorous 3-step system we have developed in this chapter selects for novel insulators functional in *A. thaliana*. The system tests sequences in 3 vectors, each containing varying enhancers, promoters, and reporter genes, to test the potential insulator sequences for insulator function defined in literature. Through our system, and through the usage of an oligonucleotide library, we have discovered 4 different *A. thaliana* insulators we believe will function in all plants: InI-3, InII-12, InIII-50, and InIII-78_5'. We have found that the length necessary for function varies between long sequences (~450bp) and short sequences (~150bp). We know that the DNA sequences themselves are not the only factor involved in insulator function, as the protein binding sites are a necessary component for function, as described by the models we discussed in the Introduction. The next step in characterizing these novel sequences would be to test for protein binding sites. Our preliminary bioinformatics analysis identified multiple protein binding sites from *D. melanogaster*, supporting our hypothesis of insulator conservation across species. This initial analysis requires an experimental assessment on the sequences and protein binding sites to determine whether actual proteins do in fact bind the sequences for insulator function in plants. In future studies this can be performed by DNAse footprinting, electrophoretic mobility shift experiments or more recent tests such as ChIP-Chip and ChIP_Seq, as previously listed.

The process we have proposed for the selection of insulator sequences can be improved. In subsequent studies the assay could involve selection/screening in pC1 and pL1 exclusively. For example, new techniques such as In-Fusion Cloning removes the problem of specific restriction sites that were found within the random oligomer, facilitating direct cloning into pL1 reducing time, resources, and minimizing data while producing the same results. We believe this assay will

be useful in other plant species as well, as the use of the napin promoter has been shown to confer seed specific expression pattern in heterologous systems (Chen and Lin, 2012; Devi et al., 2010).

This is the first assay to carry out this purpose in plants, and with a lack of plant insulators this experimental system will increase our fundamental understanding of insulator models and mechanisms. Insulator discovery and characterization in plants produces a powerful tool for transgenic plant biotechnology research and ultimately agriculture and agri-food applications, and we hope that creating this assay will allow future insulator research to become more efficient. The ability of insulators to prevent interactions between genetic elements of different genes, and to minimize or potentially prevent position effects in transgenic constructs, makes these sequences ideal in research and industrial applications involving expression of multiple gene cassettes. The study of insulators is therefore vital in the overall advancement of plant biology research.

3.0 Detailed characterization of defined non-plant insulator sequences in *Arabidopsis thaliana*

3.1. Introduction and previous work

The control of developmentally regulated genes and the partitioning of active and inactive domains in the genome is an important part in gene regulation controlled mainly by chromatin insulators and their associated protein complexes. In addition to these natural functions, the onset of misexpression in introduced transgenes due to enhancer-mediated activation or crosstalk with adjacent promoters, poses an issue in the ever-growing need for improvement of agronomic crops. In these scenarios, insulators have been shown to be ideal candidates in minimizing the unwanted interactions between regulatory elements within T-DNA borders, as well as between multiple transcriptional gene units of transgenic plants. The function of insulators is dependent on specific proteins binding to the sequence, which in turn bind nuclear proteins associated with the nuclear lamina. The primary insulator proteins have been identified in multiple species; for example in *D. melanogaster* there exists Su(Hw) (Golovnin et al., 2003; Parkhurst et al., 1988), ZW5 (Gaszner et al., 1999), GAGA Factor (Belozerov et al., 2003; Schweinsberg et al., 2004), BEAF (Sultana et al., 2011; Zhao et al., 1995), and dCTCF (Holohan et al., 2007; Moon et al., 2005). In yeast, there exists the insulator protein Rap1 (Huet and Sentenac, 1987; Steiner and Philippsen, 1994), and in vertebrates there exists the insulator protein CTCF. These proteins bind to consensus sequences on insulators, with variants among species. For example, the CTCF protein binds a consensus site in multiple vertebrates (summarized in Tables 1.1 in Chapter 1 and 2.2 in Chapter 2). This suggests that plant proteins may also bind to a consensus site on the DNA insulators we introduce.

Recent experiments have demonstrated that cloned DNAs from non-plant sources can function as insulators in the model plant *Arabidopsis thaliana* (Gudynaite-Savitch et al., 2009; Hily et al., 2009; Jiang et al., 2017; She et al., 2010a; Singer et al., 2011). Still, there remains a lack of literature about how insulators function in plants despite the importance they play. In particular, those experiments do not define the precise DNA sequences within the cloned DNA that are responsible for the measured activity. This chapter will investigate non-plant insulators through an analysis of *A. thaliana* transgenic lines with the goal of defining these DNA sequences and potentially identifying different insulator pathways in a plant. We use genetic analyses to investigate three cloned non-plant DNA fragments that have insulator activity: the fungal *Ashbya gossypii* sequence UASrpg (Bi and Broach, 1999); the human T-cell insulator BEAD1c (Zhong and Krangel, 1997); and the *D. melanogaster* gypsy insulator element (Cai and Levine, 1995; Geyer et al., 1986). Our assay involves the insertion of potential insulator fragments into the vector pL1 between a CaMV35S enhancer and seed specific napin promoter to test for enhancer-blocking activity as described in Chapter 2, Section 2.31 of this thesis.

The 250bp fragment UASrpg (upstream activation site for ribosomal protein genes) from *Ashbya gossypii* was initially shown to have insulator activity in yeast, thus signifying that a sequence from one species can function in a different one. Originally identified as part of the promoter region within *A. gossypii*'s TEF (Translation Elongation Factor) gene, when cloned into yeast it demonstrated resistance to transcriptional silencing by *HM* silencers possessing the capability of blocking the spread of heterochromatin-like structures (Bi and Broach, 1999). In gel retardation experiments with *A. gossypii* protein extracts, Steiner and Philippsen (1994) demonstrated specific protein binding of Rap1 to a homologous Rap1 consensus sequence in yeast – ACACCCRYACAYM (See Chapter 2, Table 2.2) (Graham and Chambers, 1994; Graham et al.,

1999). The two sequences: GCCCATACAT (R1) and ATCCATACAT (R2) are essential for high-level expression (Steiner and Philippsen, 1994). Mutation and deletion analysis of Rap1 protein sequences of yeast removes or reduces insulator activity (Bi and Broach, 1999; Yu et al., 2003). UASrpg has been previously tested in *A. thaliana* and found to significantly reduce non-specific interactions when placed between a CaMV35S enhancer and napin promoter (Gudynaite-Savitch et al., 2009). In an attempt to understand the function of UASrpg as an insulator in *A. thaliana*, we have chosen to recreate the mutations in the Rap1 binding sites identified by Steiner and Philippsen (1994) and Bi and Braoch (1999) that are responsible for silencer-blocking activity in yeast and to test them for activity in *A. thaliana*. If these sites are important we could then postulate that Rap1-like proteins are functional in a plant.

BEAD1c (550bp), the second sequence under examination is the core sequence of the 1.6kb BEAD1 (blocking element alpha/delta-1) element of the human T-cell receptor (TCR) α/ δ locus (Zhong and Krangel, 1997). BEAD1 is found adjacent to the 3' end of the TCR δ gene segments and adjacent to the 5' end of the TCR α Joining gene segments. The α and δ gene segments are organized within a single genetic locus, with δ gene segments nested between α gene segments, but are differentially regulated during T-cell development. Due to the close apposition of BEAD1 between the gene segments, and the predicted function of the element to prevent cis-acting effects of enhancers/promoters of one gene segment on the enhancers/promoters of the other, the BEAD1 sequence was initially tested and ultimately identified as an insulator via an enhancer-blocking chromatin-integrated construct containing an E_δ (TCR δ Enhancer) and P_δ (multiple variable δ Promoter) driving neomycin resistance gene expression in human Jurkat cells (Zhong and Krangel, 1997). Both the full length BEAD1 and its core sequence, BEAD1c, have been tested in *A. thaliana* and proved to be successful at reducing nonspecific interaction by blocking CaMV35S enhancer-

napin promoter interactions. (Gudynaite-Savitch et al., 2009), making it the obvious candidate to characterize. Our deletions analysis on BEAD1c examined the importance of a specific sequence within BEAD1c for insulator function in *A. thaliana.* This sequence, coined BEADA, binds specifically to purified chicken CTCF (CCCTC-binding factor) (Bell et al., 1999), the highly conserved insulator protein found in vertebrates which contains multiple consensus sequence variants in Eukaryotic species (Filippova et al., 1996). Deletion analysis found that BEADA alone was an effective enhancer-blocking element when positioned between a human γ-globin promoter and an upstream enhancer/locus control region (LCR) element from the mouse β-globin locus (Bell et al., 1999). In a similar fashion, we have experimentally tested deletions of BEAD1c to characterize fragments within the sequence, with and without the BEADA binding domain to investigate whether it is essential for insulator function in *A. thaliana.*

The highly characterized gypsy insulator was initially discovered as a large 7.3kb retrotransposon involved in multiple spontaneous mutations causing changes in the bithorax complex of *D. melanogaster.* (Bender et al., 1983). The element was further examined by an assay testing the insertion of gypsy in the *yellow* gene of *D. melanogaster,* whose expression is controlled by a series of tissue-specific transcriptional enhancers located in the region 5' to the gene. The expression of mutant phenotype, alteration of yellow pigment, was observed to be caused by the transcriptional blocking of gypsy on enhancers located more distal from gypsy, but produces no effects on enhancers more proximal to the promoters (Geyer et al., 1986). The mutant phenotypes were also reproduced with the insertion of a smaller 340bp fragment containing the vital Su(Hw) binding regions in a yellow gene construct, suggesting that the Su(Hw) binding region is sufficient to confer the mutant phenotype. The Su(Hw) protein is involved in the regulation of gypsy through its interaction with 12 sites on the element - twelve copies of the consensus sequence:

YRYTGCATAYYY (see Chapter 2, Table 2.2) (Adryan et al., 2007; Parkhurst et al., 1988; Spana et al., 1988). The interaction was later found to be directly responsible for the effect of gypsy, as deletions or other alterations in the Su(Hw) binding sites result in a decrease or abolishment of the mutagenic consequence of gypsy (Geyer et al., 1988; Smith and Corces, 1992). The insulating effects of this 340bp fragment became one of the most characterized gypsy sequences tested in various other species (Cai and Shen, 2001; Nagaya et al., 2001; She et al., 2010a; Singer and Cox, 2013; Singer et al., 2010). Several studies have shown that the element functions to protect against position effects in *A. thaliana* (Jiang et al., 2017; She et al., 2010b); furthermore a gypsy-like sequence found in the *A. thaliana* genome displays enhancer blocking effects when inserted between a CaMV35S enhancer and Pip (petal-and stamen-specific PISTILLATA promoter) (Singer and Cox, 2013). We have tested this 340bp gypsy insulator in our assay, to determine whether this sequence is able to block enhancer-promoter interactions between a CaMV35S enhancer and seed specific napin promoter. We have also created deletions of the twelve Su(Hw) protein binding sites to analyze whether *A. thaliana* uses a similar site for insulator function.

In addition to deletion and mutation analysis, insulator sequences under examination have also been tested for orientation dependence by cloning an inverted version of the fragment and testing for enhancer-blocking activity. This approach may provide a better understanding on the mechanism of insulator function in plants.

Through these analyses, we hope to identify regions within insulators that are crucial to plant insulator function, potentially determining sequences and mechanisms of conservation among species.

3.2. Hypothesis and Objectives

The hypothesis for this chapter is that insulators are conserved across species. This implies that the modes or mechanisms of insulator activity are conserved as well as the insulator protein binding sites. In order to test this hypothesis our experimental approach was to analyse mutations of previously identified insulator sequences, UASrpg from *Ashbya gossypii*, BEAD1c from *Homo sapiens* T-cells, and gypsy from *Drosophila melanogaster* using a transgenic *Arabidopsis thaliana* system. These sequences may potentially represent three different insulator pathways defined by UASrgp-Rap1, BEAD1c-CTCF and gypsy-Su(Hw) interactions in their original hosts.

3.3. Materials and Methods

3.3.1. Molecular techniques, plant material and growth conditions

BEAD1c_5' end, BEAD1c_3'end deletions were created from digestion of pL27 with two restriction enzymes followed by repair and ligation. All other sequences were synthesized by Bio Basic Inc.[4] subcloned into the pL1 vector (Gudynaite-Savitch et al., 2009; Figure 2.1C), and transformed into *A. thaliana* to be tested for GUS expression and insulator function. Molecular techniques, plant growth conditions and pL1 vector elements are described in Chapter 2, Section 2.3. Table 2.1 provides a description of each sequence synthesized, subcloned into pL1, and tested for insulator function. Sequences originating from UASrpg containing mutations for protein binding site as identified and analysed in yeast (Yu et al., 2003) in which 2 cytosine nucleotides were replaced with 2 adenosines. Primers for PCR amplification can be found in table 2.1, and

[4] https://www.biobasic.com/

primers used for In-Fusion cloning of ΔBEADA and UAS_3' ΔCTC can be found in Chapter 2, Table 2.1. To create inversions, PCR was performed to introduce restriction sites (HindIII and PstI) in the inverse orientation to the initial fragment.

The assay used in this chapter, analogous to Chapter 2, tests sequences on their ability to function as enhancer-blockers in the pL1 vector, specifically. Vector components were previously described in Chapter 2, Section 2.3.1 and Figure 2.1C. Sequences of previously defined insulators (BEAD1c, UASrpg and gypsy) and the generated variants were individually inserted between the CaMV 35S enhancer and the seed specific napin promoter driving a GUS reporter gene, and transgenic plant tissue samples (flowers, leaves, siliques, and seeds) were analyzed for GUS staining. Sequences tested in pL1 will produce either specific staining or non-specific staining. Samples demonstrating insulator activity will show specific staining, meaning the napin promoter will direct GUS expression specifically in seeds and no other tissues tested as the insulator will block the constitutive influence of 35S enhancer on the napin promoter. The absence of a functional insulator will produce staining in other tissues as well. In samples which contain neither specific nor non-specific staining (no staining in any tissue), we have recorded them as "none". Scoring was stringent – if even a small region of a single tissue (out of 4 tissue samples) contained GUS staining, it was recorded as positive for staining. All samples were scored independently by at least 2 individuals to reduce bias. The full scoring strategy can be found in Chapter 2, section 2.4.2. At least 5 *Arabidopsis* transformants were tested for each construct, unless samples clearly showed a lack of enhancer blocking function. All transgenic samples were analyzed by PCR with both SALK primers for DNA quality and vector specific primers for presence of insert, as well, 5 PCR products were sent for PCR sequencing to confirm correct sequence. Similarly, to the prior chapter, all DNA samples were extracted from transgenic plants with permutated insulator inserts.

For each sample, PCR amplification was carried out with SALKF and SALKR primers, and pL1 vector specific primers: pL1F and NapinSeqR. Sample electrophoresis gels for various clones with SALK primers and pL1 vector primers are shown in Figure 3.4 and Figure 3.5, respectively.

Table 3.1: Summary of all sequences sub-cloned into pL1 and tested for GUS expression for insulator function.

Sequence Name	Size (bp)	Description*	Sequence Organism
UAS_5'end	149	5' end of UASrpg	*Ashbya gossypii*
UAS_mR1	149	Dinucleotide mutation (CC → AA) on 1st Rap1 site (R1)	*Ashbya gossypii*
UAS_mR2	149	Dinucleotide mutation (CC → AA) on 2nd Rap1 site (R2)	*Ashbya gossypii*
UAS_5'INV	149	Reverse complement of UAS_5'end	*Ashbya gossypii*
UAS_3'end	109	3' end of UASrpg	*Ashbya gossypii*
UAS_3'ΔSu(Hw)	100	UAS_3'end sequence with deletion of potential Su(Hw) consensus sequence	*Ashbya gossypii*
UAS_3'ΔCTC	104	UAS_3'end sequence with deletion of potential CTCF consensus sequence	*Ashbya gossypii*
UAS_3'INV	109	Reverse complement of UAS_3'end	*Ashbya gossypii*
UASrpgINV	246	Reverse complement of UASrpg	*Ashbya gossypii*
BEAD1c_5'end	265	5' end of BEAD1c	*Homo sapiens*
BEAD1c_3'end	301	3' end of BEAD1c	*Homo sapiens*
BEAD1cINV	548	Reverse complement of BEAD1c	*Homo sapiens*
ΔBEADA	509	BEAD1c sequence with a deletion of the core binding protein sequence	*Homo sapiens*
gypsy	340	Original gypsy sequence	*Drosophila melanogaster*
gypsy_ΔSu(Hw)	286	gypsy sequence with deletion of 12 Su(Hw) sites	*Drosophila melanogaster*

*Specific mutations are described in the text. Maps indicating the position of the mutations and the deleted segments are found in Figure 3.1, 3.2, and 3.3.

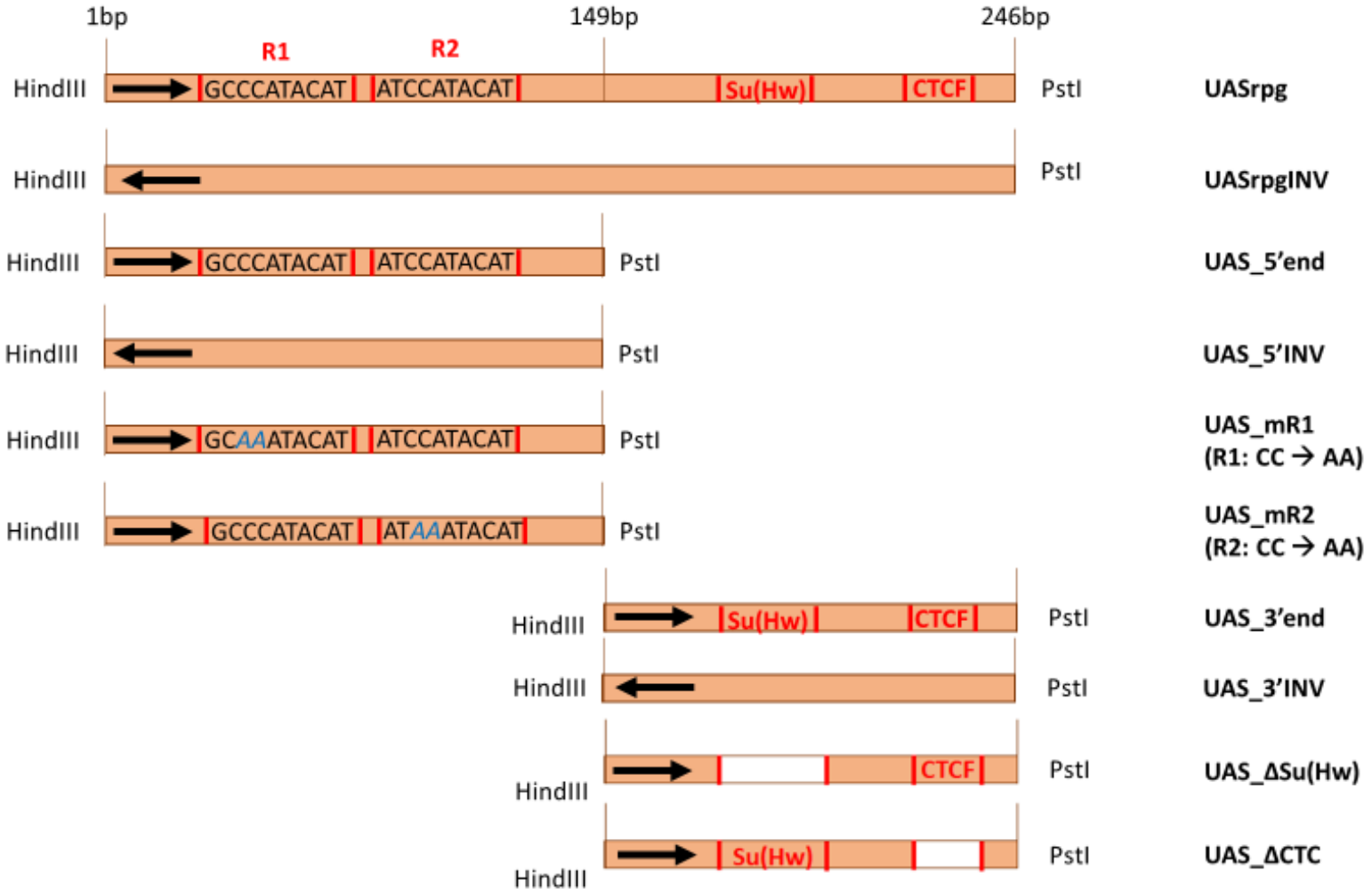


Figure 3.1 Clone map and names of all UASrpg sequence permutations. Orange rectangles represents the insulator sequence bordered by restriction sites. Length (in bp) of the full sequence is indicated at the top. Sequence orientation is represented by black arrows. Potential protein binding sequences are indicated between red bars. Deletions within the sequence is represented by white rectangles between red bars, and nucleotide mutations are shown in blue italicized text. Not drawn to scale.

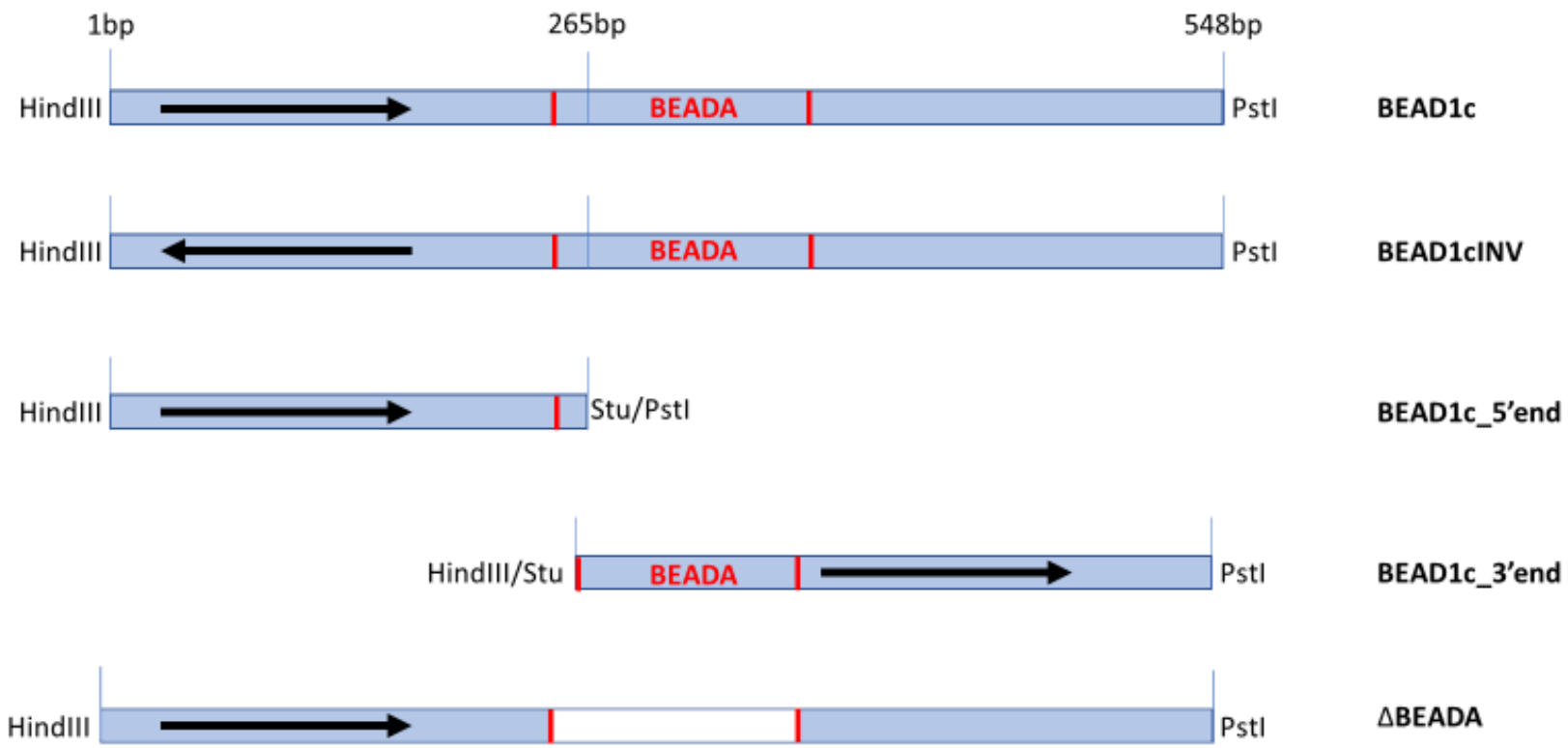


Figure 3.2 Clone map and names of all BEAD1c sequence permutations. Blue rectangles represent the insulator sequence bordered by restriction sites. Length (in bp) of the full sequence is indicated at the top. Sequence orientation is represented by black arrows. Potential protein binding sequences are indicated between red bars. Deletions within the sequence is represented by white rectangles between red bars. Not drawn to scale.

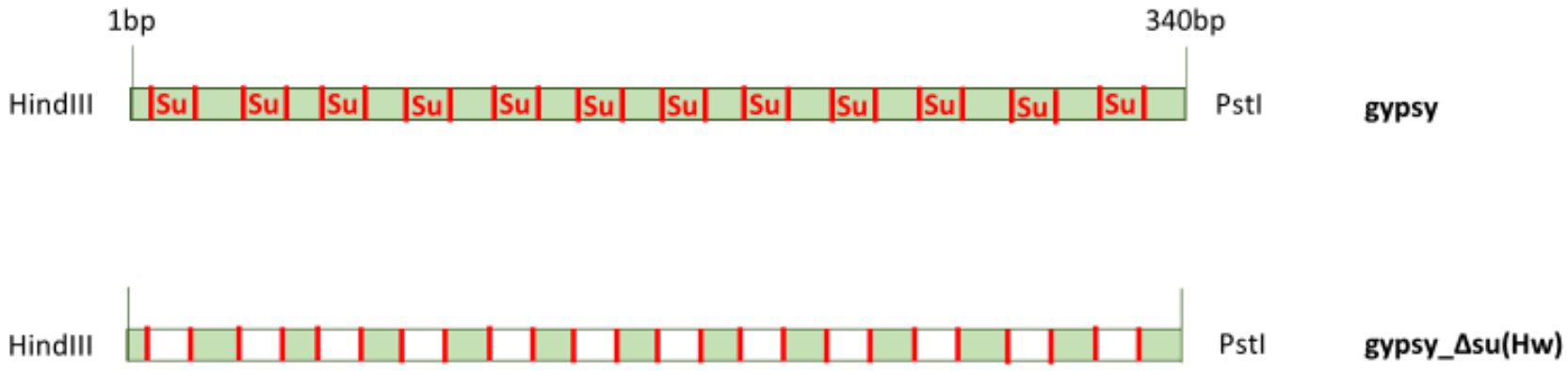


Figure 3.3 Clone map and names of all gypsy sequence permutations. Green rectangles represent the insulator sequence bordered by restriction sites. Length (in bp) of the full sequence is indicated at the top. Sequence orientation is represented by black arrows. Potential protein binding sequences are indicated between red bars. Deletions within the sequence are represented by white rectangles between red bars. Not drawn to scale.

3.3.2. Statistical analysis

Integration of T-DNA is random, thus producing pseudo-replicates. To determine significance of deleted, mutated and inverted sequences compared to their original insulator sequences, Pearson's Chi-squared test was performed, followed by the post-hoc comparison test Pairwise Nominal Independence Test for each insulator group. Tests were done in the R program, and statistically significant results were judged based on a p value of ≤ 0.05.

3.3.3. Bioinformatics/BLAST analysis

To study the association between the non-plant insulators studied and the plant genome, we performed Protein-Protein BLAST analyses (https://blast.ncbi.nlm.nih.gov/Blast.cgi) between the associated insulator binding proteins of the non-plant insulators and the *Arabidopsis* genome for potential protein orthologs. These included Rap1, (NCBI Accession No. P11938.2), CTCF (Accession No. P49711.1), and Su(Hw) (Accession No. NP524349), as well as secondary proteins Mod(mdg4) (Accession No. NP732623), CP190 (Accession No. NP524359.2, and dTopors (Accession No. NP1261083). Sequences with insulator activity possessing unidentified protein binding sites were examined for potential binding protein consensus sequences (specifically UAS_3'end). This was done in DNAman Lynnon Corporation using the software's Dot Matrix Alignment and Multiple Alignment program, as well as MEME Suites's CentriMo motif alignment program, aligning UAS_3'end sequence and known binding protein sequences from Table 2.2 in Chapter 2.

3.4. Results

3.4.1. Statement of contributions

pCAMBIA-derived pL1 plasmid vector used was previously constructed by Dr. L. Gudynaite-Savitch. Undergraduate students who contributed to the collection and analysis of transgenic plants include Hassan Badreddine, Onkar Bhanushaki, Adina Popescu, Linda Dam and Shukria Ahmadi. Adina Popescu cloned BEAD1c_5'end into pL1, Dr. Douglas Johnson cloned BEAD1c_3' end into pL1, and Hassan Baddreddine was responsible for cloning BEAD1cINV into pL1.

3.4.2. Defining insulator function

The insulator sequences, UASrpg and BEAD1c were previously tested by Gudynaite-Savitch *et al.* (2009) in *A. thaliana* for enhancer blocking properties. Their study determined that these cloned sequences significantly blocked 35S enhancer influence on seed specific napin promoter in the pL1 vector. Specifically, this study observed that out of 26 transgenic plants containing the UASrpg construct, 24/26 showed specific GUS expression and 2/26 showed non-specific GUS expression. In the case of BEAD1c, 48/55 transgenic samples expressed specific GUS staining and 7/55 expressed non-specific GUS staining. These results are indicative of strong insulator activity for both sequences. Lastly, the gypsy insulator which was tested by She *et al.* (2010b) and more recently Jiang *et al.* (2017) in *Arabidopsis*, concluded that flanking this insulator in their transgenes counteracted position effects facilitating high and precise expression of their transgenes, and generally increased independent transgene transcription levels, respectively. In addition to testing the gypsy sequence for its potential to reduce position effects, Singer et al. (Singer and Cox, 2013) tested a gypsy-like sequence from *A. thaliana* in a construct similar to the one in this thesis. Their

results showed that this gypsy-like sequence was successful in blocking enhancer-promoter interactions. We therefore want to test whether a single gypsy insulator can have similar transcriptional blocking effects, within our enhancer-blocking assay.

The assay system introduced sequences from different source species for testing in *A. thaliana*. It required that pre-existing proteins in the host that would normally interact with their plant target sequence, were able to functionally interact with a heterologous, but evolutionarily conserved introduced sequence. This may present a limitation and will be addressed in the Discussion.

3.4.3. Characterization of previously defined non-plant insulators

Three non-plant insulators were analyzed in this thesis: UASrpg from *Ashbya gossypii,* BEAD1c of *Homo sapiens*, and the gypsy insulator from *Drosophila melanogaster.* Permutations of these insulators in the form of deletions, base pair mutations, and sequence inversions, were tested to further characterize their ability to impede enhancer-promoter interactions within the pL1 vector and isolate the significant fragments within the sequence for their function. Tables 3.2A-C represents the data obtained from these analyses. To determine significance in the permutated sequences, we compared their insulator abilities (the percentage of seed specific GUS staining) to the insulator ability of the original insulator sequences.

The first candidate insulator to be investigated was UASrpg from *Ashbya gossypii* (Bi and Broach, 1999). Their analysis, which was based upon an assay testing the resistance of sequences to transcriptional repression when interposed between the HML (homothallic mating loci) α genes and the E silencer in the yeast *S. cerevisiae,* identified that multiple UASrpg sequences (from both *S. cerevisiae* and *A. gossypii*) were able to block silencer effects and function as heterochromatin barrier elements, or insulators. Specifically, they found that testing a DNA fragment containing

the Rap1 binding sites on UASrpg alone were necessary and sufficient in silencer-blocking activity (Bi and Broach, 1999). Later studies confirmed this through mutations in the Rap1 sites, identifying that the Rap1 sites were directly involved in barrier activity (Yu et al., 2003). Two regions in the 5' end of the fungal UASrpg clone that we have analyzed contains 2 binding sequences for Rap1 as identified by Steiner and Philippsen (1994): R1 (GCCCATACAT) and R2 (ATCCATACAT). Due to the high similarity between transcriptional repression at the HM loci in yeast and position effects in higher eukaryotes (reviewed by Braunstein et al. 1997), we hypothesized that the Rap1 binding sites, are important for insulator activity in *A. thaliana* as well. Following the observations made by (Gudynaite-Savitch et al., 2009) who identified that UArpg is in fact a functional and successful insulator in *A. thaliana,* we created fragments of this insulator and followed a similar mutation analysis to study the possible association of the sequence and Rap1 sites in plants.

The data in Table 3.2A represents the results for transgenic constructs containing the manipulated sequences of the UASrpg insulator. The clone UAS_5'end, containing the two candidate insulator protein binding sites R1 and R2, retains insulator function (60% of samples show seed specific staining) but at a level lower than UASrpg where 92.3% of samples show seed specific staining (Gudynaite-Savitch *et al.* 2009). The decline in insulator function of 32.3% is significant (p=0.0216) and perhaps indicates that this region alone is not sufficient for the activity for UASrpg. Despite a decline, we went on to determine whether there was still an association with the R1 and R2 sites on this fragment with insulator function. Further analysis of the 5'end of UASrpg was done to test the effects of the loss of the two Rap1 protein binding sites using the approach of Yu et al. (2003). Both UAS_mR1 (CC to AA mutation in the first Rap1 site – R1) and UAS_mR2 (CC to AA mutation in the second Rap1 site – R2) reduced enhancer-blocking activity.

When compared to UAS_5'end, UAS_mR1 showed a reduction in insulator activity of 28.79% (p=0.109) and UAS_mR2 showed a reduction of 51% in insulator activity (p=0.000589). These results indicate that although mutations in both sequences reduce insulator function, only the decline for UAS_mR2 is significant perhaps pointing to a larger role for R2 in insulator activity of UASrpg in *A. thaliana,* a heterologous interaction. At this time, we do not know why there is a difference between the R1 and R2 variants.

Our initial subcloning strategy focused on the Rap1 sequences identified by Yu et al. (2003); however, we also included a clone containing the 3' end of UASrpg as a control. To our surprise insulator activity was observed for UAS_3'end at 97%, comparable to the full UASrpg sequence (92.3%), suggesting that this fragment alone is sufficient for insulator function. To further analyze the potential influence of non-plant proteins on insulator function, we searched for sequences within UAS_3'end that may bind to other protein consensus sequences using DNAman software. Two potential candidates were identified and tested; a sequence showing homology to the binding site consensus sequence of the gypsy insulator protein Su(Hw), and another sequence showing homology to the binding site consensus sequence of the vertebrate insulator protein CTCF.

Deletion of the 9bp potential Su(Hw) binding sequence CGCTGCATA (UAS_3'ΔSu) reduced insulator activity by 54.9% (p=0.0000292) suggesting a potential role of this site in insulator function of UASrpg in *A. thaliana*. Following the line of reasoning described above for the UASrpg-Rap1-like protein interaction mediated by the sequence R2, the interaction between the 9bp Su(Hw) consensus binding sequence and a SuHw-like protein defines another insulator evolutionary-conserved system in *A. thaliana*. Deletion of the 5bp CTCF-like core binding sequence CTCCC (UAS_3'ΔCTC) produced a reduction of 14.6%, which was not statistically significant (p=0.210), thus not supporting a direct role for this sequence. More experiments will

confirm this finding, for example, creating larger deletions that remove more of the CTCF consensus sequence or investigating the insulator activity of this mutation in a UAS_3'ΔSu background. These possibilities are described in the Discussion.

The second insulator under examination is the human insulator, BEAD1c (Zhong and Krangel, 1997), which contains a sequence coined BEADA that contains high homology to the enhancer-blocking CTCF site of the 5'cHS4 chicken insulator containing the CCCTC binding core. BEADA also binds specifically to purified CTCF (Bell et al., 1999). CTCF is an eleven-zinc finger DNA-binding protein that is highly conserved in vertebrates (Filippova et al., 1996), and binds to numerous vertebrate insulators, whose consensus sites have been analysed for enhancer-blocking activity. (Bell et al., 1999). Specifically, when the BEADA sequence (CCCAGGCCTGCACTGCCGCCTGCCGGCAGGGGTCCAGTC) was analyzed by Bell *et. al.* (1999), the site alone proved to be an effective enhancer blocker as the deletion of the BEADA sequence largely eliminated insulator activity compared to the whole sequence. Interestingly, although BEADA binds to CTCF and has high homology to the CTCF site of the chicken insulator, it does not contain the specific CCCTC core. This implies high variation in consensus sequence across species. We have created several deletions to determine whether this highly conserved sequence, BEADA, was also critical in our *A. thaliana* enhancer-blocking assay.

The data obtained for analysis of the second insulator, BEAD1c, is represented in Table 3.2B, including the positive control carried out by Gudynaite-Savitch (2009). In a way similar to the previous analysis, we have created 3 deletions of the BEAD1c human insulator. When the sequence containing the 5' end of BEAD1c, BEAD1c_5'end, was tested for enhancer-blocking activity, we saw a significant reduction of specific GUS expression by 59.2% (p=0.0000000818) suggesting that this fragment may not play a large role in insulator function and the deleted

fragment plays a larger role. Testing insulator activity of the 3'end of BEAD1c, BEAD1c_3'end, which contains the majority of the BEADA sequence, showed a reduction by only 18.3% compared to the full BEAD1c sequence (p=0.0820) which is not significant. This result confirms that the fragment containing the 3'end is more involved in the insulator activity of BEAD1c in *A. thaliana*, conceivably due to BEADA sequence (the CTCF binding site) on this fragment which is responsible for insulator action in a variety of vertebrates. This led us to directly test whether the BEADA site is in fact responsible for insulator activity in *A. thaliana* by creating a sequence which removes the BEADA site completely: ΔBEADA. The sequence with this deletion reduced enhancer-blocking activity by a significant 24.1% compared to the full sequence of BEAD1c (p=0.0484), suggesting the importance of this site in insulator function in *A. thaliana.* However, it also may indicate that other sequences within the cloned fragment that we have not analysed may have insulator function as was previously observed with UASrpg.

The final insulator analysis tested the *D. melanogaster* gypsy insulator This 340bp DNA fragment has been well characterized in *D. melanogaster* and has been studied in *A. thaliana* as an element possessing the ability to protect against position effects (Jiang et al., 2017; She et al., 2010a; Singer and Cox, 2013). In addition, a gypsy-like sequence has also been tested in *A. thaliana* (Singer and Cox, 2013). Our goal is to test whether the gypsy sequence can block interactions between the CaMV35S enhancer and napin promoter to determine whether it is also able to function in this assay system. I was also interested in characterizing this sequence to understand what exactly is involved in its function, as the gypsy insulator protein, Su(Hw), has not been identified in *A. thaliana.* Our study of this sequence involved an insertion of the full sequence in our pL1 vector, as well as a sequence in which we deleted all 12 of the potential Su(Hw)-like protein binding sites. When the full gypsy sequence or the deletion was inserted between the

enhancer and promoter we observed no specific GUS expression (Table 3.2C). We can deduce from these results that this element is not a functional insulator in our assay system.

An NCBI Protein-Protein BLAST search was done on all the known insulator proteins to determine whether protein orthologs existed in the *Arabidopsis* genome. Protein accession numbers were added and *Arabidopsis thaliana* database was chosen. The search discovered no protein orthologs for Rap1, Su(Hw), or CTCF.

In this section we demonstrated that insulators from non-plant species could function successfully as enhancer-blocking elements in *A. thaliana*. The results for all three insulator sequences that were tested suggested that associated insulator protein binding sites were dependent on the fragment tested. In UASrpg, only one Rap1 binding sequence was observed to be important in insulator function. There was evidence for gypsy function when the single Su(Hw) binding site was removed from UASrpg3' but not when 12 copies of the Su(Hw) binding site were removed from gypsy. The removal of a CTCF binding site from UAS_3'end did not show loss of insulator activity while removal of the CTCF binding site from BEAD1c did reduce insulator activity significantly. These results together point to the importance of context of other sequences in the transgene as being important and will be described in the Discussion.

3.4.4. Does sequence orientation affect insulator activity?

In addition to characterizing insulator sequences with respect to their protein binding sites, studies have also tested sequences in their inverted forms (reverse sequences) to determine whether insulator function is orientation dependent. Some vertebrate studies have found insulator sequences to function better in one orientation over the other (Bell and Felsenfeld, 2000; Hark et al., 2000). Similar results were found in a plant study of the TBS (Transformation Booster

Sequence) which tested its ability to block enhancer promoter interactions between a *PIp* (petal- and stamen-specific PISTILLATA) promoter and CaMV 35S enhancer construct. Histochemical GUS analysis within floral and leaf tissues of transgenic *A. thaliana* determined that the forward TBS sequence showed significantly greater insulator activity than the same sequence when reversed (Singer et al., 2011). In a similar fashion, we have tested various sequences within the pL1 vector for insulator activities. The analysis encompassed testing the inverted sequences of BEAD1c, UASrpg, UAS_5'end, and UAS_3'end (Table 3.3). In all these examples forward and inverse are defined with respect to their orientation adjacent to the gene of interest *in situ*.

Inverting UASrpg (UASrpgINV) decreased enhancer-blocking activity by a significant 42.3% ($p=0.00462$), compared to the original UASrpg forward sequence. The inversion of UAS_5'end (UAS_5'INV) decreased enhancer-blocking activity by 52% ($p=0.000136$) compared to its forward sequence, and inverting UAS_3'end (UAS_3'INV) decreased enhancer-blocking activity by 28.6% ($p=0.0130$) compared to its forward sequence. Inversion of these three sequences significantly reduced insulator activity, suggesting a possible relationship between insulator orientation and activity of UASrpg in *A. thaliana*.

The inversion of the full length BEAD1c insulator (BEAD1cINV) only decreased the ability of the sequence to block enhancer-promoter interactions by 3.4% ($p=1.0$) compared to the forward BEAD1c sequence, indicating that orientation of the 548bp BEAD1c sequence does not play an effect on it's insulating function in our assay.

Overall, the inversion analysis demonstrated that orientation plays a substantial role in insulator function of the UASrpg sequence in our assay, while there is no significant effect in

sequence orientation in the BEAD1c sequence. The difference in size may be a factor in helping to understand and interpret these results as will be deliberated in the Discussion.

Table 3.2: The β-glucoronidase (GUS) staining results for the deletions and mutations of non-plant insulator sequences in the pL1 vector. The columns indicate the construct name and size, the number of individual transformants tested for leaf, flower, silique and seed staining, and a calculation of the specific and nonspecific expression of the GUS transgenes. * indicates data produced from Gudynaite-Savitch (2009).

A. Data representing constructs containing UASrpg sequence deletions and dinucleotide mutations.

			GUS staining				GUS expression		
Construct Name	**Size (bp)**	**No. plants tested**	**Flowers**	**Leaves**	**Siliques**	**Seeds**	**Specific**	**Nonspecific**	**None**
pL1		53	41	49	42	41	2/51	49/51	2
UASrpg*	246	26	0	0	2	26	24/26	2/26	0
UAS_5'end	149	37	0	9	8	34	21/35	14/35	2
UAS_mR1	149	18	5	10	11	13	5/16	11/16	2
UAS_mR2	149	21	12	18	14	20	2/21	19/21	0
UAS_3'end	109	41	1	1	1	33	32/33	1/33	8
UAS_ΔSu(Hw)	100	20	1	2	11	18	8/19	11/19	1
UAS_ΔCTC	104	20	0	1	4	16	14/17	3/17	3

B. Data representing constructs containing BEAD1c sequence deletions.

			GUS staining				GUS expression		
Construct Name	**Size (bp)**	**No. plants tested**	**Flowers**	**Leaves**	**Siliques**	**Seeds**	**Specific**	**Nonspecific**	**None**
BEAD1c*	548	56	3	4	7	55	48/55	7/55	1
BEAD1c_5'end	265	32	16	16	19	32	9/32	23/32	0
BEAD1c_3'end	301	31	0	1	8	29	20/29	9/29	2
ΔBEADA	509	20	2	2	7	19	12/19	7/19	1

C. Data representing constructs containing gypsy sequence deletions.

Construct Name	Size (bp)	No. plants tested	GUS staining				GUS expression		
			Flowers	Leaves	Siliques	Seeds	Specific	Nonspecific	None
gypsy	352	20	8	17	19	8	0/20	20/20	0
gypsy_ΔSu(Hw)	268	9	5	5	7	6	1/8	7/8	1

Table 3.3: The β-glucoronidase (GUS) staining results for the inversions of non-plant insulator sequences in the pL1 vector. The columns indicate the construct name and size, the number of individual transformants tested for leaf, flower, silique and seed staining, and a calculation of the specific and nonspecific expression of the GUS transgenes. * indicates data from Gudynaite-Savitch (2009).

			GUS staining				GUS expression		
Construct Name	**Size (bp)**	**No. plants tested**	**Flowers**	**Leaves**	**Siliques**	**Seeds**	**Specific**	**Nonspecific**	**None**
UASrpg*	246	26	0	0	2	26	24/26	2/26	0
UASrpgINV	246	20	1	2	10	17	9/18	9/18	2
UAS_5'end	149	37	0	9	8	34	21/35	14/35	2
UAS_5'INV	149	26	16	20	22	25	2/25	23/25	1
UAS_3'end	109	41	1	1	1	33	32/33	1/33	8
UAS_3'INV	109	21	1	5	2	17	14/20	6/20	1
BEAD1c*	548	56	3	4	7	55	48/55	7/55	1
BEAD1cINV	548	20	0	1	3	18	16/19	3/19	1

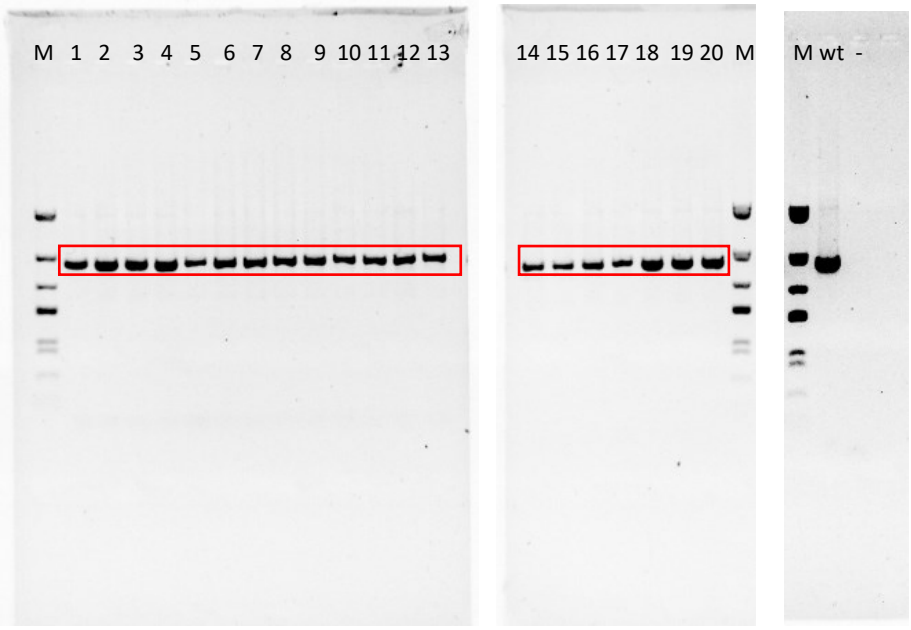


Figure 3.4: Example of agarose gel electrophoresis showing PCR amplification of clone ΔBEADA in pB31 with SALK primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate ΔBEADA insulator are shown. Amplification of DNA used SALK_049131_RP2 and SALK_049131_LP2 primers to produce a band with an expected size of 762bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449,434), 298, 267, 174, 102, 80
1-20: Transgenic plant DNA containing ΔBEADA inserts
wt: *Arabidopsis thaliana* wild type DNA
- : Negative control containing water in place of DNA

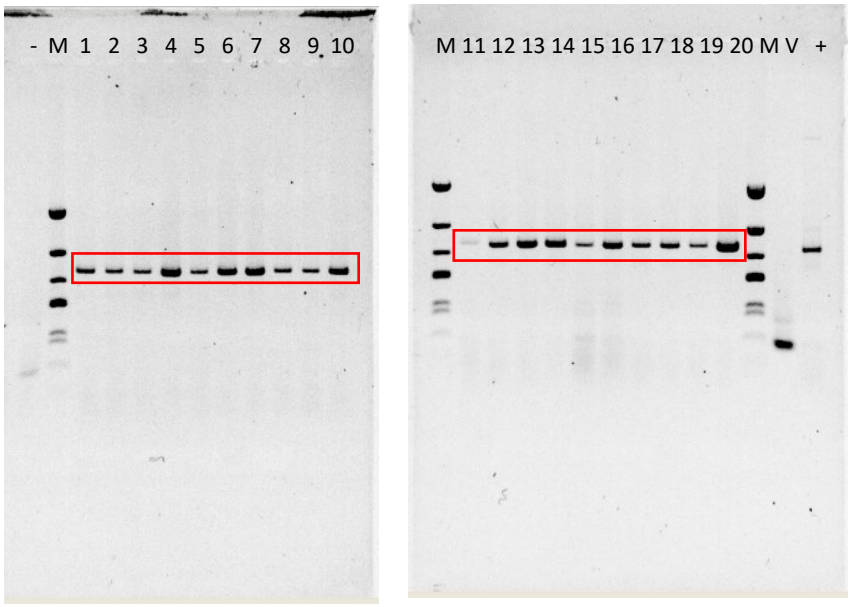


Figure 3.5: Example of agarose gel electrophoresis showing PCR amplification of clone ΔBEADA in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate ΔBEADA insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 679bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-20: Transgenic plant DNA containing ΔBEADA inserts (679)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
- : Negative control containing water in place of DNA

3.5. Discussion

The eukaryotic genome contains a multitude of factors and elements all working together to produce the various transcriptional outputs required for gene expression. Insulator sequences are a necessary element in this machinery. Insulators control gene regulation by creating boundaries in the genome preventing unwanted interactions between distinct genes. To function, these DNA sequences contain protein binding sites where specific primary and secondary proteins bind, creating loops in the chromatin to separate regulatory elements, or gaps to segregate active and inactive chromatin. The hypothesis is that the function of non-plant insulators is conserved in plants, along with their protein binding sites, ultimately sharing very similar mechanisms. In fact, the proteins bind consensus sequences on the sites, demonstrating variations of protein binding sites within eukaryotic species, and possibly between species. The function of non-plant insulators suggests that the mechanism involved may be conserved. Our goal is to study 3 non-plant insulator sequences (BEAD1c, UASrpg, and gypsy), through deletions, mutations and inversions and study the effects of their potential protein binding sites in *A. thaliana,* and ultimately describe the potential mechanism at work in *A. thaliana.*

3.5.1. Functional insulators in plants

The most comprehensive research on insulators has focused on *D. melanogaster* as a model system. Most *D. melanogaster* insulators have been studied and characterized, guiding thinking about models and mechanisms for the function of eukaryotic insulator species. The structure of chromatin is highly conserved in eukaryotes, suggesting that these mechanisms can also be applied to plants and that insulators of other species are also functional in plants. Research on insulator testing across species has been very successful, such as fungal sequences in yeast (Bi and Broach,

1999) or chicken sequences in human cell lines (Chung et al., 1993). Multiple studies tried this hypothesis by testing insulator function of non-plant sequences in plant species. Examples include testing the yeast ARS-1 SAR (Allen et al., 1993), the chicken lysozyme MAR (Mlynarova et al., 1994; Nagaya et al., 2001) and the sea urchin Ars insulator (Nagaya et al., 2001), for their ability to protect against positional effects in *N. tabacum.*

Studies have been testing both plant and non-plant insulator sequences for function in *A. thaliana.* Hily et al. (2009) initially showed that a TBS (transformation booster sequence) from *P. hybrida* functions as a successful enhancer-blocking insulator in *A. thaliana.* Their assay consisted of a TBS sequence cloned between a CaMV 35S enhancer and a flower-specific AGAMOUS second intron-derived promoter (AGIP) and assayed for histochemical and fluorometric GUS expression in one tissue type (leaves). Their analysis showed that most transgenic lines containing the TBS sequence lacked detectable GUS expression in leaves - 69% which is indicative of insulator function in their assay.

A gypsy-like sequence (2258bp) in *A. thaliana* was also found to elicit enhancer-blocking effects (Singer and Cox, 2013). The assay determined whether the Atgypsy-like fragment was able to reduce CaMV 35S enhancer-mediated constitutive activation of a nearby petal- and stamen specific PISTILLATA promoter (PIp) when situated between the two. Two tissue types, leaves and flowers, were analyzed by histochemical and fluorometric assays for GUS expression. The analysis showed 75% of lines displayed a reduction of 35S enhancer mediated activation of PIp with GUS only in flower tissue (indicative of insulator function), compared to the vector control which displayed GUS staining in leaf and flower tissue in every line.

In studies which tested non-plant insulators in *A. thaliana*, similar approaches have been made. (Gudynaite-Savitch et al., 2009) demonstrated that UASrpg (*A. gossypii*) and BEAD1c (*H. sapiens*) were successful at blocking CaMV 35S enhancer-napin promoter interactions when placed between the two. In their histochemical analysis, 92.3% of UASrpg transformed samples and 87.3% of BEAD1c transformed samples had a pattern of GUS expression indicating insulator function.

In all of these studies 100% insulator function or 0% insulator function was rarely observed. The ability of sequences to reduce, but not completely eliminate misexpression or position effects is common among insulator studies. In our assay the possibility of deletion of targeted sequences to reduce but not completely eliminate insulator function could be due to multiple sequences on the insulator. For example, deletion of CTCF site on BEAD1c (ΔBEADA) did not reduce insulator activity of BEAD1c to zero in our assay, possibly because within the sequence (5' or 3' of BEAD1c) there are other sequences contributing to its function.

The heterologous assays may produce varying results to one another for multiple reasons (for example the function of CTCF in human cells vs plant systems). It was hypothesized that variances in insulator function may be due to two competing factors: the conserved mechanism of insulator sequences to require their specific binding proteins, and the sequence divergence between plants and recognition sites (Gudynaite-Savitch et al., 2009). In order to fully determine true insulators, we believe it is necessary to thoroughly study sequences in plants and characterize them via sequence manipulations.

3.5.2. Protein binding

The sequences under examination (deleted and mutated fragments) were tested in the pL1 vector for insulator function based on their ability to block interaction between the 35S enhancer and seed specific napin promoter, when placed between them. The analysis calculated transgenic samples which expressed histochemical GUS expression in leaf, flower, silique, and seed tissues. Specific, non-specific, and no GUS staining was determined for each transgenic sample. Those samples with specific GUS staining in seeds only, signified an ability to block enhancer-promoter interactions and ultimately insulator function in plants.

The sequences under investigation come from non-plant species. They contain protein binding sites which have been found to be functional in non-plant species, and the sequences are homologous to insulator protein binding site consensus sequences. Since these insulator sequences function in *A. thaliana,* this thesis proposes that the mechanism behind insulator function is conserved in plants and therefore the proteins are conserved. However, we would expect divergence in the proteins and sites due to the evolutionary distance between plants and the source organisms.

Our overall deletions and mutations analysis of UASrpg, BEAD1c, and gypsy provided us with sufficient data to claim the presence of multiple proteins are involved in insulator function in *A. thaliana.* The characterization of UASrpg by nucleotide mutations in the protein binding sites, showed that on the 5' end of UASrpg, the second Rap1 site, R2, was significantly involved in insulator function for UASrpg, although R1 was not. And on the 3' end of UASrpg, through a computational discovery and deletion analysis, a Su(Hw) binding sequence was observed to be important for insulator function of UASrpg, yet a CTCF potential binding sequence did not. In contrast, deletion analysis of the CTCF binding site in BEAD1c, BEADA, reduced insulator

function significantly, suggesting its importance in insulator function in our assay. Interestingly, the gypsy sequence containing 12 Su(Hw) protein binding sites, did not function as an insulator in our enhancer-blocking assay, despite one Su(Hw) site playing a significant effect in UASrpg as previously mentioned.

It is unclear why in our assay, certain protein binding sites are important for insulator function in one fragment and not in another. In the analysis of UAS_5' end, a dinucleotide mutation in R2 resulted in a significant decline in insulator function. However, dinucleotide mutation of the R1 site did not produce significant decline in insulator activity. Still, we believe this does not rule out the significance of R1 in insulator function in *A. thaliana* completely, as studies have shown that the function of Rap1 sometimes relies on multiple Rap1 sites for full insulator function (Bi and Broach, 1999; Yu et al., 2003). In an enhancer-blocking study of the 3 Rap1 sites (R1, R2 and R3) in the *S. cerevisae* TEF2-UAS insulator, deletional analysis of the sequences found that the insulator function of deleted fragments would only work in particular combinations of the Rap1 sequences. For example, R1 and R2 alone would not significantly act as insulators (Bi and Broach, 1999). This suggests that their activities likely require the coordinated actions of more than one Rap1-binding sites. Similarly, a mutation study of the TEF2-UAS insulator found that mutations in two of the three Rap1 sites abolished its barrier activity, indicating that the association of Rap1 with the TEF2-UAS insulator at more than one site was required for insulator function (Yu et al., 2003). These findings suggest that the function of Rap1 likely requires the coordinated actions of multiple Rap1 binding sites. Mutation of our R1 site decreased insulator function, although it was not statistically significant. The significance of both Rap1 binding sites for insulator function in yeast compared to *A. thaliana,* which only showed significance in one site, could also be due to

differences in assay systems. Heterologous assays test insulators with different enhancer/promoters, including slight differences in chromatin structure in the example of yeast.

Further analysis would be beneficial in determining whether R1 DNA sequence in this assay is functioning in a similar way to that same R1 DNA sequence in the assays of Yu et al. (2003) and Bi and Broach et al. (1999), which binds to the Rap1 protein for function. For example, an experimental approach to test this theory would be to clone another sequence containing mutations in both R1 and R2 sites and testing it for enhancer-blocking abilities in the pL1 vector. The result would identify whether there is a difference in function compared to the mutation in R2 alone, identifying whether there is some coordinated action from R1 that is only significant with R2. It would also be beneficial to increase sample sizes for all mutation constructs.

In the case of the CTCF binding site, the observation that it functions in BEAD1c and not in UASrpg (UAS_3'end) may be due to the length of deletion. We chose to delete the minimal CTCF core sequence (CTCCC) according to the approach of Batool Gandorah and Lara Rasooli. However, true protein binding sites would likely require experimental approaches to determine whether CTCF actually binds to a sequence in UASrpg. In contrast to the experimentally determined BEADA sequence (CTCF site) in BEAD1c, the length of the sequences deleted in our UAS_3'end analysis (5bp) significantly differed to BEADA (17bp). It is possible that the CTCF core consensus sequence we have deleted may not have been a large enough deletion to have played a significant change or to have been functional. In order to test this possibility, a protein binding analysis (Chromatin Immunoprecipitation (ChIP) or DNA electrophoretic mobility shift assay (EMSA)) must be done to determine which proteins actually bind to UAS_3'end. From there, deletion of the experimentally determined site would provide a better conclusion on the importance of the CTCF site in UASrpg.

The gypsy element has been highly characterized in *D. melanogaster* and recently has been shown to function in *A. thaliana.* Studies have tested the sequence for its ability against position effects (Jiang et al., 2017; She et al., 2010b) and a gypsy-like sequence from *A. thalina* can function as an enhancer-blocker (Singer and Cox, 2013). Our analysis has shown that although the full-length gypsy with its 12 consensus sites for Su(Hw), is not functional in our assay, a site within the 3'end of UASrpg containing only one Su(Hw) consensus sequence, is significant for function. It is surprising that deletion of one consensus sequence alone significantly reduced insulator activity in one insulator and not the other. Several possibilities should be considered. For example, the function of UASrpg 3' fragment may not be due to an association with the Su(Hw) protein at all. It is possible that the deletion of the 9bp sequence (CGCTGCATA) may have affected an overlapping sequence, reducing the efficiency of an unidentified plant insulator protein to bind. In accordance to the BLAST analysis which did not find any Su(Hw) protein orthologs in *A. thaliana*, Su(Hw) may not be a factor in enhancer-blocking insulator function in our specific assay. Another possibility could be that the gypsy sequence is enhancer- or promoter-specific and non-functional with the seed specific napin promoter. Studies have shown that some insulators are specific to certain regulatory elements and even species (Gudynaite-Savitch et al., 2009; Nagaya et al., 2001). To test this, it would be necessary to clone the gypsy insulator into varying vectors containing varying enhancers and promoters to determine any differences in results.

3.5.3. The role of insulator orientation

The inversion analysis of the UASrpg elements (UASrpg, UAS_5'end, UAS_3'end) revealed significant reductions in enhancer-blocking activity in the pL1 vector for all 3 experiments when tested in the reverse orientation, suggesting that these insulators are orientation dependent in *A. thaliana*, which has been commonly found in other studies (Abhyankar et al., 2007; Barges et al.,

2000; Bell and Felsenfeld, 2000; Bell et al., 1999; Hark et al., 2000; Singer et al., 2011). A review by West *et al.* (2002) suggested that insulators often contain additional regulatory elements due to their length, such as an upstream enhancer in the case of the human *apoB* insulator (Antes et al., 2001) or the 5'HS5 insulator of the chicken β-*globin* locus (Chung et al., 1993). Since enhancer-blocking insulators are known to be functional only when situated between the enhancer and promoter, compound elements containing both enhancer and enhancer-blocking activities display polarity. When an insulator-enhancer element is orientated with the insulator closer to the promoter, that enhancer as well as others more distal will be blocked. When the combination element is reversed, the distal enhancers are still blocked, but the insulator-associated enhancer is now free to activate the reporter. However, with the case of our sequences, the length is much shorter than those previously studied. We believe that in a similar fashion our sequences are polar dependent in a way that the reverse orientation creates new end sequences which may interfere with insulator activity.

The BEAD1c sequence did not however produce an orientation dependent outcome compared to the UASrpg analysis with regards to the sequence inversion experiments. In contrast to UASrpg, the ability of reverse sequence of BEAD1c to block enhancer-promoter interactions remained intact and insulator activity was comparable and statistically similar. This result has been observed previously as well in Fab7 and gypsy insulators which were tested in forward and reverse orientations, within a construct containing an enhancer of *copia* and an *Hsp 70* promoter in *D. melanogaster* (Tchurikov et al., 2009). Larger sequences such as Fab7 (1.2kb), BEAD1c (548bp) and gypsy (340bp) have no orientation dependence for function while smaller sequences UASrpg (246bp), UAS_5'end (149bp), and UAS_3'end (109bp) demonstrate some degree of orientation

dependence. Perhaps the variation in size affects the insulator function, sequences smaller than ~300bp start to show a size effect on orientation.

3.5.4. The influence of enhancers, promoters, and other genomic elements

The effectiveness of insulators seems to be influenced by various genomic and structural factors, and not necessarily species specific. Research has shown that insulators in one context will function exceptionally, while a change in a single element, such as an enhancer, may reduce their functionality dramatically. For example, the 16bp NI29 insulator which was reported to eliminate interaction between the CaMV35S enhancer and core promoter in *A. thaliana* (Gan and Xie, 2002), however was unable to abolish interactions between the tCUP enhancer and napin promoter, or a CaMV35S enhancer and napin promoter in *A. thaliana* (Gudynaite-Savitch et al., 2009). In these cases, the activity, or lack of activity, could be due to the nature of enhancers involved. The activity of a strengthened enhancer, directing higher levels of transcription, is more difficult to block, as described by Scott *et al.* (2001) who tested insulators against a 'Mega enhancer' (FBE1 enhancer with increased number of transcription factor binding sites which increases levels of promoter activity), and found that chromatin insulation is less effective at blocking the Mega enhancer compared to the single FBE1 enhancer, indicating that the nature of regulatory interactions within a gene impacts the effectiveness of an insulator.

The inability of gypsy to function as an enhancer-blocker in our *A. thaliana* assay, despite being effective in minimizing position effects of transgenic *A. thaliana,* may not be a surprising outcome considering the insulator's complex nature, with regards to its mechanisms in mediating changes in chromatin structure and nuclear organization. The gypsy sequence seems to be more selective in plant transgenics as we see more variances in insulator function between plant species,

and within the same species. For example, although gypsy is effective in mitigating position effects in *A. thaliana*, when 2 sequences flank a transgene (Jiang et al., 2017), it is unsuccessful in achieving the same function in transgenic tobacco cells (Nagaya et al., 2001). The gypsy sequence's lack of enhancer-blocking abilities in our assay could in part be due to a requirement of multiple sequences for function in plants. Since no ortholog of Su(Hw), the gypsy insulator protein, is found in *A. thaliana*, its function is not currently understood in plants.

3.5.5. Insulator conservation across species

Insulator function across multiple species helps us predict whether there is an association between species and ultimately conservation in mechanisms. There is growing evidence of non-plant insulators able to function in plants, permitting us to believe that current insulator models may be applied to plants. The ability of the non-plant sequences investigated in this thesis to function successfully as insulators in *A. thaliana* suggests there may be some conservation of insulators across species inferring the conservation of proteins. It is clear that insulators may function differently in heterologous systems due to variances in enhancers/promoters or chromatin structure in the case of yeast. This analysis has identified multiple possible binding sites for Rap1-like, CTCF-like, and Su(Hw)-like plant proteins.

The direct mutation and deletion analysis of previously determined consensus sequences for Rap1-UASrpg (R2) and CTCF-BEAD1c (BEADA), significantly reduced insulator function in transgenic *A. thaliana.* The mode of action of these proteins in *A. thaliana* has not been studied previously. We believe though, that the function of these sequences in our assay, provides evidence that they are conserved, and possibly their mechanism of action as well. The models previously described in the Introduction of this thesis include (1) the Promoter Decoy Model, (2) the Physical

Barrier Model and (3) the Loop Domain Model. These models are not necessarily mutually exclusive. Still, there is no single inclusive model of insulator function (Matharu and Ahanger, 2015).

A growing collection of evidence has been made for the Loop Domain model, which is the model of focus in this thesis. It has best described the function of the highly characterized gypsy insulator, suggesting that insulator function is due to the formation of chromatin loops from the binding of Su(Hw), and its secondary proteins Mod(mdg4), CP190, dTopors, to the nuclear lamin. These loops separate enhancers and promoters interfering with their direct interaction. Vertebrate insulators containing CTCF binding protein sites have also been found to function by binding to nucleolar proteins found inside the nucleolus, forming chromatin loops by tethering chromatin to the surface of the nucleolus (Yusufzai et al., 2004). Similar to the gypsy element, CTCF has been found to interact with a secondary element, cohesion, a critical partner of CTCF in mediating inter- and intra- chromosomal interactions necessary for insulator function (Yang and Corces, 2011). We postulate that the function of BEAD1c in *A. thaliana* and the significance of the CTCF binding site for insulator function in our assay indicates a conserved mechanism of this protein in plants, and that the insulators which we have studied function via the Loop Domain Model. Further studies will be required to determine whether plants function following one of the other models as well, since these models are not mutually exclusive, and the function of insulators is very complex.

3.6. Conclusions

Our genetic analysis of previously defined non-plant insulators has identified multiple protein binding sites that show significance in insulator function in *A. thaliana*: Rap1 site in UASrpg, Su(Hw) site in UASrpg, and CTCF site in BEAD1c. The identification of these sites suggests

possible protein conservation between the species and therefore infer conservation of mechanisms. The identification and analysis of these protein binding sites is the first in plant systems.

Further analysis on the protein binding sites will be required to confirm whether these sequences can activate insulator activity alone or in tandem. For example, cloning the BEADA sequence and the Rap1 binding site sequences in tandem, and testing for enhancer blocking abilities in the pL1 vector would infer whether these sites alone are essential for function. Identification of the proteins that bind these sites would also clarify their roles in insulator function. Literature and our BLAST analysis has not identified protein orthologs of Rap1, Su(Hw), or CTCF in *A. thaliana*. However, *A. thaliana* does contain numerous zinc-finger proteins similar to Su(Hw) and CTCF. This suggests that there may be an *A. thaliana*-like CTCF or Su(Hw) insulator protein. Experimental analysis would be necessary and identify actual proteins binding to our DNA using techniques such as Chromatin Immunoprecipitation (ChIP) or DNA electrophoretic mobility shift (EMSA).

Once more research on plant insulators and their binding proteins are readily available, we can build stronger models to support their function. The idea that no one insulator model describes all insulator function is evident in the data produced here. As we have observed from the deletion analysis in UASrpg in which the 3'end alone produced strong and comparable insulator activity compared to the full UASrpg sequence, multiple systems may be at work in a single insulator, and possibly interfering with one another. This suggests that perhaps insulators in plants have multiple roles as in the example of the scs (special chromatin structure) insulator from *D. melanogastor* containing both enhancer-blocking and barrier element functions (Cai and Levine, 1995; Kellum and Schedl, 1992; Vazquez and Schedl, 1994), as was described in the Introduction. The ability of the insulators investigated here to function in *A. thaliana* suggests that insulators are conserved

across species and like other species, plants contain multiple, functional insulators. The protein binding sites that were found to be significant to their function suggests that in turn their protein binding sites are also conserved. This implies that the mechanisms by which other species function may be used to describe plant insulation once further research is done.

4.0 Thesis summary and future directions

4.1. Chapter 2 and Chapter 3 results summary and implications

We hypothesized that plants have multiple insulator types able to block unwanted interactions between enhancers and promoters, and that insulators are evolutionarily conserved across species. The implications of this statement are that the mechanism(s) behind insulator function as well as the protein binding sites and the insulator proteins themselves are conserved. We described in this thesis a selection method which not only tests but also identifies novel insulator sequences in *A. thaliana* using a random oligonucleotide library. The results of this three-step analysis identified four insulator sequences we believe will be functional in all plants: InI-3, InII-12, InIII-50 and InIII-78. We also believe that we have identified essential sites of the three previously defined non-plant insulator sequences, UASrpg from *Ashbya gossypii*, BEAD1c from *Homo sapiens,* and gypsy from *Drosophila melanogaster*. To date, few studies have identified plant insulators or their associated insulator proteins. The data obtained in this thesis will be beneficial to understanding and identifying the mechanisms behind insulators.

In Chapter 2, we presented an in-depth analysis to select for insulator sequences *de novo* using our three-step selection method. Candidate sequences were cloned sequentially into three selection/screening vectors: pC1, pB31, and pL1. Each containing the cauliflower mosaic virus (CaMV) 35S enhancer and various promoters fused to a reporter gene. Each vector has been chosen

to test the sequence for different properties and addressing possible limitations of one another, such as mutations during transformations in pC1, or repressor competition in pL1. In addition, our system rigorously tests sequences in multiple tissue types, assessing for tissue specificity, compared to most published studies. The novel sequences we have revealed to possess insulator function in *A. thaliana* are approximately 150bp and 450bp long, and through a preliminary bioinformatics study we have identified numerous factors which may bind to these novel insulators. The goal of this bioinformatics search was to find potential protein binding sites of insulators from non-plant species to test the hypothesis of conservation of insulator function across species. Unfortunately, the results for potential protein binding sites on our sequences tested positive for sequences from our analysis which had no insulator function as well. In addition, matches were not statistically significant, and therefore results remain to be ineffectual and suggest the possibility of false positives. Furthermore, our search for repressor sites produced similar outcomes. Supplementary analysis for potential protein binding sites on these novel sequences must be done in order to discover true potential protein associations. We also point to ways in which the technology can be streamlined and improved.

The results in Chapter 2 show a way forward that can be used to detect novel, functional insulator sequences, as a step towards detecting multiple insulator pathways and further test our hypothesis. Unlike the results presented in Chapter 3, no candidate sequences need to be identified prior to testing.

In Chapter 3, we present a comprehensive characterization of previously defined non-plant insulators in *A. thaliana.* Our goal was to localize sites important in insulator function, specifically potential protein binding consensus sites. In particular, we have identified three protein binding sites potentially involved in insulator function in *A. thaliana.* A Rap1 site (R2) on UASrpg, a

Su(Hw) site on UASrpg, and a CTCF site (BEADA) on BEAD1c. It is unclear why in our assay certain protein binding sites are important for insulator function in one fragment and not in another. In UASrpg, although the sequence contains two Rap1 proteins, our results suggest that only R2 is significantly important in insulator function of UASrpg. Further mutational analysis of the Rap1 sites should be done to make any conclusion on the significance of R1, since the mutation on the R1 site did in fact reduce some insulator activity in UASrpg. We propose creating a sequence with mutations in both Rap1 sites, as well as cloning R1 and R2 in tandem and testing the abilities of these fragments to act as insulators. This would provide a more rounded analysis on the sequence.

The significance of CTCF protein site in BEAD1c and not in the UASrpg sequence could be attributed to multiple reasons. First, it is clear that the experimentally determined site on BEAD1c is much longer in length (17bp) compared to the one we deleted on UASrpg (5bp). This difference in size may be a factor for why the observations differed between the two insulators. It is possible that the CTCCC deletion we produced is much too short in size for any function of the protein. Second, it may be possible that there is actually no CTCF site on UASrpg at all. In order to test this theory experimental analysis would have to be done such as Chromatin Immunoprecipitation (ChIP) or DNA electrophoretic mobility shift assay (EMSA) to determine true protein binding factors to our sequences.

It was also interesting to find that gypsy and its Su(Hw) binding site (twelve tandem sequences of the Su(Hw) consensus sequences) was not able to function as an insulator in our assay. The gypsy insulator has been shown to be functional in multiple *A. thaliana* studies. It is possible that this sequence is enhancer or promoter specific, as has been shown in numerous insulator studies (Gudynaite-Savitch et al., 2009; Nagaya et al., 2001). Or, gypsy may not be an enhancer-blocking

insulator in plants, as most studies have tested the gypsy sequence or gypsy-like sequences for positional effects only (Jiang et al., 2017; She et al., 2010a; Singer and Cox, 2013)

Although the genetic analysis of Chapter 3 identified an importance in multiple insulator binding sites, it is necessary to carry out future research on protein binding of these sites to determine the specific plant protein involved. As studies have shown, proteins will bind to consensus sequences, indicating possible variations of sites between species. The function of the sequences identified here suggests that plant insulation is conserved across species and proteins or their orthologs are conserved.

The results from this thesis have helped us propose potential models to explain the mechanisms at work. From the investigation of the human insulator, BEAD1c, specifically the BEADA site deletion analysis, we have identified this site to be critical in *A. thaliana.* The BLAST search on the CTCF protein identified similar zinc proteins present in *A. thaliana.* This highly conserved protein has been revealed to be critical in multiple insulator species and shown to interact with multiple nuclear substrates contributing to chromatin localization and organization in the nucleus, producing chromatin loops (Yang and Corces, 2011). Through the evidence presented here and the published literature on CTCF, we postulate that plant insulation may also follow the "chromatin loop model". This model suggests that insulators and their binding proteins (both primary and secondary) are responsible for creating physical loops with chromatin fiber, separating enhancers and promoters into independent domains. This separation blocks the two regulatory elements from communicating with one another. The next step in testing whether this concept is true, is to continue experimental research to identify the specific plant protein which binds to the BEADA site, according to Bell et al.'s (1999) method. After identification of the protein which results have

suggested to be CTCF, a knockdown of its nuclear associated proteins to determine whether insulator function becomes impaired would be necessary.

From the data obtained from UASrpg, we may expect to find multiple models involved in plant insulation. The deletion analysis pertaining to UAS_3'end, which resulted in comparable insulator activity to the full UASrpg sequence, suggested the possibility of multiple systems occurring in the insulator involved in function for the whole sequence and for the 3'end alone. This possibility requires multiple models to explain its function, and further research on insulator sequences and primary and secondary protein binding sites must be done to fully understand the mechanisms at work.

The results from Chapter 3 demonstrate that multiple protein binding sites are conserved in plants, providing a direction to identifying conserved models to explain and further test our insulator hypotheses.

4.2. Future directions

Chromatin insulators are an important class of DNA regulatory elements found in multiple species from fungi, to yeast, to humans. The overall role of insulators and their associated proteins in the genome of most eukaryotes is to organize the massive amounts of chromatin into structurally and functionally independent domains. Using what we know about insulators in flies and vertebrates, and the research presented here, we can expand our fundamental understanding of gene regulation and organization. Moreover, insulators can be valuable tools in biotechnology as they prove to function well in transgenics and successful at protecting against position effects.

In addition to genetic analyses, as the one here, the continuation of this research should identify specific protein factors associated with the novel insulators identified in Chapter 2, as well as ones

associated with the protein binding sites identified in Chapter 3. Well defined approaches to determine these proteins include Gel Shift Assay or Electrophoretic Mobility Shift Assay (EMSA) and Yeast One-Hybrid to discover associated primary proteins, and Yeast Two-Hybrid to discover secondary proteins.

EMSA is an *in vitro* approach to identify primary insulator binding proteins in nuclear extracts by a reduction in mobility of DNA and determining their migration patterns on polyacrylamide or agarose gels. Functional insulator sequences determined in our enhancer-blocking assay can be PCR-amplified with biotinylated primers and attached to streptavidin-beads for isolation of the binding protein(s). Eluted proteins even in mixtures can be identified (Desveaux et al., 2000) by reference to databases following electrospray mass spectrometry which gives the protein sequence of the generated fragments. Yeast One-Hybrid is another possible *in vivo* approach to identify primary insulator binding proteins where the bait is the insulator sequence (single or double stranded). An *A. thaliana* cDNA library will be created from leaf polyA + RNA as per the manufacturer's recommendation and putative positives will be sequenced. Advances in DNA sequencing allow the screening of a large number of clones.

Isolated candidates would be sequenced (DNA, Yeast 1-hybrid) or the sequence recovered form the peptide sequence (EMSA-assisted purification) and analysed for DNA binding motifs, not previously well-characterized etc. cDNAs for the best candidates will be expressed and used for antiserum production (Schnell et al., 2010) and GFP fusions will be made (Angers et al., 2006) to generate the tools for further analyses. EMSA will confirm the DNA binding ability with microscopy and immunochemistry used to localize the protein within the nucleus and compare to previous distributions. Identification of both insulator sequence and protein provides strong support for the evolutionary conservation of insulation.

The hypothesis that several proteins are involved in the function of insulators, as in Drosophila, would be tested by identifying secondary proteins associated with the primary proteins discovered in the above experiments (EMSA and Yeast One-Hybrid). The identification of interacting proteins would be attempted by using a yeast 2-hybrd system in which the bait becomes the insulator binding factor. The cDNA will be cloned using the ProQuest™ system and clones with putative interacting proteins sequenced. They will be analysed in a similar manner as for the 1-hybrid system. Promising candidates could be tested for interaction with purified insulator binding factor by a pull down assay that uses the antibody previously generated. The identification of proteins interacting with In-BF would further support the evolutionary hypothesis and link insulation to chromatin remodeling.

We believe the insulators we have found in this thesis are ideal sizes for transgenic experiments, as the genetic engineering of plants require elements to prevent inappropriate enhancer-promoter interactions or position effects should be relatively short sequences and easy to clone. With the integration of multiple genes, increasing the necessity to block misexpression, it is clear the future of agriculture and crop science relies on these special regulatory sequences.

Reference

Abhyankar, M. M., C. Urekar, and P. P. Reddi, 2007, A novel CpG-free vertebrate insulator silences the testis-specific SP-10 gene in somatic tissues: Journal of Biological Chemistry, v. 282, p. 36143-36154.

Adryan, B., G. Woerfel, I. Birch-Machin, S. Gao, M. Quick, L. Meadows, S. Russell, and R. White, 2007, Genomic mapping of Suppressor of Hairy-wing binding sites in Drosophila: Genome Biology, v. 8, p. 16.

Allen, G. C., G. Hall, S. Michalowski, W. Newman, S. Spiker, A. K. Weissinger, and W. F. Thompson, 1996, High-level transgene expression in plant cells: Effects of a strong scaffold attachment region from tobacco: Plant Cell, v. 8, p. 899-913.

Allen, G. C., G. E. Hall, L. C. Childs, A. K. Weissinger, S. Spiker, and W. F. Thompson, 1993, scaffold attachment regions increase reporter gene-expression in stably transformed plant-cells: Plant Cell, v. 5, p. 603-613.

Andersen, L., M. Kilstrup, and J. Neuhard, 1989, pyrimidine, purine and nitrogen control of cytosine deaminase synthesis in escherichia-coli-k12 - involvement of the glnlg and purr genes in the regulation of coda expression: Archives of Microbiology, v. 152, p. 115-118.

Angers, S., C. J. Thorpe, T. L. Biechele, S. J. Goldenberg, N. Zheng, M. J. MacCoss, and R. T. Moon, 2006, The KLHL12-Cullin-3 ubiquitin ligase negatively regulates the Wnt-beta-catenin pathway by targeting Dishevelled for degradation: Nature Cell Biology, v. 8, p. 348-U16.

Antes, T. J., S. J. Namciu, R. E. K. Fournier, and B. Levy-Wilson, 2001, The 5 ' boundary of the human apolipoprotein B chromatin domain in intestinal cells: Biochemistry, v. 40, p. 6731-6742.

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Y. Ren, W. W. Li, and W. S. Noble, 2009, MEME SUITE: tools for motif discovery and searching: Nucleic Acids Research, v. 37, p. W202-W208.

Barges, S., J. Mihaly, M. Galloni, K. Hagstrom, M. Muller, G. Shanower, P. Schedl, H. Gyurkovics, and F. Karch, 2000, The Fab-8 boundary defines the distal limit of the bithorax complex iab-7 domain and insulates iab-7 from initiation elements and a PRE in the adjacent iab-8 domain: Development, v. 127, p. 779-790.

Bell, A. C., and G. Felsenfeld, 2000, Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene: Nature, v. 405, p. 482-485.

Bell, A. C., A. G. West, and G. Felsenfeld, 1999, The protein CTCF is required for the enhancer blocking activity of vertebrate insulators: Cell, v. 98, p. 387-396.

Belozerov, V. E., P. Majumder, P. Shen, and H. N. Cai, 2003, A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of Drosophila: Embo Journal, v. 22, p. 3113-3121.

Bender, W., M. Akam, F. Karch, P. A. Beachy, M. Peifer, P. Spierer, E. B. Lewis, and D. S. Hogness, 1983, Molecular-Genetics of the Bithorax Complex in Drosophila-Melanogaster: Science, v. 221, p. 23-29.

Bi, X., and J. R. Broach, 1999, UAS(rpg) can function as a heterochromatin boundary element in yeast: Genes & Development, v. 13, p. 1089-1101.

Bickmore, W. A., 2013, The Spatial Organization of the Human Genome, *in* A. Chakravarti, and E. Green, eds., Annual Review of Genomics and Human Genetics, Vol 14: Annual Review of Genomics and Human Genetics, v. 14: Palo Alto, Annual Reviews, p. 67-84.

Breyne, P., M. Vanmontagu, A. Depicker, and G. Gheysen, 1992, characterization of a plant scaffold attachment region in a dna fragment that normalizes transgene expression in tobacco: Plant Cell, v. 4, p. 463-471.

Cai, H., and V. Levine, 1995, Modulation of Enhancer-Promoter Interactions by Insulators in the Drosophila Embryo: Nature, v. 376, p. 533-536.

Cai, H. N., and P. Shen, 2001, Effects of cis arrangement of chromatin insulators on enhancer-blocking activity: Science, v. 291, p. 493-495.

Carlos Roman, A., F. J. Gonzalez-Rico, E. Molto, H. Hernando, A. Neto, C. Vicente-Garcia, E. Ballestar, J. L. Gomez-Skarmeta, J. Vavrova-Anderson, R. J. White, L. Montoliu, and P. M. Fernandez-Salguero, 2011, Dioxin receptor and SLUG transcription factors regulate the insulator activity of B1 SINE retrotransposons via an RNA polymerase switch: Genome research, v. 21, p. 422-432.

Catarino, R. R., C. Neumayr, and A. Stark, 2017, Promoting transcription over long distances: Nature Genetics, v. 49, p. 972-973.

Chen, G. Q., and J.-T. Lin, 2012, A napin promoter activates gene expression in developing seeds of Lesquerella fendleri: *OnLine Journal of Biological Sciences*, v. 12, p. 4.

Cheng, C. L., G. N. Acedo, J. Dewdney, H. M. Goodman, and M. A. Conkling, 1991, differential expression of the 2 arabidopsis nitrate reductase genes: Plant Physiology, v. 96, p. 275-279.

Chetverina, D. A., P. V. Elizaryev, P. G. Georgiev, and M. M. Erokhin, 2013, 1A2 Insulator can interact with promoter of hsp70 gene in D-melanogaster: Russian Journal of Genetics, v. 49, p. 371-379.

Chung, J. H., M. Whiteley, and G. Felsenfeld, 1993, A 5' Element of the Chicken Beta-Globin Domain Serves as an Insulator in Human Erythroid-Cells and Protects Against Position Effect in Drosophila: Cell, v. 74, p. 505-514.

Ciabrelli, F., and G. Cavalli, 2015, Chromatin-Driven Behavior of Topologically Associating Domains: Journal of Molecular Biology, v. 427, p. 608-625.

Covey, S. N., G. P. Lomonossoff, and R. Hull, 1981, characterization of cauliflower mosaic-virus dna-sequences which encode major polyadenylated transcripts: Nucleic Acids Research, v. 9, p. 6735-6747.

Cremer, T., M. Cremer, B. Huebner, H. Strickfaden, D. Smeets, J. Popken, M. Sterr, Y. Markaki, K. Rippe, and C. Cremer, 2015, The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments: Febs Letters, v. 589, p. 2931-2943.

Dai, S., R. Carcamo, Z. Zhang, S. Chen, and R. N. Beachy, 2001, The bacterial cytosine deaminase gene used as a conditional negative selection marker in transgenic rice plants: Plant Cell Reports, v. 20, p. 738-743.

Desveaux, D, Després, C, Joyeux, A, Rajagopal Subramaniam, R & Brisson, N 2000 PBF-2 Is a Novel Single-Stranded DNA Binding Factor Implicated in PR-10a Gene Activation in Potato. The Plant Cell 12: 1477–1489

Devi, S., D. Kumar, N. Ansari, and A. Sivasankar, 2010, Studies on the expression pattern of seed-specific napin promoter (BcNAI) in transgenic (Nicotiana tabacum L.) tobacco: International Journal of Environmental Science and Development, v. 1, p. 4.

Donze, D., C. R. Adams, J. Rine, and R. T. Kamakaka, 1999, The boundaries of the silenced HMR domain in Saccharomyces cerevisiae: Genes & development, v. 13, p. 698-708.

Donze, D., and R. T. Kamakaka, 2001, RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae: Embo Journal, v. 20, p. 520-531.

Dorsett, D., 1990, potentiation of a polyadenylylation site by a downstream protein dna interaction: Proceedings of the National Academy of Sciences of the United States of America, v. 87, p. 4373-4377.

Dubeau, M. P., M. G. Ghinet, P. E. Jacques, N. Clermont, C. Beaulieu, and R. Brzezinski, 2009, Cytosine Deaminase as a Negative Selection Marker for Gene Disruption and Replacement in the Genus Streptomyces and Other Actinobacteria: Applied and Environmental Microbiology, v. 75, p. 1211-1214.

Ericson, M. L., E. Muren, H. O. Gustavsson, L. G. Josefsson, and L. Rask, 1991, analysis of the promoter region of napin genes from brassica-napus demonstrates binding of nuclear-protein invitro to a conserved sequence motif: European Journal of Biochemistry, v. 197, p. 741-746.

Filippova, G. N., S. Fagerlie, E. M. Klenova, C. Myers, Y. Dehner, G. Goodwin, P. E. Neiman, S. J. Collins, and V. V. Lobanenkov, 1996, An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes: Molecular and cellular biology, v. 16, p. 2802-2813.

Forsbach, A., D. Schubert, B. Lechtenberg, M. Gils, and R. Schmidt, 2003, A comprehensive characterization of single-copy T-DNA insertions in the Arabidopsis thaliana genome: Plant Molecular Biology, v. 52, p. 161-176.

Fourel, G., E. Revardel, C. E. Koering, and E. Gilson, 1999, Cohabitation of insulators and silencing elements in yeast subtelomeric regions: Embo Journal, v. 18, p. 2522-2537.

Gan, S. and Xie, M. (2002) Genetic insulator for preventing influence by another gene promoter. USA patent # 20060174370.

Gaszner, M., and G. Felsenfeld, 2006, Insulators: exploiting transcriptional and epigenetic mechanisms: Nature Reviews Genetics, v. 7, p. 703-713.

Gaszner, M., J. Vazquez, and P. Schedl, 1999, The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction: Genes & Development, v. 13, p. 2098-2107.

Geyer, P. K., M. M. Green, and V. G. Corces, 1988, Reversion of a gypsy-induced mutation at the yellow (y) locus of drosophila-melanogaster is associated with the insertion of a newly defined transposable element: Proceedings of the National Academy of Sciences of the United States of America, v. 85, p. 3938-3942.

Geyer, P. K., C. Spana, and V. G. Corces, 1986, On the molecular mechanism of gypsy-induced mutations at the yellow locus of drosophila-melanogaster: Embo Journal, v. 5, p. 2657-2662.

Golovnin, A., I. Birukova, O. Romanova, M. Silicheva, A. Parshikov, E. Savitskaya, V. Pirrotta, and P. Georgiev, 2003, An endogenous Su(Hw) insulator separates the yellow gene from the Achaete-scute gene complex in Drosophila: Development, v. 130, p. 3249-3258.

Graham, I. R., and A. Chambers, 1994, use of a selection technique to identify the diversity of binding-sites for the yeast rap1 transcription factor: Nucleic Acids Research, v. 22, p. 124-130.

Graham, I. R., R. A. Haw, K. G. Spink, K. A. Halden, and A. Chambers, 1999, In vivo analysis of functional regions within yeast Rap1p: Molecular and Cellular Biology, v. 19, p. 7481-7490.

Gruzdeva, N., O. Kyrchanova, A. Parshikov, A. Kullyev, and P. Georgiev, 2005, The Mcp element from the bithorax complex contains an insulator that is capable of pairwise interactions and can facilitate enhancer-promoter communication: Molecular and Cellular Biology, v. 25, p. 3682-3689.

Gudynaite-Savitch, L., D. A. Johnson, and B. L. A. Miki, 2009, Strategies to mitigate transgene-promoter interactions: Plant Biotechnology Journal, v. 7, p. 472-485.

Gyurkovics, H., J. Gausz, J. Kummer, and F. Karch, 1990, A new homeotic mutation in the drosophila bithorax complex removes a boundary separating 2 domains of regulation: Embo Journal, v. 9, p. 2579-2585.

Hagstrom, K., M. Muller, and P. Schedl, 1996, Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the Drosophila bithorax complex: Genes & Development, v. 10, p. 3202-3215.

Hark, A. T., C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. Levorse, and S. M. Tilghman, 2000, CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus: Nature, v. 405, p. 486-489.

Higo, K., Y. Ugawa, M. Iwamoto, and T. Korenaga, 1999, Plant cis-acting regulatory DNA elements (PLACE) database: 1999: Nucleic Acids Research, v. 27, p. 297-300.

Hily, J.-M., S. D. Singer, Y. Yang, and Z. Liu, 2009, A transformation booster sequence (TBS) from Petunia hybrida functions as an enhancer-blocking insulator in Arabidopsis thaliana: Plant Cell Reports, v. 28, p. 1095-1104.

Holohan, E. E., C. Kwong, B. Adryan, M. Bartkuhn, M. Herold, R. Renkawitz, S. Russell, and R. White, 2007, CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex: Plos Genetics, v. 3, p. 1211-1222.

Huang, S. C., H. Z. Liu, G. H. He, and F. G. Yu, 2007, An improved method to identify the T-DNA insertion site in transgenic Arabidopsis thaliana genome: Russian Journal of Plant Physiology, v. 54, p. 822-826.

Huet, J., and A. Sentenac, 1987, Tuf, the Yeast Dna-Binding Factor Specific for Uasrpg Upstream Activating Sequences - Identification of the Protein and its Dna-Binding Domain: Proceedings of the National Academy of Sciences of the United States of America, v. 84, p. 3648-3652.

Jefferson, R. A., T. A. Kavanagh, and M. W. Bevan, 1987, beta-glucuronidase (gus) as a sensitive and versatile gene fusion marker in plants: Journal of Cellular Biochemistry, p. 57-57.

Jiang, W., L. Sun, X. Yang, M. Wang, N. Esmaeili, N. Pehlivan, R. Zhao, H. Zhang, and Y. Zhao, 2017, The effects of Transcription Directions of Transgenes and the gypsy Insulators on the Transcript Levels of Transgenes in Transgenic Arabidopsis: Scientific Reports, v. 7.

Karch, F., B. Weiffenbach, M. Peifer, W. Bender, I. Duncan, S. Celniker, M. Crosby, and E. B. Lewis, 1985, the abdominal region of the bithorax complex: Cell, v. 43, p. 81-96.

Kellum, R., and P. Schedl, 1991, A Position-Effect Assay for Boundaries of Higher-Order Chromosomal Domains: Cell, v. 64, p. 941-950.

Kellum, R., and P. Schedl, 1992, A Group of Scs Elements Function as Domain Boundaries in an Enhancer-Blocking Assay: Molecular and cellular biology, v. 12, p. 2424-2431.

Khan, A., O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Chèneby, S. R. Kulkarni, G. Tan, D. Baranasic, D. J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W. W. Wasserman, F. Parcy, and A. Mathelier, 2017, JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework.: Nucleic Acids Research.

Kim, S. I., Veena, and S. B. Gelvin, 2007, Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions: Plant Journal, v. 51, p. 779-791.

Kyrchanova, O., S. Toshchakov, A. Parshikov, and P. Georgiev, 2007, Study of the functional interaction between Mcp insulators from the Drosophila bithorax complex: Effects of insulator pairing on enhancer-promoter communication: Molecular and Cellular Biology, v. 27, p. 3035-3043.

Lescot, M., P. Dehais, G. Thijs, K. Marchal, Y. Moreau, Y. Van de Peer, P. Rouze, and S. Rombauts, 2002, PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences: Nucleic Acids Research, v. 30, p. 325-327.

Li, X. G., Z. X. Lu, L. Chen, G. F. Xiao, and Z. Zhu, 2001, The influence of matrix attachment regions on transgene expression in transgenic tobaccos: Acta Botanica Sinica, v. 43, p. 405-408.

Lunyak, V. V., G. G. Prefontaine, E. Nunez, T. Cramer, B.-G. Ju, K. A. Ohgi, K. Hutt, R. Roy, A. Garcia-Diaz, X. Zhu, Y. Yung, L. Montoliu, C. K. Glass, and M. G. Rosenfeld, 2007, Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis: Science, v. 317, p. 248-251.

Majumder, P., and H. N. Cai, 2003, The functional analysis of insulator interactions in the Drosophila embryo: Proceedings of the National Academy of Sciences of the United States of America, v. 100, p. 5223-5228.

Malik, K., K. Wu, X. Q. Li, T. Martin-Heller, M. Hu, E. Foster, L. Tian, C. Wang, K. Ward, M. Jordan, D. Brown, S. Gleddie, D. Simmonds, S. Zheng, J. Simmonds, and B. Miki, 2002, A constitutive gene expression system derived from the tCUP cryptic promoter elements: Theoretical and Applied Genetics, v. 105, p. 505-514.

Markstein, M., C. Pitsouli, C. Villalta, S. E. Celniker, and N. Perrimon, 2008, Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes: Nature Genetics, v. 40, p. 476-483.

Matharu, N. K., and S. H. Ahanger, 2015, Chromatin Insulators and Topological Domains: Adding New Dimensions to 3D Genome Architecture: Genes, v. 6, p. 790-811.

McKinney, E. C., N. Ali, A. Traut, K. A. Feldmann, D. A. Belostotsky, J. M. McDowell, and R. B. Meagher, 1995, sequence-based identification of t-dna insertion mutations in arabidopsis - actin mutants act2-1 and act4-1: Plant Journal, v. 8, p. 613-622.

Mirkovitch, J., M. E. Mirault, and U. K. Laemmli, 1984, organization of the higher-order chromatin loop - specific dna attachment sites on nuclear scaffold: Cell, v. 39, p. 223-232.

Misteli, T., 2007, Beyond the sequence: Cellular organization of genome function: Cell, v. 128, p. 787-800.

Mlynarova, L., A. Loonen, J. Heldens, R. C. Jansen, P. Keizer, W. J. Stiekema, and J. P. Nap, 1994, reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region: Plant Cell, v. 6, p. 417-426.

Moon, H., G. Filippova, D. Loukinov, E. Pugacheva, Q. Chen, S. T. Smith, A. Munhall, B. Grewe, M. Bartkuhn, R. Arnold, L. J. Burke, R. Renkawitz-Pohl, R. Ohlsson, J. M. Zhou, R. Renkawitz, and V. Lobanenkov, 2005, CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator: Embo Reports, v. 6, p. 165-170.

Mullen, C. A., M. Kilstrup, and R. M. Blaese, 1992, transfer of the bacterial gene for cytosine deaminase to mammalian-cells confers lethal sensitivity to 5-fluorocytosine - a negative selection system: Proceedings of the National Academy of Sciences of the United States of America, v. 89, p. 33-37.

Nagaya, S., K. Yoshida, K. Kato, K. Akasaka, and A. Shinmyo, 2001, An insulator element from the sea urchin Hemicentrotus pulcherrimus suppresses variation in transgene expression in cultured tobacco cells: Molecular Genetics and Genomics, v. 265, p. 405-413.

Parkhurst, S. M., D. A. Harrison, M. P. Remington, C. Spana, R. L. Kelley, R. S. Coyne, and V. G. Corces, 1988, the drosophila su(hw) gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative dna-binding protein: Genes & Development, v. 2, p. 1205-1215.

Perera, R. J., C. G. Linard, and E. R. Signer, 1993, cytosine deaminase as a negative selective marker for arabidopsis: Plant Molecular Biology, v. 23, p. 793-799.

Raab, J. R., and R. T. Kamakaka, 2010, OPINION Insulators and promoters: closer than we think: Nature Reviews Genetics, v. 11, p. 439-446.

Recillas-Targa, F., M. J. Pikaart, B. Burgess-Beusse, A. C. Bell, M. D. Litt, A. G. West, M. Gaszner, and G. Felsenfeld, 2002, Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities: Proceedings of the National Academy of Sciences of the United States of America, v. 99, p. 6883-6888.

Robinett, C. C., A. Oconnor, and M. Dunaway, 1997, The repeat organizer, a specialized insulator element within the intergenic spacer of the Xenopus rRNA genes: Molecular and Cellular Biology, v. 17, p. 2866-2875.

Schlaman, H. R. M., and P. J. J. Hooykaas, 1997, Effectiveness of the bacterial gene codA encoding cytosine deaminase as a negative selectable marker in Agrobacterium-mediated plant transformation: Plant Journal, v. 11, p. 1377-1385.

Schnell, J. A., S. Y. Han, B. L. Miki, and D. A. Johnson, 2010, Soybean peroxidase propeptides are functional signal peptides and increase the yield of a foreign protein: Plant Cell Reports, v. 29, p. 987-996.

Schoffl, F., G. Schroder, M. Kliem, and M. Rieping, 1993, an sar-sequence containing 395 bp-dna fragment mediates enhanced, gene-dosage-correlated expression of a chimeric heat-shock gene in transgenic tobacco plants: Transgenic Research, v. 2, p. 93-100.

Schweinsberg, S., K. Hagstrom, D. Gohl, P. Schedl, R. P. Kumar, R. Mishra, and F. Karch, 2004, The enhancer-blocking activity of the Fab-7 boundary from the Drosophila bithorax complex requires GAGA-factor-binding sites: Genetics, v. 168, p. 1371-1384.

Scott, K. C., A. D. Taubman, and P. K. Geyer, 1999, Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength: Genetics, v. 153, p. 787-798.

Shao, M., J. M. Michno, S. K. Hotton, A. Blechl, and J. Thomson, 2015, A bacterial gene codA encoding cytosine deaminase is an effective conditional negative selectable marker in Glycine max: Plant Cell Reports, v. 34, p. 1707-1716.

She, W., W. Lin, Y. Zhu, Y. Chen, W. Jin, Y. Yang, N. Han, H. Bian, M. Zhu, and J. Wang, 2010a, The gypsy Insulator of Drosophila melanogaster, Together With Its Binding Protein Suppressor of Hairy-Wing, Facilitate High and Precise Expression of Transgenes in Arabidopsis thaliana: Genetics, v. 185, p. 1141-U27.

She, W. J., W. Q. Lin, Y. B. Zhu, Y. Chen, W. Y. Jin, Y. J. Yang, N. Han, H. W. Bian, M. Y. Zhu, and J. H. Wang, 2010b, The gypsy Insulator of Drosophila melanogaster, Together With Its Binding Protein Suppressor of Hairy-Wing, Facilitate High and Precise Expression of Transgenes in Arabidopsis thaliana: Genetics, v. 185, p. 1141-U27.

Shen, Y., F. Yue, D. F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov, and B. Ren, 2012, A map of the cis-regulatory sequences in the mouse genome: Nature, v. 488, p. 116-120.

Singer, S. D., and K. D. Cox, 2013, A gypsy-like sequence from Arabidopsis thaliana exhibits enhancer-blocking activity in transgenic plants: Journal of Plant Biochemistry and Biotechnology, v. 22, p. 35-42.

Singer, S. D., J.-M. Hily, and K. D. Cox, 2011, Analysis of the enhancer-blocking function of the TBS element from Petunia hybrida in transgenic Arabidopsis thaliana and Nicotiana tabacum: Plant Cell Reports, v. 30, p. 2013-2025.

Singer, S. D., J.-M. Hily, and Z. Liu, 2010, A 1-kb Bacteriophage Lambda Fragment Functions as an Insulator to Effectively Block Enhancer-Promoter Interactions in Arabidopsis thaliana: Plant Molecular Biology Reporter, v. 28, p. 69-76.

Smith, P. A., and V. G. Corces, 1992, The suppressor of hairy-wing binding region is required for gypsy mutagenesis: Molecular & General Genetics, v. 233, p. 65-70.

Spana, C., D. A. Harrison, and V. G. Corces, 1988, the drosophila-melanogaster-suppressor of hairy-wing protein binds to specific sequences of the gypsy retrotransposon: Genes & Development, v. 2, p. 1414-1423.

Steiner, S., and P. Philippsen, 1994, sequence and promoter analysis of the highly expressed tef gene of the filamentous fungus ashbya-gossypii: Molecular & General Genetics, v. 242, p. 263-271.

Stougaard, J., 1993, substrate-dependent negative selection in plants using a bacterial cytosine deaminase gene: Plant Journal, v. 3, p. 755-761.

Sultana, H., S. Verma, and R. K. Mishra, 2011, A BEAF dependent chromatin domain boundary separates myoglianin and eyeless genes of Drosophila melanogaster: Nucleic Acids Research, v. 39, p. 3543-3557.

Tchurikov, N. A., O. V. Kretova, E. D. Moiseeva, and D. V. Sosin, 2009, Evidence for RNA synthesis in the intergenic region between enhancer and promoter and its inhibition by insulators in Drosophila melanogaster: Nucleic Acids Research, v. 37, p. 111-122.

Udvardy, A., E. Maine, and P. Schedl, 1985, The 87A7 chromomere: Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains: Journal of Molecular Biology, v. 185, p. 341-358.

Vain, P., B. Worland, A. Kohli, J. W. Snape, P. Christou, G. C. Allen, and W. F. Thompson, 1999, Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny: Plant Journal, v. 18, p. 233-242.

Vandergeest, A. H. M., G. E. Hall, S. Spiker, and T. C. Hall, 1994, the beta-phaseolin gene is flanked by matrix attachment regions: Plant Journal, v. 6, p. 413-423.

Vazquez, J., and P. Schedl, 1994, sequences required for enhancer blocking activity of scs are located within 2 nuclease-hypersensitive regions: Embo Journal, v. 13, p. 5984-5993.

Wang, J., C. Vicente-Garcia, D. Seruggia, E. Molto, A. Fernandez-Minan, A. Neto, E. Lee, J. L. Gomez-Skarmeta, L. Montoliu, V. V. Lunyak, and I. K. Jordan, 2015, MIR retrotransposon sequences provide insulators to the human genome: Proceedings of the National Academy of Sciences of the United States of America, v. 112, p. E4428-E4437.

West, A. G., M. Gaszner, and G. Felsenfeld, 2002, Insulators: many functions, many mechanisms: Genes & development, v. 16, p. 271-288.

Xue, H., Y. T. Yang, C. A. Wu, G. D. Yang, M. M. Zhang, and C. C. Zheng, 2005, TM2, a novel strong matrix attachment region isolated from tobacco, increases transgene expression in transgenic rice calli and plants: Theoretical and Applied Genetics, v. 110, p. 620-627.

Yang, J. P., and V. G. Corces, 2011, Chromatin Insulators: A Role in Nuclear Organization and Gene Expression: Advances in Cancer Research, Vol 110, v. 110, p. 43-76.

Yu, Q., R. X. Qiu, T. B. Foland, D. Griesen, C. S. Galloway, Y. H. Chiu, J. Sandmeier, J. R. Broach, and X. Bi, 2003, Rap1p and other transcriptional regulators can function in defining distinct domains of gene expression: Nucleic Acids Research, v. 31, p. 1224-1233.

Yusufzai, T. M., H. Tagami, Y. Nakatani, and G. Felsenfeld, 2004, CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species: Molecular Cell, v. 13, p. 291-298.

Zhao, K., C. M. Hart, and U. K. Laemmli, 1995, visualization of chromosomal domains with boundary element-associated factor BEAF-32: Cell, v. 81, p. 879-889.

Zhong, X.-P., and M. S. Krangel, 1997, An enhancer-blocking element between alpha and delta gene segments within the human T cell receptor alpha/delta locus: Proceedings of the National Academy of Sciences of the United States of America, v. 94, p. 5219-5224.

Zhou, J. M., S. Barolo, P. Szymanski, and M. Levine, 1996, The Fab-7 element of the bithorax complex attenuates enhancer-promoter interactions in the Drosophila embryo: Genes & Development, v. 10, p. 3195-3201.

Appendix – Chapter 2

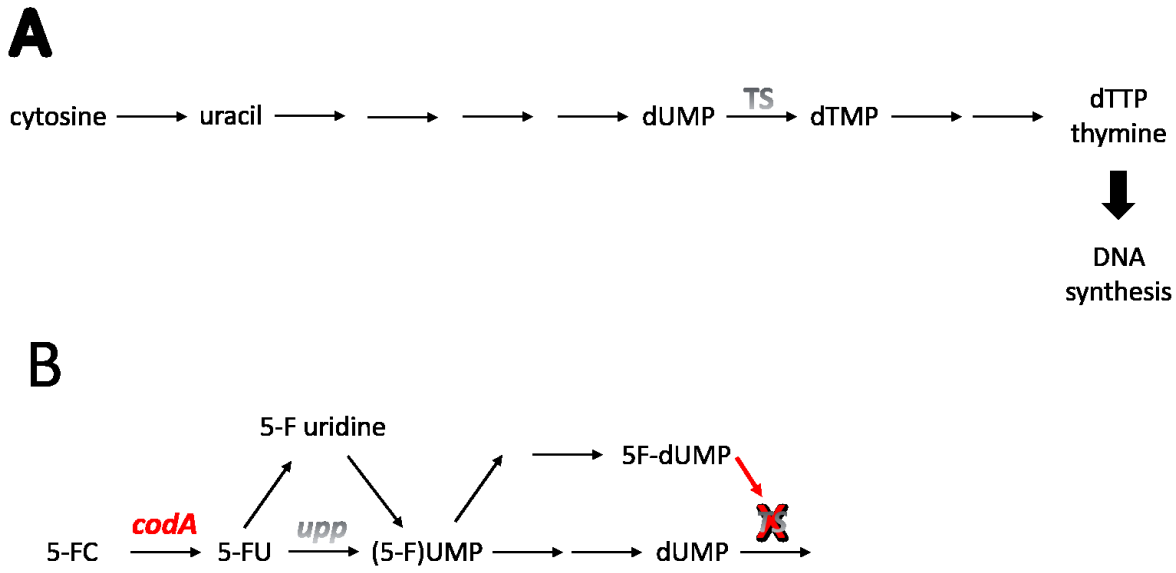


Figure A2.1: The toxic metabolic pathway of 5-Flurocytosine (5-FC) to 5-Fluorouracil (5-FU) catalyzed by the product of *coda* gene expression. (A) Normally cytosine is metabolized to uracil and eventually to deoxythymine triphosphate (dTTP) critical for DNA synthesis. (B) Cytosine deaminase enzyme encoded by *codA* converts 5-FC to 5-FU, which is processed to 5-fluoro-deoxyuridine monophosphate (5F-dUMP) indirectly via the intermediate 5-F uridine or directly by uracil phosphoribosyltransferase (upp). 5F-dUMP irreversibly inhibits thymidylate synthase (TS) activity, and as a result the cells are deprived of deoxythymidine triphosphate (dTTP) necessary for DNA synthesis. Modified from http://aac.asm.org/content/47/4/1275/F3.expansion.html.

DNA Analysis of *Arabidopsis thaliana* transformed with InI-6pB31 sequence

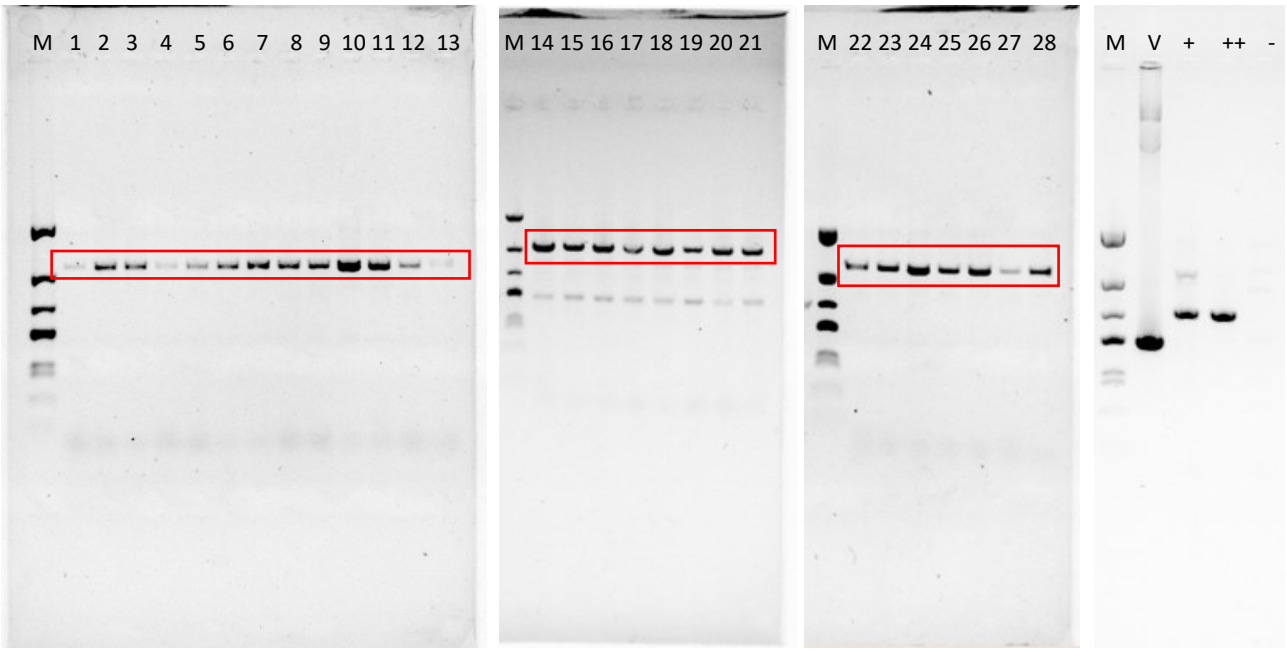


Figure A2.2: Example of agarose gel electrophoresis showing PCR amplification of InI-6 in pB31 with 1381F and GUS5'Rev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate insulator InI-6pB31 are shown. Amplification of DNA used 1300LacZF and GUS5'Rev primers with an expected band size of 888bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80

1-28: Transgenic plant DNA containing InI-6pB31 inserts (888bp)

V : pB31 control vector (no insert, 450bp)

+ : II-10 in pB31 sequenced control (604bp)

++ : I-3Δ in pB31 sequenced control (592bp)

- : Negative control containing water in replace of DNA

Alignment of "InI-6pB31" samples and consensus sequence output from Sequencher 4.10.1

```
<Input_InI-6pB31    #1              GGATCCGAGA CAAGCCGCGC CCAAAGACCT GGCCGGAATA AGGGGGGCGC
>InI-6pB31_S2       #1              GGATCCGAGA CAAGCCGCGC CCAAAGACCT GGCCGGAATA AGGGGGGCGC
>InI-6pB31_S7       #1              GGATCCGAGA CAAGCCGCGC CCAAAGACCT GGCCGGAATA AGGGGGGCGC
>InI-6pB31_S8       #1              GGATCCGAGA CAAGCCGCGC CCAAAGACCT GGCCGGAATA AGGGGGGCGC
>InI-6pB31_S9       #1              GGATCCGAGA CAAGCCGCGC CCAAAGACCT GGCCGGAATA AGGGGGGCGC
>InI-6pB31_S10      #1              GGATCCGAGA CAAGCCGCGC CCAAAGACCT GGCCGGAATA AGGGGGGCGC
                                    ..................................................
 InI-6pB31          #1              GGATCCGAGA CAAGCCGCGC CCAAAGACCT GGCCGGAATA AGGGGGGCGC
 Consensus

<Input_InI-6pB31    #51             TTGTGATCAC ACGGGGAGGC GCAGGGACAC GAGAAGCGGC AGGGGGTCGT
>InI-6pB31_S2       #51             TTGTGATCAC ACGGGGAGGC GCAGGGACAC GAGAAGCGGC AGGGGGTCGT
>InI-6pB31_S7       #51             TTGTGATCAC ACGGGGAGGC GCAGGGACAC GAGAAGCGGC AGGGGGTCGT
>InI-6pB31_S8       #51             TTGTGATCAC ACGGGGAGGC GCAGGGACAC GAGAAGCGGC AGGGGGTCGT
>InI-6pB31_S9       #51             TTGTGATCAC ACGGGGAGGC GCAGGGACAC GAGAAGCGGC AGGGGGTCGT
>InI-6pB31_S10      #51             TTGTGATCAC ACGGGGAGGC GCAGGGACAC GAGAAGCGGC AGGGGGTCGT
                                    ..................................................
 InI-6pB31          #51             TTGTGATCAC ACGGGGAGGC GCAGGGACAC GAGAAGCGGC AGGGGGTCGT


<Input_InI-6pB31    #101            AACTGGGGAA AGGCCGGGTG GGGGACAGCG GTGGGTATGC CTCCTCCTGA
>InI-6pB31_S2       #101            AACTGGGGAA AGGCCGGGTG GGGGACAGCG GTGGGTATGC CTCCTCCTGA
>InI-6pB31_S7       #101            AACTGGGGAA AGGCCGGGTG GGGGACAGCG GTGGGTATGC CTCCTCCTGA
>InI-6pB31_S8       #101            AACTGGGGAA AGGCCGGGTG GGGGACAGCG GTGGGTATGC CTCCTCCTGA
>InI-6pB31_S9       #101            AACTGGGGAA AGGCCGGGTG GGGGACAGCG GTGGGTATGC CTCCTCCTGA
>InI-6pB31_S10      #101            AACTGGGGAA AGGCCGGGTG GGGGACAGCG GTGGGTATGC CTCCTCCTGA
                                    ..................................................
 InI-6pB31          #101            AACTGGGGAA AGGCCGGGTG GGGGACAGCG GTGGGTATGC CTCCTCCTGA


<Input_InI-6pB31    #151            ATTCAGGAGG AGGTTGCACG CGCATCCCGA ATGAACCGCT AACCTCTCCT
>InI-6pB31_S2       #151            ATTCAGGAGG AGGTTGCACG CGCATCCCGA ATGAACCGCT AACCTCTCCT
>InI-6pB31_S7       #151            ATTCAGGAGG AGGTTGCACG CGCATCCCGA ATGAACCGCT AACCTCTCCT
>InI-6pB31_S8       #151            ATTCAGGAGG AGGTTGCACG CGCATCCCGA ATGAACCGCT AACCTCTCCT
>InI-6pB31_S9       #151            ATTCAGGAGG AGGTTGCACG CGCATCCCGA ATGAACCGCT AACCTCTCCT
>InI-6pB31_S10      #151            ATTCAGGAGG AGGTTGCACG CGCATCCCGA ATGAACCGCT AACCTCTCCT
```

```
                                                  ..................................................
 InI-6pB31          #151                          ATTCAGGAGG AGGTTGCACG CGCATCCCGA ATGAACCGCT AACCTCTCCT


<Input_InI-6pB31    #201                          TCCTTGTTTT CTTCCCACTC TTCGCCTCCG CCTTGCAGCT AATTCACTTT
>InI-6pB31_S2       #201                          TCCTTGTTTT CTTCCCACTC TTCGCCTCCG CCTTGCAGCT AATTCACTTT
>InI-6pB31_S7       #201                          TCCTTGTTTT CTTCCCACTC TTCGCCTCCG CCTTGCAGCT AATTCACTTT
>InI-6pB31_S8       #201                          TCCTTGTTTT CTTCCCACTC TTCGCCTCCG CCTTGCAGCT AATTCACTTT
>InI-6pB31_S9       #201                          TCCTTGTTTT CTTCCCACTC TTCGCCTCCG CCTTGCAGCT AATTCACTTT
>InI-6pB31_S10      #201                          TCCTTGTTTT CTTCCCACTC TTCGCCTCCG CCTTGCAGCT AATTCACTTT
                                                  ..................................................
 InI-6pB31          #201                          TCCTTGTTTT CTTCCCACTC TTCGCCTCCG CCTTGCAGCT AATTCACTTT


<Input_InI-6pB31    #251                          ACTGCTTTAT TTGTACACAT AATTTGCTTG TCTCGGATCC GAGACAAGCG
>InI-6pB31_S2       #251                          ACTGCTTTAT TTGTACACAT AATTTGCTTG TCTCGGATCC GAGACAAGCG
>InI-6pB31_S7       #251                          ACTGCTTTAT TTGTACACAT AATTTGCTTG TCTCGGATCC GAGACAAGCG
>InI-6pB31_S8       #251                          ACTGCTTTAT TTGTACACAT AATTTGCTTG TCTCGGATCC GAGACAAGCG
>InI-6pB31_S9       #251                          ACTGCTTTAT TTGTACACAT AATTTGCTTG TCTCGGATCC GAGACAAGCG
>InI-6pB31_S10      #251                          ACTGCTTTAT TTGTACACAT AATTTGCTTG TCTCGGATCC GAGACAAGCG
                                                  ..................................................
 InI-6pB31          #251                          ACTGCTTTAT TTGTACACAT AATTTGCTTG TCTCGGATCC GAGACAAGCG


<Input_InI-6pB31    #301                          GGAGGTTGGG GAGCCATGAG CGCCATGAAC GCGGAAACGG CGGTGAGAAG
>InI-6pB31_S2       #301                          GGAGGTTGGG GAGCCATGAG CGCCATGAAC GCGGAAACGG CGGTGAGAAG
>InI-6pB31_S7       #301                          GGAGGTTGGG GAGCCATGAG CGCCATGAAC GCGGAAACGG CGGTGAGAAG
>InI-6pB31_S8       #301                          GGAGGTTGGG GAGCCATGAG CGCCATGAAC GCGGAAACGG CGGTGAGAAG
>InI-6pB31_S9       #301                          GGAGGTTGGG GAGCCATGAG CGCCATGAAC GCGGAAACGG CGGTGAGAAG
>InI-6pB31_S10      #301                          GGAGGTTGGG GAGCCATGAG CGCCATGAAC GCGGAAACGG CGGTGAGAAG
                                                  ..................................................
 InI-6pB31          #301                          GGAGGTTGGG GAGCCATGAG CGCCATGAAC GCGGAAACGG CGGTGAGAAG


<Input_InI-6pB31    #351                          ATCGGAGTGG TAGGTGTATA CGGAGTTGTG CAGCTGGATG ACAGAAAAGG
>InI-6pB31_S2       #351                          ATCGGAGTGG TAGGTGTATA CGGAGTTGTG CAGCTGGATG ACAGAAAAGG
>InI-6pB31_S7       #351                          ATCGGAGTGG TAGGTGTATA CGGAGTTGTG CAGCTGGATG ACAGAAAAGG
>InI-6pB31_S8       #351                          ATCGGAGTGG TAGGTGTATA CGGAGTTGTG CAGCTGGATG ACAGAAAAGG
>InI-6pB31_S9       #351                          ATCGGAGTGG TAGGTGTATA CGGAGTTGTG CAGCTGGATG ACAGAAAAGG
```

```
>InI-6pB31_S10      #351              ATCGGAGTGG TAGGTGTATA CGGAGTTGTG CAGCTGGATG ACAGAAAAGG
                                      ..................................................
 InI-6pB31          #351              ATCGGAGTGG TAGGTGTATA CGGAGTTGTG CAGCTGGATG ACAGAAAAGG


<Input_InI-6pB31    #401              GAGGGGTATT AGAGCGTAAA GGGCCTCCTC CTGAATTC
>InI-6pB31_S2       #401              GAGGGGTATT AGAGCGTAAA GGGCCTCCTC CTGAATTC
>InI-6pB31_S7       #401              GAGGGGTATT AGAGCGTAAA GGGCCTCCTC CTGAATTC
>InI-6pB31_S8       #401              GAGGGGTATT AGAGCGTAAA GGGCCTCCTC CTGAATTC
>InI-6pB31_S9       #401              GAGGGGTATT AGAGCGTAAA GGGCCTCCTC CTGAATTC
>InI-6pB31_S10      #401              GAGGGGTATT AGAGCGTAAA GGGCCTCCTC CTGAATTC
                                      ................................................
 InI-6pB31          #401              GAGGGGTATT AGAGCGTAAA GGGCCTCCTC CTGAATTC
```

DNA Analysis of *Arabidopsis thaliana* transformed with InII-10pB31 sequence

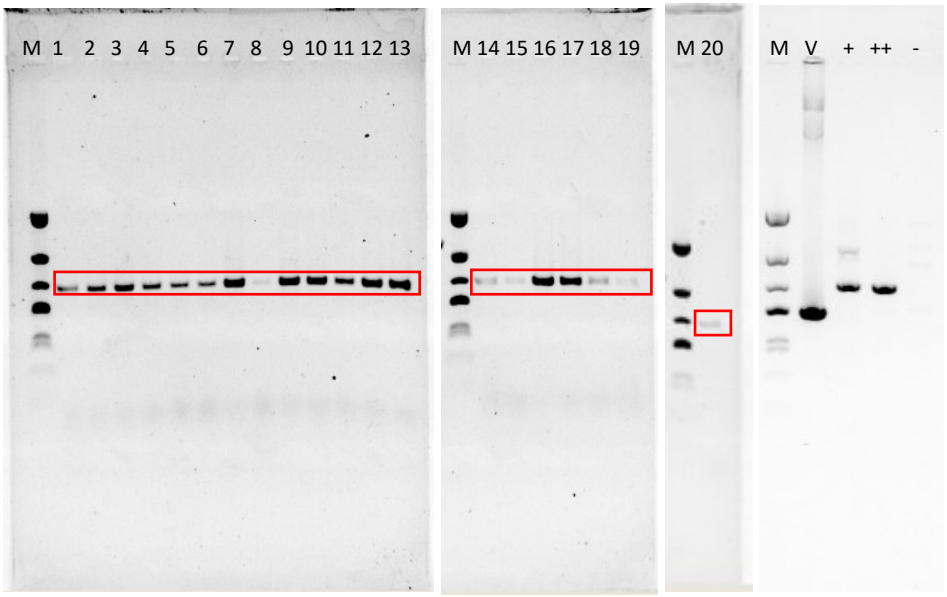


Figure A2.3: Example of agarose gel electrophoresis showing PCR amplification of InII-10 in pB31 with 1381F and GUS5'Rev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate insulator InII-10pB31 are shown. Amplification of DNA used 1300LacZF and GUS5'Rev primers with an expected band size of 604bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80

1-20: Transgenic plant DNA containing InII-10pB31 inserts (604bp)

V : pB31 control vector (no insert, 450bp)

+ : II-10 in pB31 sequenced control (604bp)

++ : I-3Δ in pB31 sequenced control (587bp)

- : Negative control containing water in replace of DNA

Alignment of "InII-10pB31" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_InII-10pB31  #1                GAATTCAGGA GGAGGCATAT GTCCCCATTG ACCCCTCCGC CCCTTTCCGC
<II-10pB31_S3       #1                GAATTCAGGA GGAGGCATAT GTCCCCATTG ACCCCTCCGC CCCTTTCCGC
<II-10pB31_S4       #1                GAATTCAGGA GGAGGCATAT GTCCCCATTG ACCCCTCCGC CCCTTTCCGC
<II-10pB31_S8       #1                GAATTCAGGA GGAGGCATAT GTCCCCATTG ACCCCTCCGC CCCTTTCCGC
<II-10pB31_S9       #1                GAATTCAGGA GGAGGCATAT GTCCCCATTG ACCCCTCCGC CCCTTTCCGC
<II-10pB31_S11      #1                GAATTCAGGA GGAGGCATAT GTCCCCATTG ACCCCTCCGC CCCTTTCCGC
                                      ..................................................
 InII-10pB31        #1                GAATTCAGGA GGAGGCATAT GTCCCCATTG ACCCCTCCGC CCCTTTCCGC
 Consensus

>Input_InII-10pB31  #51               TGGATCTACC ACCTCCGCCC TAGCCCCCGT GTATCGGCCT TTCGAATGTG
<II-10pB31_S3       #51               TGGATCTACC ACCTCCGCCC TAGCCCCCGT GTATCGGCCT TTCGAATGTG
<II-10pB31_S4       #51               TGGATCTACC ACCTCCGCCC TAGCCCCCGT GTATCGGCCT TTCGAATGTG
<II-10pB31_S8       #51               TGGATCTACC ACCTCCGCCC TAGCCCCCGT GTATCGGCCT TTCGAATGTG
<II-10pB31_S9       #51               TGGATCTACC ACCTCCGCCC TAGCCCCCGT GTATCGGCCT TTCGAATGTG
<II-10pB31_S11      #51               TGGATCTACC ACCTCCGCCC TAGCCCCCGT GTATCGGCCT TTCGAATGTG
                                      ..................................................
 InII-10pB31        #51               TGGATCTACC ACCTCCGCCC TAGCCCCCGT GTATCGGCCT TTCGAATGTG


>Input_InII-10pB31  #101              CGTAATGAAC CCTCCCCCCC TTCGTCGCTC CTACCCCTCG CTTGTCTCGG
<II-10pB31_S3       #101              CGTAATGAAC CCTCCCCCCC TTCGTCGCTC CTACCCCTCG CTTGTCTCGG
<II-10pB31_S4       #101              CGTAATGAAC CCTCCCCCCC TTCGTCGCTC CTACCCCTCG CTTGTCTCGG
<II-10pB31_S8       #101              CGTAATGAAC CCTCCCCCCC TTCGTCGCTC CTACCCCTCG CTTGTCTCGG
<II-10pB31_S9       #101              CGTAATGAAC CCTCCCCCCC TTCGTCGCTC CTACCCCTCG CTTGTCTCGG
<II-10pB31_S11      #101              CGTAATGAAC CCTCCCCCCC TTCGTCGCTC CTACCCCTCG CTTGTCTCGG
                                      ..................................................
 InII-10pB31        #101              CGTAATGAAC CCTCCCCCCC TTCGTCGCTC CTACCCCTCG CTTGTCTCGG

>Input_InII-10pB31  #151              ATCC
<II-10pB31_S3       #151              ATCC
<II-10pB31_S4       #151              ATCC
<II-10pB31_S8       #151              ATCC
<II-10pB31_S9       #151              ATCC
<II-10pB31_S11      #151              ATCC
                                      ..................................................
 InII-10pB31        #151              ATCC
```

DNA Analysis of *Arabidopsis thaliana* transformed with InII-12pB31 sequence

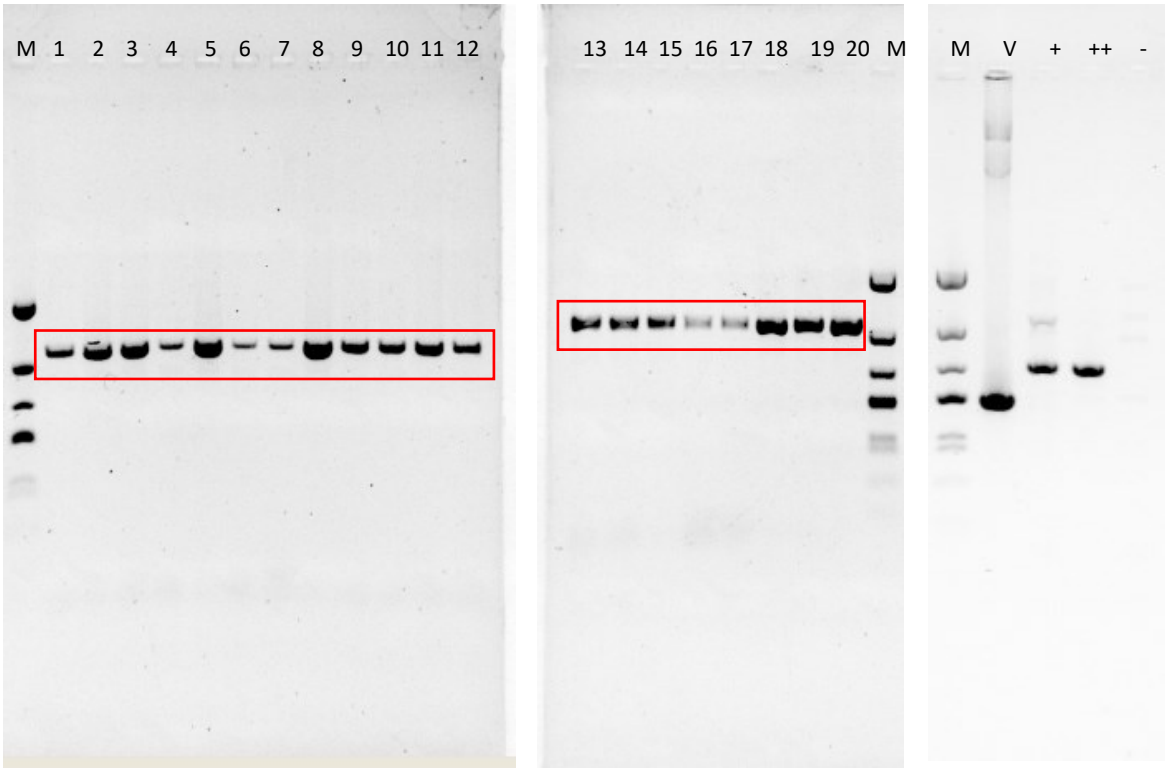


Figure A2.4: Example of agarose gel electrophoresis showing PCR amplification of InII-12 in pB31 with 1381F and GUS5'Rev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate insulator InII-12pB31 are shown. Amplification of DNA used 1300LacZF and GUS5'Rev primers with an expected band size of 877bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80

1-20: Transgenic plant DNA containing InI-12pB31 inserts (877bp)

V : pB31 control vector (no insert, 450bp)

+ : II-10 in pB31 sequenced control (604bp)

++ : I-3Δ in pB31 sequenced control (587bp)

- : Negative control containing water in replace of DNA

Alignment of "InII-12pB31" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_InII-12pB31  #1                GAATTCAGGA GGAGGTACCT TTCTCCATAG ACTCTCGGAT
>II-12pB31_S5       #1                GAATTCAGGA GGAGGTACCT TTCTCCATAG ACTCTCGGAT
>II-12pB31_S4       #1                GAATTCAGGA GGAGGTACCT TTCTCCATAG ACTCTCGGAT
>II-12pB31_S3       #1                GAATTCAGGA GGAGGTACCT TTCTCCATAG ACTCTCGGAT
>II-12pB31_S1       #1                GAATTCAGGA GGAGGTACCT TTCTCCATAG ACTCTCGGAT
>II-12pB31_S2       #1                GAATTCAGGA GGAGGTACCT TTCTCCATAG ACTCTCGGAT
                                      ..........................................
 InII-12pB31        #1                GAATTCAGGA GGAGGTACCT TTCTCCATAG ACTCTCGGAT
 Consensus

>Input_InII-12pB31  #41               GCTTGAATTC CATATTGCTA CGTATGGTTG AATCCTCCCC
>II-12pB31_S5       #41               GCTTGAATTC CATATTGCTA CGTATGGTTG AATCCTCCCC
>II-12pB31_S4       #41               GCTTGAATTC CATATTGCTA CGTATGGTTG AATCCTCCCC
>II-12pB31_S3       #41               GCTTGAATTC CATATTGCTA CGTATGGTTG AATCCTCCCC
>II-12pB31_S1       #41               GCTTGAATTC CATATTGCTA CGTATGGTTG AATCCTCCCC
>II-12pB31_S2       #41               GCTTGAATTC CATATTGCTA CGTATGGTTG AATCCTCCCC
                                      ..........................................
 InII-12pB31        #41               GCTTGAATTC CATATTGCTA CGTATGGTTG AATCCTCCCC


>Input_InII-12pB31  #81               TCCTGCCTCC CCACCCAACT CTCCACGTGC TTTTAAGTGC
>II-12pB31_S5       #81               TCCTGCCTCC CCACCCAACT CTCCACGTGC TTTTAAGTGC
>II-12pB31_S4       #81               TCCTGCCTCC CCACCCAACT CTCCACGTGC TTTTAAGTGC
>II-12pB31_S3       #81               TCCTGCCTCC CCACCCAACT CTCCACGTGC TTTTAAGTGC
>II-12pB31_S1       #81               TCCTGCCTCC CCACCCAACT CTCCACGTGC TTTTAAGTGC
>II-12pB31_S2       #81               TCCTGCCTCC CCACCCAACT CTCCACGTGC TTTTAAGTGC
                                      ..........................................
 InII-12pB31        #81               TCCTGCCTCC CCACCCAACT CTCCACGTGC TTTTAAGTGC


>Input_InII-12pB31  #121              TTGGACGGAG CTTGTCTCGG ATCCGAGACA AGCGACTTGT
>II-12pB31_S5       #121              TTGGACGGAG CTTGTCTCGG ATCCGAGACA AGCGACTTGT
>II-12pB31_S4       #121              TTGGACGGAG CTTGTCTCGG ATCCGAGACA AGCGACTTGT
>II-12pB31_S3       #121              TTGGACGGAG CTTGTCTCGG ATCCGAGACA AGCGACTTGT
>II-12pB31_S1       #121              TTGGACGGAG CTTGTCTCGG ATCCGAGACA AGCGACTTGT
>II-12pB31_S2       #121              TTGGACGGAG CTTGTCTCGG ATCCGAGACA AGCGACTTGT
                                      ..........................................
```

```
 InII-12pB31        #121                TTGGACGGAG CTTGTCTCGG ATCCGAGACA AGCGACTTGT


>Input_InII-12pB31  #161                GACCAAATAG ACCGGGTAGA CAGCTGGCAA ATGCCGTGGA
>II-12pB31_S5       #161                GACCAAATAG ACCGGGTAGA CAGCTGGCAA ATGCCGTGGA
>II-12pB31_S4       #161                GACCAAATAG ACCGGGTAGA CAGCTGGCAA ATGCCGTGGA
>II-12pB31_S3       #161                GACCAAATAG ACCGGGTAGA CAGCTGGCAA ATGCCGTGGA
>II-12pB31_S1       #161                GACCAAATAG ACCGGGTAGA CAGCTGGCAA ATGCCGTGGA
>II-12pB31_S2       #161                GACCAAATAG ACCGGGTAGA CAGCTGGCAA ATGCCGTGGA
                                        ..........................................
 InII-12pB31        #161                GACCAAATAG ACCGGGTAGA CAGCTGGCAA ATGCCGTGGA


>Input_InII-12pB31  #201                ACGAAGGTGG GGTGCGGTTG GCCGGCGGGA GGGAAGGAAG
>II-12pB31_S5       #201                ACGAAGGTGG GGTGCGGTTG GCCGGCGGGA GGGAAGGAAG
>II-12pB31_S4       #201                ACGAAGGTGG GGTGCGGTTG GCCGGCGGGA GGGAAGGAAG
>II-12pB31_S3       #201                ACGAAGGTGG GGTGCGGTTG GCCGGCGGGA GGGAAGGAAG
>II-12pB31_S1       #201                ACGAAGGTGG GGTGCGGTTG GCCGGCGGGA GGGAAGGAAG
>II-12pB31_S2       #201                ACGAAGGTGG GGTGCGGTTG GCCGGCGGGA GGGAAGGAAG
                                        ..........................................
 InII-12pB31        #201                ACGAAGGTGG GGTGCGGTTG GCCGGCGGGA GGGAAGGAAG


>Input_InII-12pB31  #241                GACAGGGGAG AGGAGAATAA TCAGCCATAC GCGGGTACCT
>II-12pB31_S5       #241                GACAGGGGAG AGGAGAATAA TCAGCCATAC GCGGGTACCT
>II-12pB31_S4       #241                GACAGGGGAG AGGAGAATAA TCAGCCATAC GCGGGTACCT
>II-12pB31_S3       #241                GACAGGGGAG AGGAGAATAA TCAGCCATAC GCGGGTACCT
>II-12pB31_S1       #241                GACAGGGGAG AGGAGAATAA TCAGCCATAC GCGGGTACCT
>II-12pB31_S2       #241                GACAGGGGAG AGGAGAATAA TCAGCCATAC GCGGGTACCT
                                        ..........................................
 InII-12pB31        #241                GACAGGGGAG AGGAGAATAA TCAGCCATAC GCGGGTACCT


>Input_InII-12pB31  #281                CCTCCTGAAT TCAGGAGGAG GCGTTGTTTC TCCCCTCACA
>II-12pB31_S5       #281                CCTCCTGAAT TCAGGAGGAG GCGTTGTTTC TCCCCTCACA
>II-12pB31_S4       #281                CCTCCTGAAT TCAGGAGGAG GCGTTGTTTC TCCCCTCACA
>II-12pB31_S3       #281                CCTCCTGAAT TCAGGAGGAG GCGTTGTTTC TCCCCTCACA
>II-12pB31_S1       #281                CCTCCTGAAT TCAGGAGGAG GCGTTGTTTC TCCCCTCACA
>II-12pB31_S2       #281                CCTCCTGAAT TCAGGAGGAG GCGTTGTTTC TCCCCTCACA
```

```
                                        ........................................
 InII-12pB31        #281                CCTCCTGAAT TCAGGAGGAG GCGTTGTTTC TCCCCTCACA


>Input_InII-12pB31  #321                TATTGATTTT CTTGTGCCAC CTCTGCTGTC TATCTCTTTC
>II-12pB31_S5       #321                TATTGATTTT CTTGTGCCAC CTCTGCTGTC TATCTCTTTC
>II-12pB31_S4       #321                TATTGATTTT CTTGTGCCAC CTCTGCTGTC TATCTCTTTC
>II-12pB31_S3       #321                TATTGATTTT CTTGTGCCAC CTCTGCTGTC TATCTCTTTC
>II-12pB31_S1       #321                TATTGATTTT CTTGTGCCAC CTCTGCTGTC TATCTCTTTC
>II-12pB31_S2       #321                TATTGATTTT CTTGTGCCAC CTCTGCTGTC TATCTCTTTC
                                        ........................................
 InII-12pB31        #321                TATTGATTTT CTTGTGCCAC CTCTGCTGTC TATCTCTTTC


>Input_InII-12pB31  #361                ACACTCTTCC CTCCTTTTAG CCTCTACGTG CTATGTGGAT
>II-12pB31_S5       #361                ACACTCTTCC CTCCTTTTAG CCTCTACGTG CTATGTGGAT
>II-12pB31_S4       #361                ACACTCTTCC CTCCTTTTAG CCTCTACGTG CTATGTGGAT
>II-12pB31_S3       #361                ACACTCTTCC CTCCTTTTAG CCTCTACGTG CTATGTGGAT
>II-12pB31_S1       #361                ACACTCTTCC CTCCTTTTAG CCTCTACGTG CTATGTGGAT
>II-12pB31_S2       #361                ACACTCTTCC CTCCTTTTAG CCTCTACGTG CTATGTGGAT
                                        ........................................
 InII-12pB31        #361                ACACTCTTCC CTCCTTTTAG CCTCTACGTG CTATGTGGAT


>Input_InII-12pB31  #401                TTCCTGGCAT AGGCTTGTCT CGGATCC
>II-12pB31_S5       #401                TTCCTGGCAT AGGCTTGTCT CGGATCC
>II-12pB31_S4       #401                TTCCTGGCAT AGGCTTGTCT CGGATCC
>II-12pB31_S3       #401                TTCCTGGCAT AGGCTTGTCT CGGATCC
>II-12pB31_S1       #401                TTCCTGGCAT AGGCTTGTCT CGGATCC
>II-12pB31_S2       #401                TTCCTGGCAT AGGCTTGTCT CGGATCC
                                        ........................................
 InII-12pB31        #401                TTCCTGGCAT AGGCTTGTCT CGGATCC
```

DNA Analysis of *Arabidopsis thaliana* transformed with InIII-50pB31 sequence

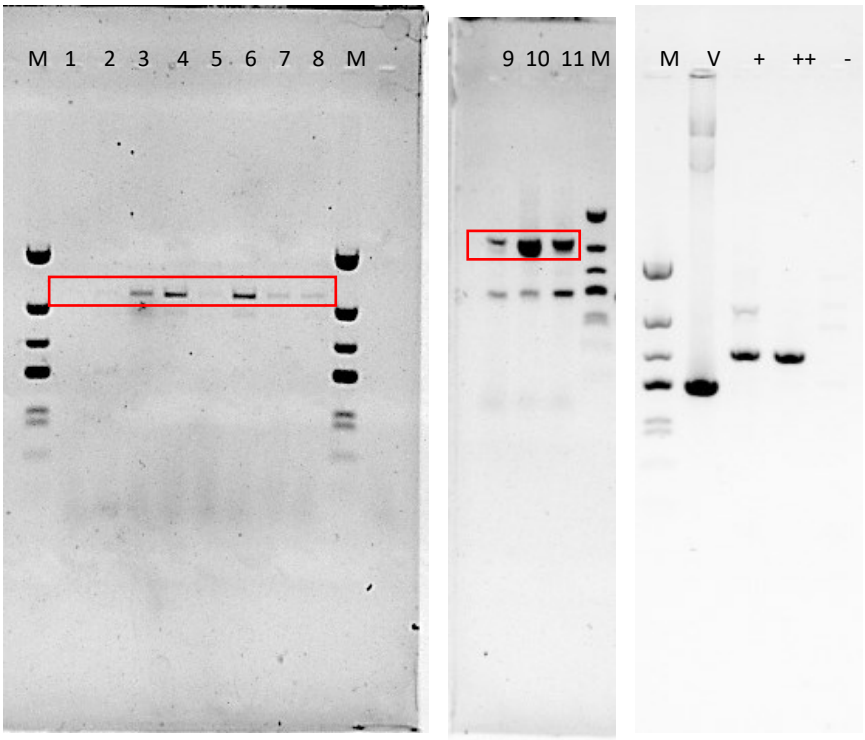


Figure A2.5: Example of agarose gel electrophoresis showing PCR amplification of InIII-50 in pB31 with 1381F and GUS5'Rev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate insulator InII-50pB31 are shown. Amplification of DNA used 1300LacZF and GUS5'Rev primers with an expected band size of 893bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80

1-11: Transgenic plant DNA containing InIII-50pB31 inserts (893bp)

V : pB31 control vector (no insert, 450bp)

+ : II-10 in pB31 sequenced control (604bp)

++ : I-3Δ in pB31 sequenced control (587bp)

- : Negative control containing water in replace of DNA

Alignment of "InIII-50pB31" samples and consensus sequence output from Sequencher 4.10.1

```
<Input_InIII-50pB31 #1              GGATCCGAGA CAAGCCGGCA AGCTCGGGCG CGCAGGCGCC ACGTGATAGA
>III-50pB31_S8      #1              GGATCCGAGA CAAGCCGGCA AGCTCGGGCG CGCAGGCGCC ACGTGATAGA
>III-50pB31_S6      #1              GGATCCGAGA CAAGCCGGCA AGCTCGGGCG CGCAGGCGCC ACGTGATAGA
>III-50pB31_S4      #1              GGATCCGAGA CAAGCCGGCA AGCTCGGGCG CGCAGGCGCC ACGTGATAGA
>III-50pB31_S2      #1              GGATCCGAGA CAAGCCGGCA AGCTCGGGCG CGCAGGCGCC ACGTGATAGA
>III-50pB31_S3      #1              GGATCCGAGA CAAGCCGGCA AGCTCGGGCG CGCAGGCGCC ACGTGATAGA
                                    ..................................................
 InIII-50pB31       #1              GGATCCGAGA CAAGCCGGCA AGCTCGGGCG CGCAGGCGCC ACGTGATAGA
 Consensus

<Input_InIII-50pB31 #51             AGGAGAGGAA GTGAGAGATA AACAGAAGGA GGAGGCCCGA TATGGTTGAG
>III-50pB31_S8      #51             AGGAGAGGAA GTGAGAGATA AACAGAAGGA GGAGGCCCGA TATGGTTGAG
>III-50pB31_S6      #51             AGGAGAGGAA GTGAGAGATA AACAGAAGGA GGAGGCCCGA TATGGTTGAG
>III-50pB31_S4      #51             AGGAGAGGAA GTGAGAGATA AACAGAAGGA GGAGGCCCGA TATGGTTGAG
>III-50pB31_S2      #51             AGGAGAGGAA GTGAGAGATA AACAGAAGGA GGAGGCCCGA TATGGTTGAG
>III-50pB31_S3      #51             AGGAGAGGAA GTGAGAGATA AACAGAAGGA GGAGGCCCGA TATGGTTGAG
                                    ..................................................
 InIII-50pB31       #51             AGGAGAGGAA GTGAGAGATA AACAGAAGGA GGAGGCCCGA TATGGTTGAG


<Input_InIII-50pB31 #101            GCGTCGCAGT GCGACGGGAA AAGTAAGGCC ACGCGGAGAC CTCCTCCTGA
>III-50pB31_S8      #101            GCGTCGCAGT GCGACGGGAA AAGTAAGGCC ACGCGGAGAC CTCCTCCTGA
>III-50pB31_S6      #101            GCGTCGCAGT GCGACGGGAA AAGTAAGGCC ACGCGGAGAC CTCCTCCTGA
>III-50pB31_S4      #101            GCGTCGCAGT GCGACGGGAA AAGTAAGGCC ACGCGGAGAC CTCCTCCTGA
>III-50pB31_S2      #101            GCGTCGCAGT GCGACGGGAA AAGTAAGGCC ACGCGGAGAC CTCCTCCTGA
>III-50pB31_S3      #101            GCGTCGCAGT GCGACGGGAA AAGTAAGGCC ACGCGGAGAC CTCCTCCTGA
                                    ..................................................
 InIII-50pB31       #101            GCGTCGCAGT GCGACGGGAA AAGTAAGGCC ACGCGGAGAC CTCCTCCTGA


<Input_InIII-50pB31 #151            ATTCAGGAGG AGGCCGACTT TGCGGGCCTT TGTACTCAGC TAACAACTCG
>III-50pB31_S8      #151            ATTCAGGAGG AGGCCGACTT TGCGGGCCTT TGTACTCAGC TAACAACTCG
>III-50pB31_S6      #151            ATTCAGGAGG AGGCCGACTT TGCGGGCCTT TGTACTCAGC TAACAACTCG
>III-50pB31_S4      #151            ATTCAGGAGG AGGCCGACTT TGCGGGCCTT TGTACTCAGC TAACAACTCG
>III-50pB31_S2      #151            ATTCAGGAGG AGGCCGACTT TGCGGGCCTT TGTACTCAGC TAACAACTCG
>III-50pB31_S3      #151            ATTCAGGAGG AGGCCGACTT TGCGGGCCTT TGTACTCAGC TAACAACTCG
                                    ..................................................
```

```
 InIII-50pB31       #151              ATTCAGGAGG AGGCCGACTT TGCGGGCCTT TGTACTCAGC TAACAACTCG


<Input_InIII-50pB31 #201              CCACGGTCGC ATATTTCCCG TGTAAAGCGT TTCTACGCGT CGTCCTTGCG
>III-50pB31_S8      #201              CCACGGTCGC ATATTTCCCG TGTAAAGCGT TTCTACGCGT CGTCCTTGCG
>III-50pB31_S6      #201              CCACGGTCGC ATATTTCCCG TGTAAAGCGT TTCTACGCGT CGTCCTTGCG
>III-50pB31_S4      #201              CCACGGTCGC ATATTTCCCG TGTAAAGCGT TTCTACGCGT CGTCCTTGCG
>III-50pB31_S2      #201              CCACGGTCGC ATATTTCCCG TGTAAAGCGT TTCTACGCGT CGTCCTTGCG
>III-50pB31_S3      #201              CCACGGTCGC ATATTTCCCG TGTAAAGCGT TTCTACGCGT CGTCCTTGCG
                                      ..................................................
 InIII-50pB31       #201              CCACGGTCGC ATATTTCCCG TGTAAAGCGT TTCTACGCGT CGTCCTTGCG


<Input_InIII-50pB31 #251              CGACCTCCTG CGTGACCGAG CCACCTCGGA ACCGTTACTT GTCTCGGATC
>III-50pB31_S8      #251              CGACCTCCTG CGTGACCGAG CCACCTCGGA ACCGTTACTT GTCTCGGATC
>III-50pB31_S6      #251              CGACCTCCTG CGTGACCGAG CCACCTCGGA ACCGTTACTT GTCTCGGATC
>III-50pB31_S4      #251              CGACCTCCTG CGTGACCGAG CCACCTCGGA ACCGTTACTT GTCTCGGATC
>III-50pB31_S2      #251              CGACCTCCTG CGTGACCGAG CCACCTCGGA ACCGTTACTT GTCTCGGATC
>III-50pB31_S3      #251              CGACCTCCTG CGTGACCGAG CCACCTCGGA ACCGTTACTT GTCTCGGATC
                                      ..................................................
 InIII-50pB31       #251              CGACCTCCTG CGTGACCGAG CCACCTCGGA ACCGTTACTT GTCTCGGATC


<Input_InIII-50pB31 #301              CGAGACAAGC AGATTACGCG AGGACGGGAG GGGCACAAAT GCACGAAATA
>III-50pB31_S8      #301              CGAGACAAGC AGATTACGCG AGGACGGGAG GGGCACAAAT GCACGAAATA
>III-50pB31_S6      #301              CGAGACAAGC AGATTACGCG AGGACGGGAG GGGCACAAAT GCACGAAATA
>III-50pB31_S4      #301              CGAGACAAGC AGATTACGCG AGGACGGGAG GGGCACAAAT GCACGAAATA
>III-50pB31_S2      #301              CGAGACAAGC AGATTACGCG AGGACGGGAG GGGCACAAAT GCACGAAATA
>III-50pB31_S3      #301              CGAGACAAGC AGATTACGCG AGGACGGGAG GGGCACAAAT GCACGAAATA
                                      ..................................................
 InIII-50pB31       #301              CGAGACAAGC AGATTACGCG AGGACGGGAG GGGCACAAAT GCACGAAATA


<Input_InIII-50pB31 #351              AAGCGAGTGC ATGCCGCCAG TAGCAACATC CTGGTGTACG CGCTATGTGC
>III-50pB31_S8      #351              AAGCGAGTGC ATGCCGCCAG TAGCAACATC CTGGTGTACG CGCTATGTGC
>III-50pB31_S6      #351              AAGCGAGTGC ATGCCGCCAG TAGCAACATC CTGGTGTACG CGCTATGTGC
>III-50pB31_S4      #351              AAGCGAGTGC ATGCCGCCAG TAGCAACATC CTGGTGTACG CGCTATGTGC
>III-50pB31_S2      #351              AAGCGAGTGC ATGCCGCCAG TAGCAACATC CTGGTGTACG CGCTATGTGC
>III-50pB31_S3      #351              AAGCGAGTGC ATGCCGCCAG TAGCAACATC CTGGTGTACG CGCTATGTGC
```

```
                                                  ..................................................
 InIII-50pB31       #351                          AAGCGAGTGC ATGCCGCCAG TAGCAACATC CTGGTGTACG CGCTATGTGC


<Input_InIII-50pB31 #401                          GTGGTACGGA CGCGGGTCGA GCAGAGGGCC TCCTCCTGAA TTC
>III-50pB31_S8      #401                          GTGGTACGGA CGCGGGTCGA GCAGAGGGCC TCCTCCTGAA TTC
>III-50pB31_S6      #401                          GTGGTACGGA CGCGGGTCGA GCAGAGGGCC TCCTCCTGAA TTC
>III-50pB31_S4      #401                          GTGGTACGGA CGCGGGTCGA GCAGAGGGCC TCCTCCTGAA TTC
>III-50pB31_S2      #401                          GTGGTACGGA CGCGGGTCGA GCAGAGGGCC TCCTCCTGAA TTC
>III-50pB31_S3      #401                          GTGGTACGGA CGCGGGTCGA GCAGAGGGCC TCCTCCTGAA TTC
                                                  ..................................................
 InIII-50pB31       #401                          GTGGTACGGA CGCGGGTCGA GCAGAGGGCC TCCTCCTGAA TTC
```

DNA Analysis of *Arabidopsis thaliana* transformed with InIII-78pB31 sequence

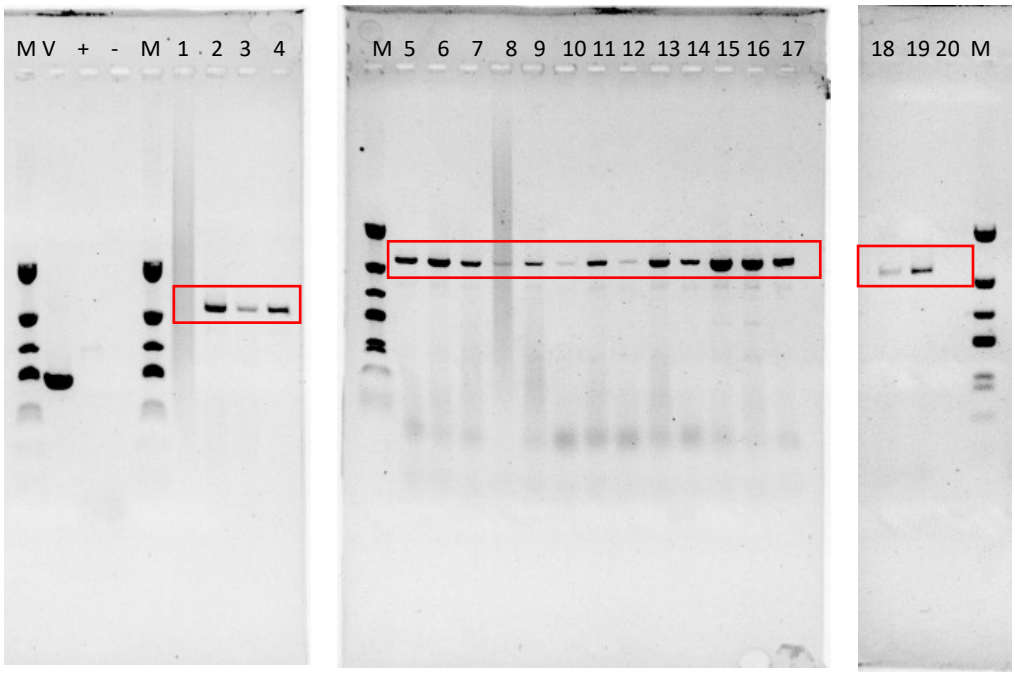


Figure A2.6: Example of agarose gel electrophoresis showing PCR amplification of InIII-78 in pB31 with 1381F and GUS5'Rev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate insulator InIII-78pB31 are shown. Amplification of DNA used 1300LacZF and GUS5'Rev primers with an expected band size of 883bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80

1-20 : Transgenic plant DNA containing InIII-78pB31 inserts (833bp)

V : pB31 control vector (no insert, 450bp)

+ : I-3Δ in pB31 sequenced control (587bp)

- : Negative control containing water in replace of DNA

Alignment of "InIII-78pB31" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_InIII-78     #1      GAATTCAGGA GGAGGGTTCG TTTGGATTTC ATGCAGCTCC GCTCCGGATC
>III-78pB31_S2      #1      GAATTCAGGA GGAGGGTTCG TTTGGATTTC ATGCAGCTCC GCTCCGGATC
<III-78pB31_S4      #1      GAATTCAGGA GGAGGGTTCG TTTGGATTTC ATGCAGCTCC GCTCCGGATC
>III-78pB31_S5      #1      GAATTCAGGA GGAGGGTTCG TTTGGATTTC ATGCAGCTCC GCTCCGGATC
>III-78pB31_S6      #1      GAATTCAGGA GGAGGGTTCG TTTGGATTTC ATGCAGCTCC GCTCCGGATC
>III-78pB31_S7      #1      GAATTCAGGA GGAGGGTTCG TTTGGATTTC ATGCAGCTCC GCTCCGGATC
                            ..................................................
 InIII-78pB31       #1      GAATTCAGGA GGAGGGTTCG TTTGGATTTC ATGCAGCTCC GCTCCGGATC
 Consensus

>Input_InIII-78     #51     AATCTTCCAT TCTAGCTTAG TGCCTCTTCC ATTTGTTGCT CCTGGGCGTC
>III-78pB31_S2      #51     AATCTTCCAT TCTAGCTTAG TGCCTCTTCC ATTTGTTGCT CCTGGGCGTC
<III-78pB31_S4      #51     AATCTTCCAT TCTAGCTTAG TGCCTCTTCC ATTTGTTGCT CCTGGGCGTC
>III-78pB31_S5      #51     AATCTTCCAT TCTAGCTTAG TGCCTCTTCC ATTTGTTGCT CCTGGGCGTC
>III-78pB31_S6      #51     AATCTTCCAT TCTAGCTTAG TGCCTCTTCC ATTTGTTGCT CCTGGGCGTC
>III-78pB31_S7      #51     AATCTTCCAT TCTAGCTTAG TGCCTCTTCC ATTTGTTGCT CCTGGGCGTC
                            ..................................................
 InIII-78pB31       #51     AATCTTCCAT TCTAGCTTAG TGCCTCTTCC ATTTGTTGCT CCTGGGCGTC


>Input_InIII-78     #101    TCCCCCGGGA CTAGTCTGTG CCGTTGTGCT TGTCTCGGAT CCGAGACAAG
>III-78pB31_S2      #101    TCCCCCGGGA CTAGTCTGTG CCGTTGTGCT TGTCTCGGAT CCGAGACAAG
<III-78pB31_S4      #101    TCCCCCGGGA CTAGTCTGTG CCGTTGTGCT TGTCTCGGAT CCGAGACAAG
>III-78pB31_S5      #101    TCCCCCGGGA CTAGTCTGTG CCGTTGTGCT TGTCTCGGAT CCGAGACAAG
>III-78pB31_S6      #101    TCCCCCGGGA CTAGTCTGTG CCGTTGTGCT TGTCTCGGAT CCGAGACAAG
>III-78pB31_S7      #101    TCCCCCGGGA CTAGTCTGTG CCGTTGTGCT TGTCTCGGAT CCGAGACAAG
                            ..................................................
 InIII-78pB31       #101    TCCCCCGGGA CTAGTCTGTG CCGTTGTGCT TGTCTCGGAT CCGAGACAAG


>Input_InIII-78     #151    CTTACGGCGC GATGAGGGGG AGACCAAATA ACGAAATCAC CGGACGTCGT
>III-78pB31_S2      #151    CTTACGGCGC GATGAGGGGG AGACCAAATA ACGAAATCAC CGGACGTCGT
<III-78pB31_S4      #151    CTTACGGCGC GATGAGGGGG AGACCAAATA ACGAAATCAC CGGACGTCGT
>III-78pB31_S5      #151    CTTACGGCGC GATGAGGGGG AGACCAAATA ACGAAATCAC CGGACGTCGT
>III-78pB31_S6      #151    CTTACGGCGC GATGAGGGGG AGACCAAATA ACGAAATCAC CGGACGTCGT
>III-78pB31_S7      #151    CTTACGGCGC GATGAGGGGG AGACCAAATA ACGAAATCAC CGGACGTCGT
                            ..................................................
```

```
 InIII-78pB31       #151              CTTACGGCGC GATGAGGGGG AGACCAAATA ACGAAATCAC CGGACGTCGT


>Input_InIII-78     #201              GCATAAGCGT CGCGCAAGGG TCTCAGGGAG AGAGGTGTAA CAGGCGTGTC
>III-78pB31_S2      #201              GCATAAGCGT CGCGCAAGGG TCTCAGGGAG AGAGGTGTAA CAGGCGTGTC
<III-78pB31_S4      #201              GCATAAGCGT CGCGCAAGGG TCTCAGGGAG AGAGGTGTAA CAGGCGTGTC
>III-78pB31_S5      #201              GCATAAGCGT CGCGCAAGGG TCTCAGGGAG AGAGGTGTAA CAGGCGTGTC
>III-78pB31_S6      #201              GCATAAGCGT CGCGCAAGGG TCTCAGGGAG AGAGGTGTAA CAGGCGTGTC
>III-78pB31_S7      #201              GCATAAGCGT CGCGCAAGGG TCTCAGGGAG AGAGGTGTAA CAGGCGTGTC
                                      ..................................................
 InIII-78pB31       #201              GCATAAGCGT CGCGCAAGGG TCTCAGGGAG AGAGGTGTAA CAGGCGTGTC


>Input_InIII-78     #251              GCTCCGCGGG GGTTAAGGGT GCCTCCTCCT GAATTCAGGA GGAGGTCCAC
>III-78pB31_S2      #251              GCTCCGCGGG GGTTAAGGGT GCCTCCTCCT GAATTCAGGA GGAGGTCCAC
<III-78pB31_S4      #251              GCTCCGCGGG GGTTAAGGGT GCCTCCTCCT GAATTCAGGA GGAGGTCCAC
>III-78pB31_S5      #251              GCTCCGCGGG GGTTAAGGGT GCCTCCTCCT GAATTCAGGA GGAGGTCCAC
>III-78pB31_S6      #251              GCTCCGCGGG GGTTAAGGGT GCCTCCTCCT GAATTCAGGA GGAGGTCCAC
>III-78pB31_S7      #251              GCTCCGCGGG GGTTAAGGGT GCCTCCTCCT GAATTCAGGA GGAGGTCCAC
                                      ..................................................
 InIII-78pB31       #251              GCTCCGCGGG GGTTAAGGGT GCCTCCTCCT GAATTCAGGA GGAGGTCCAC


>Input_InIII-78     #301              CTCCCTCTCT CGCGCTATTC GCCCGCTCCC ACTCCCTCTT ACCCGCTTCT
>III-78pB31_S2      #301              CTCCCTCTCT CGCGCTATTC GCCCGCTCCC ACTCCCTCTT ACCCGCTTCT
<III-78pB31_S4      #301              CTCCCTCTCT CGCGCTATTC GCCCGCTCCC ACTCCCTCTT ACCCGCTTCT
>III-78pB31_S5      #301              CTCCCTCTCT CGCGCTATTC GCCCGCTCCC ACTCCCTCTT ACCCGCTTCT
>III-78pB31_S6      #301              CTCCCTCTCT CGCGCTATTC GCCCGCTCCC ACTCCCTCTT ACCCGCTTCT
>III-78pB31_S7      #301              CTCCCTCTCT CGCGCTATTC GCCCGCTCCC ACTCCCTCTT ACCCGCTTCT
                                      ..................................................
 InIII-78pB31       #301              CTCCCTCTCT CGCGCTATTC GCCCGCTCCC ACTCCCTCTT ACCCGCTTCT


>Input_InIII-78     #351              TCAAGCTAGC GTCCCTGTAC ACGTAACCAT CTTCCCTTTC GCCCGTCCTG
>III-78pB31_S2      #351              TCAAGCTAGC GTCCCTGTAC ACGTAACCAT CTTCCCTTTC GCCCGTCCTG
<III-78pB31_S4      #351              TCAAGCTAGC GTCCCTGTAC ACGTAACCAT CTTCCCTTTC GCCCGTCCTG
>III-78pB31_S5      #351              TCAAGCTAGC GTCCCTGTAC ACGTAACCAT CTTCCCTTTC GCCCGTCCTG
>III-78pB31_S6      #351              TCAAGCTAGC GTCCCTGTAC ACGTAACCAT CTTCCCTTTC CCCCGTCCTG
>III-78pB31_S7      #351              TCAAGCTAGC GTCCCTGTAC ACGTAACCAT CTTCCCTTTC CCCCGTCCTG
```

```
                                    ..................................................
 InIII-78pB31       #351            TCAAGCTAGC GTCCCTGTAC ACGTAACCAT CTTCCCTTTC GCCCGTCCTG
                                                                                    *

>Input_InIII-78     #401            GCTATCCGCG TGACGTGAGC TTGTCTCGGA TCC
>III-78pB31_S2      #401            GDTHTCCGGG TTAWYCNNCC TTGMCTAGAA RCC
<III-78pB31_S4      #401            GCTATCCGCG TGACGTGAGC TTGTCTCGGA TCC
>III-78pB31_S5      #401            GCTATCCGCG TGACGTGAGC TTGTCTCGGA TCC
>III-78pB31_S6      #401            GCTATCCGCG TGACGTGAGC TTGTCTCGGA TCC
>III-78pB31_S7      #401            GCTATCCGCG TGACGTGATC TTGTCTCGGA TCC
                                    ..................................................
 InIII-78pB31       #401            GCTATCCGCG TGACGTGAGC TTGTCTCGGA TCC
                                     + +    *   * ++*++*     +  * *  +
```

DNA Analysis of *Arabidopsis thaliana* transformed with InI-3pL1 sequence

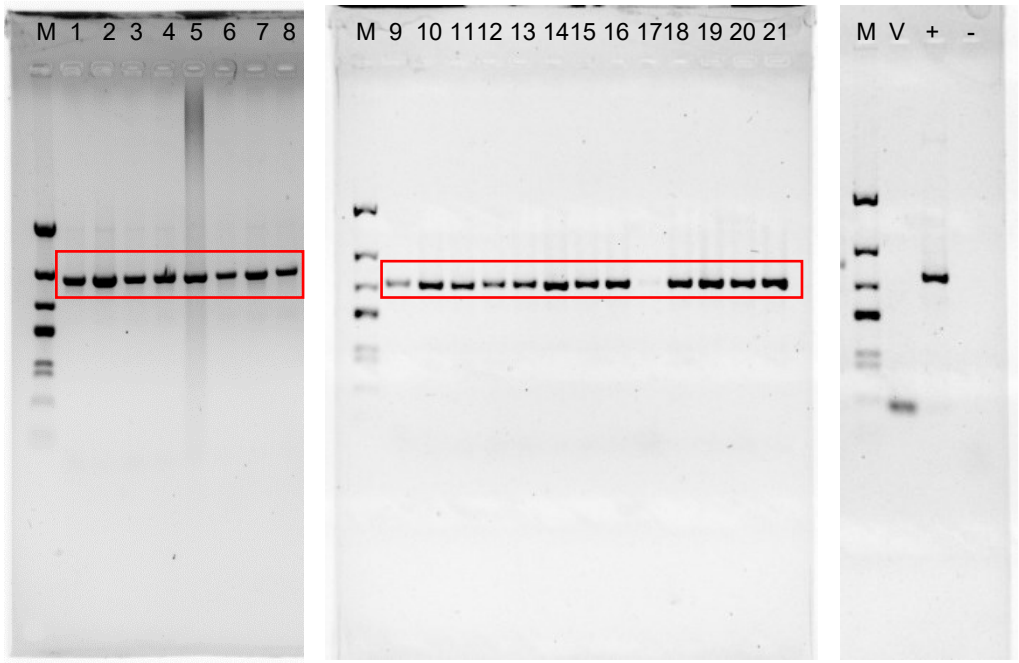


Figure A2.7: Example of agarose gel electrophoresis showing PCR amplification of clone InI-3 in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate InI-3pL1 insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 620bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-21: Transgenic plant DNA containing InI-3pL1 inserts (620bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "InI-3pL1" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_InI-3pL1     #1                AAGCTTGAAT TCAGGAGGAG GCTCTCAGCC CGCCCTTCTC CCGGCTACGT
<InI-3pL1_S2        #1                AAGCTTGAAT TCAGGAGGAG GCTCTCAGCC CGCCCTTCTC CCGGCTACGT
<InI-3pL1_S3        #1                AAGCTTGAAT TCAGGAGGAG GCTCTCAGCC CGCCCTTCTC CCGGCTACGT
<InI-3pL1_S4        #1                AAGCTTGAAT TCAGGAGGAG GCTCTCAGCC CGCCCTTCTC CCGGCTACGT
>InI-3pL1_S6        >#1>                        TCAGGAGGAG GCTCTCAGCC CGCCCTTCTC CCGGCTACGT
<InI-3pL1_S8        #1                AAGCTTGAAT TCAGGAGGAG GCTCTCAGCC CGCCCTTCTC CCGGCTACGT
                                      ..................................................
 InI-3pL1           #1                AAGCTTGAAT TCAGGAGGAG GCTCTCAGCC CGCCCTTCTC CCGGCTACGT
 Consensus

>Input_InI-3pL1     #51               GTCTTTTACC CCCCCCTCTG TCCCCTTGTG GCATTGAGCA TATCGTTTCT
<InI-3pL1_S2        #51               GTCTTTTACC CCCCCCTCTG TCCCCTTGTG GCATTGAGCA TATCGTTTCT
<InI-3pL1_S3        #51               GTCTTTTACC CCCCCCTCTG TCCCCTTGTG GCATTGAGCA TATCGTTTCT
<InI-3pL1_S4        #51               GTCTTTTACC CCCCCCTCTG TCCCCTTGTG GCATTGAGCA TATCGTTTCT
>InI-3pL1_S6        #41               GTCTTTTACC CCCCCCTCTG TCCCCTTGTG GCATTGAGCA TATCGTTTCT
<InI-3pL1_S8        #51               GTCTTTTACC CCCCCCTCTG TCCCCTTGTG GCATTGAGCA TATCGTTTCT
                                      ..................................................
 InI-3pL1           #51               GTCTTTTACC CCCCCCTCTG TCCCCTTGTG GCATTGAGCA TATCGTTTCT


>Input_InI-3pL1     #101              AAAGTCCGCC CGTCTTAGGC ATTTGGGACC CAGCTTTAAC TGTTGCTTGT
<InI-3pL1_S2        #101              AAAGTCCGCC CGTCTTAGGC ATTTGGGACC CAGCTTTAAC TGTTGCTTGT
<InI-3pL1_S3        #101              AAAGTCCGCC CGTCTTAGGC ATTTGGGACC CAGCTTTAAC TGTTGCTTGT
<InI-3pL1_S4        #101              AAAGTCCGCC CGTCTTAGGC ATTTGGGACC CAGCTTTAAC TGTTGCTTGT
>InI-3pL1_S6        #91               AAAGTCCGCC CGTCTTAGGC ATTTGGGACC CAGCTTTAAC TGTTGCTTGT
<InI-3pL1_S8        #101              AAAGTCCGCC CGTCTTAGGC ATTTGGGACC CAGCTTTAAC TGTTGCTTGT
                                      ..................................................
 InI-3pL1           #101              AAAGTCCGCC CGTCTTAGGC ATTTGGGACC CAGCTTTAAC TGTTGCTTGT


>Input_InI-3pL1     #151              CTCGGATCCG AGACAAGCAG TAATAGAAAG GAATGGTAAC TGGAGAAGTA
<InI-3pL1_S2        #151              CTCGGATCCG AGACAAGCAG TAATAGAAAG GAATGGTAAC TGGAGAAGTA
<InI-3pL1_S3        #151              CTCGGATCCG AGACAAGCAG TAATAGAAAG GAATGGTAAC TGGAGAAGTA
<InI-3pL1_S4        #151              CTCGGATCCG AGACAAGCAG TAATAGAAAG GAATGGTAAC TGGAGAAGTA
```

```
>InI-3pL1_S6        #141              CTCGGATCCG AGACAAGCAG TAATAGAAAG GAATGGTAAC TGGAGAAGTA
<InI-3pL1_S8        #151              CTCGGATCCG AGACAAGCAG TAATAGAAAG GAATGGTAAC TGGAGAAGTA
                                      ..................................................
 InI-3pL1           #151              CTCGGATCCG AGACAAGCAG TAATAGAAAG GAATGGTAAC TGGAGAAGTA


>Input_InI-3pL1     #201              AGGCCGGGAA GCTTCACGAA GCCGCGGGAG GGAGAGAATA TGCACGCCAG
<InI-3pL1_S2        #201              AGGCCGGGAA GCTTCACGAA GCCGCGGGAG GGAGAGAATA TGCACGCCAG
<InI-3pL1_S3        #201              AGGCCGGGAA GCTTCACGAA GCCGCGGGAG GGAGAGAATA TGCACGCCAG
<InI-3pL1_S4        #201              AGGCCGGGAA GCTTCACGAA GCCGCGGGAG GGAGAGAATA TGCACGCCAG
>InI-3pL1_S6        #191              AGGCCGGGAA GCTTCACGAA GCCGCGGGAG GGAGAGAATA TGCACGCCAG
<InI-3pL1_S8        #201              AGGCCGGGAA GCTTCACGAA GCCGCGGGAG GGAGAGAATA TGCACGCCAG
                                      ..................................................
 InI-3pL1           #201              AGGCCGGGAA GCTTCACGAA GCCGCGGGAG GGAGAGAATA TGCACGCCAG


>Input_InI-3pL1     #251              GTCAGGGAGG ACCCGAAGGG AAAAAGTCGG GGAAGACCTG GGTCCTCCTC
<InI-3pL1_S2        #251              GTCAGGGAGG ACCCGAAGGG AAAAAGTCGG GGAAGACCTG GGTCCTCCTC
<InI-3pL1_S3        #251              GTCAGGGAGG ACCCGAAGGG AAAAAGTCGG GGAAGACCTG GGTCCTCCTC
<InI-3pL1_S4        #251              GTCAGGGAGG ACCCGAAGGG AAAAAGTCGG GGAAGACCTG GGTCCTCCTC
>InI-3pL1_S6        #241              GTCAGGGAGG ACCCGAAGGG AAAAAGTCGG GGAAGACCTG GGTCCTCCTC
<InI-3pL1_S8        #251              GTCAGGGAGG ACCCGAAGGG AAAAAGTCGG GGAAGACCTG GGTCCTCCTC
                                      ..................................................
 InI-3pL1           #251              GTCAGGGAGG ACCCGAAGGG AAAAAGTCGG GGAAGACCTG GGTCCTCCTC


>Input_InI-3pL1     #301              CTGAATTCAG GAGGAGGTCA ATGGCTTGTC GCACTAACCC ACCCTCGCTC
<InI-3pL1_S2        #301              CTGAATTCAG GAGGAGGTCA ATGGCTTGTC GCACTAACCC ACCCTCGCTC
<InI-3pL1_S3        #301              CTGAATTCAG GAGGAGGTCA ATGGCTTGTC GCACTAACCC ACCCTCGCTC
<InI-3pL1_S4        #301              CTGAATTCAG GAGGAGGTCA ATGGCTTGTC GCACTAACCC ACCCTCGCTC
>InI-3pL1_S6        #291              CTGAATTCAG GAGGAGGTCA ATGGCTTGTC GCACTAACCC ACCCTCGCTC
<InI-3pL1_S8        #301              CTGAATTCAG GAGGAGGTCA ATGGCTTGTC GCACTAACCC ACCCTCGCTC
                                      ..................................................
 InI-3pL1           #301              CTGAATTCAG GAGGAGGTCA ATGGCTTGTC GCACTAACCC ACCCTCGCTC


>Input_InI-3pL1     #351              ACCATTTTCC CCGCCGTGCA GCGCGTTATC TGACTTCCTC GCGATCTTTA
```

```
<InI-3pL1_S2        #351              ACCATTTTCC CCGCCGTGCA GCGCGTTATC TGACTTCCTC GCGATCTTTA
<InI-3pL1_S3        #351              ACCATTTTCC CCGCCGTGCA GCGCGTTATC TGACTTCCTC GCGATCTTTA
<InI-3pL1_S4        #351              ACCATTTTCC CCGCCGTGCA GCGCGTTATC TGACTTCCTC GCGATCTTTA
>InI-3pL1_S6        #341              ACCATTTTCC CCGCCGTGCA GCGCGTTATC TGACTTCCTC GCGATCTTTA
<InI-3pL1_S8        #351              ACCATTTTCC CCGCCGTGCA GCGCGTTATC TGACTTCCTC GCGATCTTTA
                                      ..................................................
 InI-3pL1           #351              ACCATTTTCC CCGCCGTGCA GCGCGTTATC TGACTTCCTC GCGATCTTTA


>Input_InI-3pL1     #401              CCCGCGGTTT GGAATAGCTG CGGTACCTCG CTTGTCTCGG ATCCCTGCAG
<InI-3pL1_S2        #401              CCCGCGGTTT GGAATAGCTG CGGTACCTCG CTTGTCTCGG ATCCCTGCAG
<InI-3pL1_S3        #401              CCCGCGGTTT GGAATAGCTG CGGTACCTCG CTTGTCTCGG ATCCCTGCAG
<InI-3pL1_S4        #401              CCCGCGGTTT GGAATAGCTG CGGTACCTCG CTTGTCTCGG ATCCCTGCAG
>InI-3pL1_S6        #391              CCCGCGGTTT GGAATAGCTG CGGTACCTCG CTTGTCTCGG ATCCCTGCAG
<InI-3pL1_S8        #401              CCCGCGGTTT GGAATAGCTG CGGTACCTCG CTTGTCTCGG ATCCCTGCAG
                                      ..................................................
 InI-3pL1           #401              CCCGCGGTTT GGAATAGCTG CGGTACCTCG CTTGTCTCGG ATCCCTGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with InII-12pL1 sequence

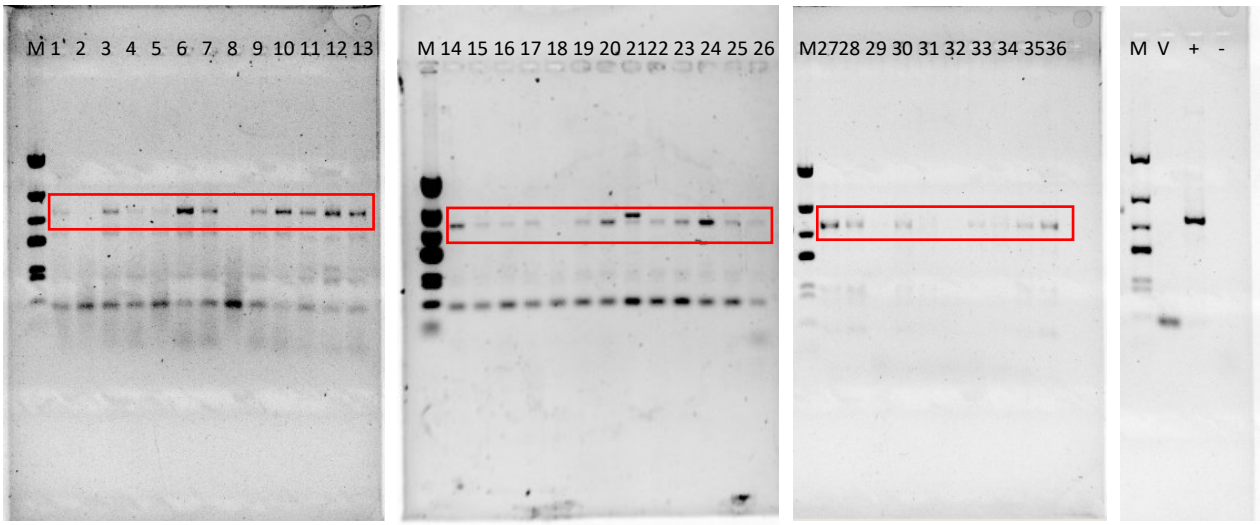


Figure A2.8: Example of agarose gel electrophoresis showing PCR amplification of clone InII-12 in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate InII-12pL1 insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 620bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-36 : Transgenic plant DNA containing InI-3pL1 inserts (620bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "InII-12pL1" samples and consensus sequence output from Sequencher 4.10.1

```
<Input_InII-12pL1   #1                CTGCAGGGAT CCGAGACAAG CCTATGCCAG GAAATCCACA TAGCACGTAG
>InII-12pL1_S4      #1                CTGCAGGGAT CCGAGACAAG CCTATGCCAG GAAATCCACA TAGCACGTAG
>InII-12pL1_S5      #1                CTGCAGGGAT CCGAGACAAG CCTATGCCAG GAAATCCACA TAGCACGTAG
>InII-12pL1_S6      #1                CTGCAGGGAT CCGAGACAAG CCTATGCCAG GAAATCCACA TAGCACGTAG
>InII-12pL1_S12     #1                CTGCAGGGAT CCGAGACAAG CCTATGCCAG GAAATCCACA TAGCACGTAG
>InII-12pL1_S13     #1                CTGCAGGGAT CCGAGACAAG CCTATGCCAG GAAATCCACA TAGCACGTAG
                                      ..................................................
 InII-12pL1         #1                CTGCAGGGAT CCGAGACAAG CCTATGCCAG GAAATCCACA TAGCACGTAG
 Consensus

<Input_InII-12pL1   #51               AGGCTAAAAG GAGGGAAGAG TGTGAAAGAG ATAGACAGCA GAGGTGGCAC
>InII-12pL1_S4      #51               AGGCTAAAAG GAGGGAAGAG TGTGAAAGAG ATAGACAGCA GAGGTGGCAC
>InII-12pL1_S5      #51               AGGCTAAAAG GAGGGAAGAG TGTGAAAGAG ATAGACAGCA GAGGTGGCAC
>InII-12pL1_S6      #51               AGGCTAAAAG GAGGGAAGAG TGTGAAAGAG ATAGACAGCA GAGGTGGCAC
>InII-12pL1_S12     #51               AGGCTAAAAG GAGGGAAGAG TGTGAAAGAG ATAGACAGCA GAGGTGGCAC
>InII-12pL1_S13     #51               AGGCTAAAAG GAGGGAAGAG TGTGAAAGAG ATAGACAGCA GAGGTGGCAC
                                      ..................................................
 InII-12pL1         #51               AGGCTAAAAG GAGGGAAGAG TGTGAAAGAG ATAGACAGCA GAGGTGGCAC


<Input_InII-12pL1   #101              AAGAAAATCA ATATGTGAGG GGAGAAACAA CGCCTCCTCC TGAATTCAGG
>InII-12pL1_S4      #101              AAGAAAATCA ATATGTGAGG GGAGAAACAA CGCCTCCTCC TGAATTCAGG
>InII-12pL1_S5      #101              AAGAAAATCA ATATGTGAGG GGAGAAACAA CGCCTCCTCC TGAATTCAGG
>InII-12pL1_S6      #101              AAGAAAATCA ATATGTGAGG GGAGAAACAA CGCCTCCTCC TGAATTCAGG
>InII-12pL1_S12     #101              AAGAAAATCA ATATGTGAGG GGAGAAACAA CGCCTCCTCC TGAATTCAGG
>InII-12pL1_S13     #101              AAGAAAATCA ATATGTGAGG GGAGAAACAA CGCCTCCTCC TGAATTCAGG
                                      ..................................................
 InII-12pL1         #101              AAGAAAATCA ATATGTGAGG GGAGAAACAA CGCCTCCTCC TGAATTCAGG


<Input_InII-12pL1   #151              AGGAGGTACC CGCGTATGGC TGATTATTCT CCTCTCCCCT GTCCTTCCTT
>InII-12pL1_S4      #151              AGGAGGTACC CGCGTATGGC TGATTATTCT CCTCTCCCCT GTCCTTCCTT
>InII-12pL1_S5      #151              AGGAGGTACC CGCGTATGGC TGATTATTCT CCTCTCCCCT GTCCTTCCTT
>InII-12pL1_S6      #151              AGGAGGTACC CGCGTATGGC TGATTATTCT CCTCTCCCCT GTCCTTCCTT
>InII-12pL1_S12     #151              AGGAGGTACC CGCGTATGGC TGATTATTCT CCTCTCCCCT GTCCTTCCTT
>InII-12pL1_S13     #151              AGGAGGTACC CGCGTATGGC TGATTATTCT CCTCTCCCCT GTCCTTCCTT
```

```
                                          ..................................................
 InII-12pL1          #151                 AGGAGGTACC CGCGTATGGC TGATTATTCT CCTCTCCCCT GTCCTTCCTT


<Input_InII-12pL1    #201                 CCCTCCCGCC GGCCAACCGC ACCCCACCTT CGTTCCACGG CATTTGCCAG
>InII-12pL1_S4       #201                 CCCTCCCGCC GGCCAACCGC ACCCCACCTT CGTTCCACGG CATTTGCCAG
>InII-12pL1_S5       #201                 CCCTCCCGCC GGCCAACCGC ACCCCACCTT CGTTCCACGG CATTTGCCAG
>InII-12pL1_S6       #201                 CCCTCCCGCC GGCCAACCGC ACCCCACCTT CGTTCCACGG CATTTGCCAG
>InII-12pL1_S12      #201                 CCCTCCCGCC GGCCAACCGC ACCCCACCTT CGTTCCACGG CATTTGCCAG
>InII-12pL1_S13      #201                 CCCTCCCGCC GGCCAACCGC ACCCCACCTT CGTTCCACGG CATTTGCCAG
                                          ..................................................
 InII-12pL1          #201                 CCCTCCCGCC GGCCAACCGC ACCCCACCTT CGTTCCACGG CATTTGCCAG


<Input_InII-12pL1    #251                 CTGTCTACCC GGTCTATTTG GTCACAAGTC GCTTGTCTCG GATCCGAGAC
>InII-12pL1_S4       #251                 CTGTCTACCC GGTCTATTTG GTCACAAGTC GCTTGTCTCG GATCCGAGAC
>InII-12pL1_S5       #251                 CTGTCTACCC GGTCTATTTG GTCACAAGTC GCTTGTCTCG GATCCGAGAC
>InII-12pL1_S6       #251                 CTGTCTACCC GGTCTATTTG GTCACAAGTC GCTTGTCTCG GATCCGAGAC
>InII-12pL1_S12      #251                 CTGTCTACCC GGTCTATTTG GTCACAAGTC GCTTGTCTCG GATCCGAGAC
>InII-12pL1_S13      #251                 CTGTCTACCC GGTCTATTTG GTCACAAGTC GCTTGTCTCG GATCCGAGAC
                                          ..................................................
 InII-12pL1          #251                 CTGTCTACCC GGTCTATTTG GTCACAAGTC GCTTGTCTCG GATCCGAGAC


<Input_InII-12pL1    #301                 AAGCTCCGTC CAAGCACTTA AAAGCACGTG GAGAGTTGGG TGGGGAGGCA
>InII-12pL1_S4       #301                 AAGCTCCGTC CAAGCACTTA AAAGCACGTG GAGAGTTGGG TGGGGAGGCA
>InII-12pL1_S5       #301                 AAGCTCCGTC CAAGCACTTA AAAGCACGTG GAGAGTTGGG TGGGGAGGCA
>InII-12pL1_S6       #301                 AAGCTCCGTC CAAGCACTTA AAAGCACGTG GAGAGTTGGG TGGGGAGGCA
>InII-12pL1_S12      #301                 AAGCTCCGTC CAAGCACTTA AAAGCACGTG GAGAGTTGGG TGGGGAGGCA
>InII-12pL1_S13      #301                 AAGCTCCGTC CAAGCACTTA AAAGCACGTG GAGAGTTGGG TGGGGAGGCA
                                          ..................................................
 InII-12pL1          #301                 AAGCTCCGTC CAAGCACTTA AAAGCACGTG GAGAGTTGGG TGGGGAGGCA


<Input_InII-12pL1    #351                 GGAGGGGAGG ATTCAACCAT ACGTAGCAAT ATGGAATTCA AGCATCCGAG
>InII-12pL1_S4       #351                 GGAGGGGAGG ATTCAACCAT ACGTAGCAAT ATGGAATTCA AGCATCCGAG
>InII-12pL1_S5       #351                 GGAGGGGAGG ATTCAACCAT ACGTAGCAAT ATGGAATTCA AGCATCCGAG
>InII-12pL1_S6       #351                 GGAGGGGAGG ATTCAACCAT ACGTAGCAAT ATGGAATTCA AGCATCCGAG
>InII-12pL1_S12      #351                 GGAGGGGAGG ATTCAACCAT ACGTAGCAAT ATGGAATTCA AGCATCCGAG
```

```
>InII-12pL1_S13     #351              GGAGGGGAGG ATTCAACCAT ACGTAGCAAT ATGGAATTCA AGCATCCGAG
                                      ..................................................
 InII-12pL1         #351              GGAGGGGAGG ATTCAACCAT ACGTAGCAAT ATGGAATTCA AGCATCCGAG


<Input_InII-12pL1   #401              AGTCTATGGA GAAAGGTACC TCCTCCTGAA TTCAAGCTT
>InII-12pL1_S4      #401              AGTCTATGGA GAAAGGTACC TCCTCCTGAA TTCAAGCTT
>InII-12pL1_S5      #401              AGTCTATGGA GAAAGGTACC TCCTCCTGAA TTCAAGCTT
>InII-12pL1_S6      #401              AGTCTATGGA GAAAGGTACC TCCTCCTGAA TTCAAGCTT
>InII-12pL1_S12     #401              AGTCTATGGA GAAAGGTACC TCCTCCTGAA TTCAAGCTT
>InII-12pL1_S13     #401              AGTCTATGGA GAAAGGTACC TCCTCCTGAA TTCAAGCTT
                                      .................................................
 InII-12pL1         #401              AGTCTATGGA GAAAGGTACC TCCTCCTGAA TTCAAGCTT
```

DNA Analysis of *Arabidopsis thaliana* transformed with InIII-50pL1 sequence

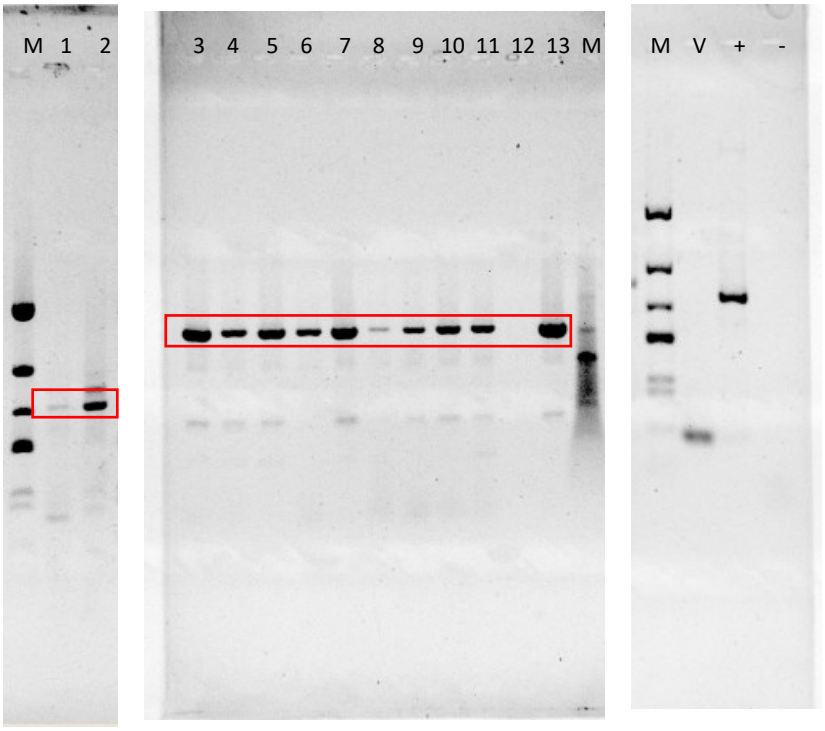


Figure A2.9: Example of agarose gel electrophoresis showing PCR amplification of clone InIII-50 in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate InIII-50pL1 insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 625bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-13 : Transgenic plant DNA containing InIII-50pL1 inserts (625bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "InIII-50pL1" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_InIII-50pL1  #1              AAGCTTGAAT TCAGGAGGAG GCCCTCTGCT CGACCCGCGT CCGTACCACG
<III-50pL1_S1       #1              AAGCTTGAAT TCAGGAGGAG GCCCTCTGCT CGACCCGCGT CCGTACCACG
<III-50pL1_S2       #1              AAGCTTGAAT TCAGGAGGAG GCCCTCTGCT CGACCCGCGT CCGTACCACG
<III-50pL1_S3       #1              AAGCTTGAAT TCAGGAGGAG GCCCTCTGCT CGACCCGCGT CCGTACCACG
<III-50pL1_S5       #1              AAGCTTGAAT TCAGGAGGAG GCCCTCTGCT CGACCCGCGT CCGTACCACG
>III-50pL1_S5       >#1>                           GAGGAG GCCCTCTGCT CGACCCGCGT CCGTACCACG
                                    ..................................................
 InIII-50pL1        #1              AAGCTTGAAT TCAGGAGGAG GCCCTCTGCT CGACCCGCGT CCGTACCACG
 Consensus

>Input_InIII-50pL1  #51             CACATAGCGC GTACACCAGG ATGTTGCTAC TGGCGGCATG CACTCGCTTT
<III-50pL1_S1       #51             CACATAGCGC GTACACCAGG ATGTTGCTAC TGGCGGCATG CACTCGCTTT
<III-50pL1_S2       #51             CACATAGCGC GTACACCAGG ATGTTGCTAC TGGCGGCATG CACTCGCTTT
<III-50pL1_S3       #51             CACATAGCGC GTACACCAGG ATGTTGCTAC TGGCGGCATG CACTCGCTTT
<III-50pL1_S5       #51             CACATAGCGC GTACACCAGG ATGTTGCTAC TGGCGGCATG CACTCGCTTT
>III-50pL1_S5       #37             CACATAGCGC GTACACCAGG ATGTTGCTAC TGGCGGCATG CACTCGCTTT
                                    ..................................................
 InIII-50pL1        #51             CACATAGCGC GTACACCAGG ATGTTGCTAC TGGCGGCATG CACTCGCTTT


>Input_InIII-50pL1  #101            ATTTCGTGCA TTTGTGCCCC TCCCGTCCTC GCGTAATCTG CTTGTCTCGG
<III-50pL1_S1       #101            ATTTCGTGCA TTTGTGCCCC TCCCGTCCTC GCGTAATCTG CTTGTCTCGG
<III-50pL1_S2       #101            ATTTCGTGCA TTTGTGCCCC TCCCGTCCTC GCGTAATCTG CTTGTCTCGG
<III-50pL1_S3       #101            ATTTCGTGCA TTTGTGCCCC TCCCGTCCTC GCGTAATCTG CTTGTCTCGG
<III-50pL1_S5       #101            ATTTCGTGCA TTTGTGCCCC TCCCGTCCTC GCGTAATCTG CTTGTCTCGG
>III-50pL1_S5       #87             ATTTCGTGCA TTTGTGCCCC TCCCGTCCTC GCGTAATCTG CTTGTCTCGG
                                    ..................................................
 InIII-50pL1        #101            ATTTCGTGCA TTTGTGCCCC TCCCGTCCTC GCGTAATCTG CTTGTCTCGG


>Input_InIII-50pL1  #151            ATCCGAGACA AGTAACGGTT CCGAGGTGGC TCGGTCACGC AGGAGGTCGC
<III-50pL1_S1       #151            ATCCGAGACA AGTAACGGTT CCGAGGTGGC TCGGTCACGC AGGAGGTCGC
<III-50pL1_S2       #151            ATCCGAGACA AGTAACGGTT CCGAGGTGGC TCGGTCACGC AGGAGGTCGC
<III-50pL1_S3       #151            ATCCGAGACA AGTAACGGTT CCGAGGTGGC TCGGTCACGC AGGAGGTCGC
<III-50pL1_S5       #151            ATCCGAGACA AGTAACGGTT CCGAGGTGGC TCGGTCACGC AGGAGGTCGC
>III-50pL1_S5       #137            ATCCGAGACA AGTAACGGTT CCGAGGTGGC TCGGTCACGC AGGAGGTCGC
```

```
                                          ..................................................
 InIII-50pL1        #151                  ATCCGAGACA AGTAACGGTT CCGAGGTGGC TCGGTCACGC AGGAGGTCGC


>Input_InIII-50pL1  #201                  GCAAGGACGA CGCGTAGAAA CGCTTTACAC GGGAAATATG CGACCGTGGC
<III-50pL1_S1       #201                  GCAAGGACGA CGCGTAGAAA CGCTTTACAC GGGAAATATG CGACCGTGGC
<III-50pL1_S2       #201                  GCAAGGACGA CGCGTAGAAA CGCTTTACAC GGGAAATATG CGACCGTGGC
<III-50pL1_S3       #201                  GCAAGGACGA CGCGTAGAAA CGCTTTACAC GGGAAATATG CGACCGTGGC
<III-50pL1_S5       #201                  GCAAGGACGA CGCGTAGAAA CGCTTTACAC GGGAAATATG CGACCGTGGC
>III-50pL1_S5       #187                  GCAAGGACGA CGCGTAGAAA CGCTTTACAC GGGAAATATG CGACCGTGGC
                                          ..................................................
 InIII-50pL1        #201                  GCAAGGACGA CGCGTAGAAA CGCTTTACAC GGGAAATATG CGACCGTGGC


>Input_InIII-50pL1  #251                  GAGTTGTTAG CTGAGTACAA AGGCCCGCAA AGTCGGCCTC CTCCTGAATT
<III-50pL1_S1       #251                  GAGTTGTTAG CTGAGTACAA AGGCCCGCAA AGTCGGCCTC CTCCTGAATT
<III-50pL1_S2       #251                  GAGTTGTTAG CTGAGTACAA AGGCCCGCAA AGTCGGCCTC CTCCTGAATT
<III-50pL1_S3       #251                  GAGTTGTTAG CTGAGTACAA AGGCCCGCAA AGTCGGCCTC CTCCTGAATT
<III-50pL1_S5       #251                  GAGTTGTTAG CTGAGTACAA AGGCCCGCAA AGTCGGCCTC CTCCTGAATT
>III-50pL1_S5       #237                  GAGTTGTTAG CTGAGTACAA AGGCCCGCAA AGTCGGCCTC CTCCTGAATT
                                          ..................................................
 InIII-50pL1        #251                  GAGTTGTTAG CTGAGTACAA AGGCCCGCAA AGTCGGCCTC CTCCTGAATT


>Input_InIII-50pL1  #301                  CAGGAGGAGG TCTCCGCGTG GCCTTACTTT TCCCGTCGCA CTGCGACGCC
<III-50pL1_S1       #301                  CAGGAGGAGG TCTCCGCGTG GCCTTACTTT TCCCGTCGCA CTGCGACGCC
<III-50pL1_S2       #301                  CAGGAGGAGG TCTCCGCGTG GCCTTACTTT TCCCGTCGCA CTGCGACGCC
<III-50pL1_S3       #301                  CAGGAGGAGG TCTCCGCGTG GCCTTACTTT TCCCGTCGCA CTGCGACGCC
<III-50pL1_S5       #301                  CAGGAGGAGG TCTCCGCGTG GCCTTACTTT TCCCGTCGCA CTGCGACGCC
>III-50pL1_S5       #287                  CAGGAGGAGG TCTCCGCGTG GCCTTACTTT TCCCGTCGCA CTGCGACGCC
                                          ..................................................
 InIII-50pL1        #301                  CAGGAGGAGG TCTCCGCGTG GCCTTACTTT TCCCGTCGCA CTGCGACGCC


>Input_InIII-50pL1  #351                  TCAACCATAT CGGGCCTCCT CCTTCTGTTT ATCTCTCACT TCCTCTCCTT
<III-50pL1_S1       #351                  TCAACCATAT CGGGCCTCCT CCTTCTGTTT ATCTCTCACT TCCTCTCCTT
<III-50pL1_S2       #351                  TCAACCATAT CGGGCCTCCT CCTTCTGTTT ATCTCTCACT TCCTCTCCTT
<III-50pL1_S3       #351                  TCAACCATAT CGGGCCTCCT CCTTCTGTTT ATCTCTCACT TCCTCTCCTT
<III-50pL1_S5       #351                  TCAACCATAT CGGGCCTCCT CCTTCTGTTT ATCTCTCACT TCCTCTCCTT
```

```
>III-50pL1_S5       #337              TCAACCATAT CGGGCCTCCT CCTTCTGTTT ATCTCTCACT TCCTCTCCTT
                                      ..................................................
 InIII-50pL1        #351              TCAACCATAT CGGGCCTCCT CCTTCTGTTT ATCTCTCACT TCCTCTCCTT


>Input_InIII-50pL1  #401              CTATCACGTG GCGCCTGCGC GCCCGAGCTT GCCGGCTTGT CTCGGATCCC
<III-50pL1_S1       #401              CTATCACGTG GCGCCTGCGC GCCCGAGCTT GCCGGCTTGT CTCGGATCCC
<III-50pL1_S2       #401              CTATCACGTG GCGCCTGCGC GCCCGAGCTT GCCGGCTTGT CTCGGATCCC
<III-50pL1_S3       #401              CTATCACGTG GCGCCTGCGC GCCCGAGCTT GCCGGCTTGT CTCGGATCCC
<III-50pL1_S5       #401              CTATCACGTG GCGCCTGCGC GCCCGAGCTT GCCGGCTTGT CTCGGATCCC
>III-50pL1_S5       #387              CTATCACGTG GCGCCTGCGC GCCCGAGCTT GCCGGCTTGT CTCGGATCCC
                                      ..................................................
 InIII-50pL1        #401              CTATCACGTG GCGCCTGCGC GCCCGAGCTT GCCGGCTTGT CTCGGATCCC


>Input_InIII-50pL1  #451              TGCAG
<III-50pL1_S1       #451              TGCAG
<III-50pL1_S2       #451              TGCAG
<III-50pL1_S3       #451              TGCAG
<III-50pL1_S5       #451              TGCAG
>III-50pL1_S5       #437              TGCAG
                                      .....................................................
 InIII-50pL1        #451              TGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with InIII-78_5'pL1 sequence

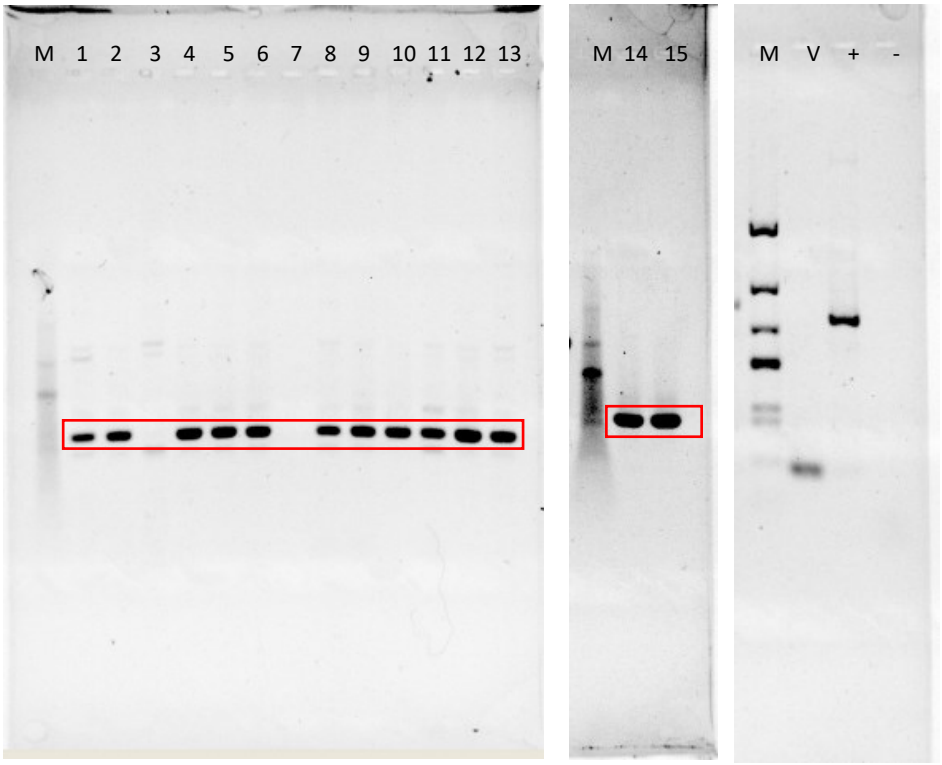


Figure A2.10: Example of agarose gel electrophoresis showing PCR amplification of clone InIII-78_5' in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate InIII-78_5'pL1 insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 323bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-15 : Transgenic plant DNA containing InIII-78_5'pL1 inserts (323bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "InIII-78_5'pL1" samples and consensus sequence output from Sequencher 4.10.1

```
<Input_InIII-78_5'  #1                  AAGCTTGTCT CGGATCCGAG ACAAGCACAA CGGCACAGAC TAGTCCCGGG
>III-78-5'_S4       >#1>                 AGCTTGTCT CGGATCCGAG ACAAGCACAA CGGCACAGAC TAGTCCCGGG
>InIII-78_5'_S8     #1                  AAGCTTGTCT CGGATCCGAG ACAAGCACAA CGGCACAGAC TAGTCCCGGG
>InIII-78_5'_S12    #1                  CAGCTTGTCT CGGATCCGAG ACAAGCACAA CGGCACAGAC TAGTCCCGGG
>InIII-78_5'_S13    #1                  CAGCTTGTCT CGGATCCGAG ACAAGCACAA CGGCACAGAC TAGTCCCGGG
>InIII-78_5'_S15    #1                  CAGCTTGTCT CGGATCCGAG ACAAGCACAA CGGCACAGAC TAGTCCCGGG
                                        ..................................................
 InIII-78_5'pL1     #1                  CAGCTTGTCT CGGATCCGAG ACAAGCACAA CGGCACAGAC TAGTCCCGGG
 Consensus                                      *

<Input_InIII-78_5'  #51                 GGAGACGCCC AGGAGCAACA AATGGAAGAG GCACTAAGCT AGAATGGAAG
>III-78-5'_S4       #50                 GGAGACGCCC AGGAGCAACA AATGGAAGAG GCACTAAGCT AGAATGGAAG
>InIII-78_5'_S8     #51                 GGAGACGCCC AGGAGCAACA AATGGAAGAG GCACTAAGCT AGAATGGAAG
>InIII-78_5'_S12    #51                 GGAGACGCCC AGGAGCAACA AATGGAAGAG GCACTAAGCT AGAATGGAAG
>InIII-78_5'_S13    #51                 GGAGACGCCC AGGAGCAACA AATGGAAGAG GCACTAAGCT AGAATGGAAG
>InIII-78_5'_S15    #51                 GGAGACGCCC AGGAGCAACA AATGGAAGAG GCACTAAGCT AGAATGGAAG
                                        ..................................................
 InIII-78_5'pL1     #51                 GGAGACGCCC AGGAGCAACA AATGGAAGAG GCACTAAGCT AGAATGGAAG


<Input_InIII-78_5'  #101                ATTGATCCGG AGCGGAGCTG CATGAAATCC AAACGAACCC TCCTCCTGAA
>III-78-5'_S4       #100                ATTGATCCGG AGCGGAGCTG CATGAAATCC AAACGAACCC TCCTCCTGAA
>InIII-78_5'_S8     #101                ATTGATCCGG AGCGGAGCTG CATGAAATCC AAACGAACCC TCCTCCTGAA
>InIII-78_5'_S12    #101                ATTGATCCGG AGCGGAGCTG CATGAAATCC AAACGAACCC TCCTCCTGAA
>InIII-78_5'_S13    #101                ATTGATCCGG AGCGGAGCTG CATGAAATCC AAACGAACCC TCCTCCTGAA
>InIII-78_5'_S15    #101                ATTGATCCGG AGCGGAGCTG CATGAAATCC AAACGAACCC TCCTCCTGAA
                                        ..................................................
 InIII-78_5'pL1     #101                ATTGATCCGG AGCGGAGCTG CATGAAATCC AAACGAACCC TCCTCCTGAA


<Input_InIII-78_5'  #151                TTC
>III-78-5'_S4       #150                TTC
>InIII-78_5'_S8     #151                TTC
>InIII-78_5'_S12    #151                TTC
>InIII-78_5'_S13    #151                TTC
>InIII-78_5'_S15    #151                TTC
                                        ...
 InIII-78_5'pL1     #151                TTC
```

Table A2.1: GUS staining in different tissues of individual *Arabidopsis thaliana* transgenic lines. The intensity was assessed visually using the following scale: 3 (strong), 2 (medium), 1 (week), and 0 (none), as shown in Figure 2.2A. Data represents scoring of tissues in pB31 vector for all sequences tested.

Construct Name	Size (bp)	Sample #	Flower	Leaf	Silique 1	Silique 2
pB31 control		1	3	3	3	0
		2	0	0	0	0
		3	3	0	2	0
		4	3	0	3	0
		5	0	3	3	0
		6	0	3	3	1
		7	0	0	0	2
		8	3	0	3	1
		9	0	3	3	0
		10	0	0	0	0
		11	0	2	1	0
		12	0	1	0	0
		13	0	2	2	0
		14	0	1	0	0
		15	0	3	2	0
		16	0	1	1	0
		17	1	3	3	0
		18	0	1	1	0
		19	1	3	3	0
		20	0	1	1	0
		21	2	3	3	0
		22	0	0	0	0
		23	0	0	0	0
		24	0	1	1	0
		25	1	3	3	0
		26	0	0	0	0
		27	0	0	0	0
		28	0	0	0	0
		29	0	3	0	0
		30	0	2	3	0
		31	0	3	3	0
		32	0	1	1	0
		33	0	3	3	0
		34	2	3	3	0
		35	0	1	0	0
		36	0	0	0	0
		37	1	3	3	0
		38	0	1	0	0
		39	3	3	3	0
		40	0	1	1	0
		41	0	1	0	0
		42	0	2	2	0
		43	0	0	0	0

		44	0	0	0	0
		45	0	0	0	0
		46	0	3	2	2
		47	0	0	0	0
		48	0	0	0	0
		49	0	3	2	1
		50	0	0	0	1
		51	0	0	0	0
		52	0	2	1	1
		53	0	3	0	0
		54	0	0	0	0
		55	0	3	1	1
		56	0	1	0	0
		57	0	1	1	0
		58	0	0	0	0
		59	0	0	0	1
		60	0	3	0	3
		61	0	3	1	0
		62	0	0	1	0
		Total	11	39	34	10
		Staining:	46			
		None:	16			
CLO I-2 (pB31: pCAM 1300-35S46-GUS)	446	1	1	0	0	0
		2	1	0	0	0
		3	1	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	0	0	0	0
		Total	3	0	0	0
		Stained:	3			
		None:	3			
CLO I-3 (pB31: pCAM 1300-35S46-GUS)	438	1	0	0	0	0
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	0	0	0	0
		7	0	0	0	0
		8	0	0	0	0
		9	0	0	0	0
		10	0	0	0	0
		11	0	0	0	0
		12	0	0	0	0
		13	0	0	0	0
		14	0	0	0	0
		15	0	0	0	0
		16	0	0	0	0
		17	0	0	0	0
		18	0	0	0	0
		19	0	0	0	0
		20	0	0	0	0

		21	0	0	0	0
		22	0	0	0	0
		23	0	0	0	0
		24	0	0	0	0
		25	0	0	0	0
		26	0	0	0	0
		27	0	0	0	0
		28	0	0	0	0
		29	0	0	0	0
		30	0	0	0	0
		31	0	0	0	0
		32	0	0	0	0
		33	0	0	0	0
		34	0	0	0	0
		35	0	0	0	0
		36	0	0	0	0
		37	0	0	0	0
		38	0	0	0	0
		39	0	0	0	0
		40	0	0	0	0
		41	0	0	0	0
		42	0	0	0	0
		43	0	0	0	0
		44	0	0	0	0
		45	0	0	0	0
		46	0	0	0	0
		47	0	0	0	0
		48	0	0	0	0
		49	0	0	0	0
		50	0	0	0	0
		51	0	0	0	0
		Total	0	0	0	0
		Stained:	0			
		None:	51			
CLOI-6 (pB31: pCAM 1300-35S46-GUS)	438	1	0	0	0	0
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	0	0	0	0
		7	0	0	0	0
		8	0	0	0	0
		9	0	0	0	0
		10	0	0	0	0
		11	0	0	0	0
		12	0	0	0	0
		13	0	0	0	0
		14	0	0	0	0
		15	0	0	0	0
		16	0	0	0	0
		17	0	0	0	0
		18	0	0	0	0

		19	0	0	0	0
		20	0	0	0	0
		21	0	0	0	0
		22	0	0	0	0
		23	0	0	0	0
		24	0	0	0	0
		Total	0	0	0	0
		Stained:	0			
		None:	24			
CLO II-3 (pB31: pCAM 1300-35S46-GUS)	151	1	1	2	3	0
		2	1	0	1	0
		3	1	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	0	1	3	3
		7	1	0	1	0
		8	0	2	1	2
		9	1	2	3	3
		10	1	0	0	0
		11	0	2	3	3
		12	0	0	0	0
		13	1	3	0	0
		14	0	1	0	0
		15	1	2	2	0
		16	1	0	1	0
		17	0	0	0	0
		Total	9	6	9	4
		Stained:	13			
		None:	4			
CLO II-7 (pB31: pCAM 1300-35S46-GUS)	~150	1	0	0	0	0
		2	0	1	1	2
		3	0	1	1	1
		4	0	1	1	0
		5	1	0	0	1
		6	0	2	0	0
		7	0	0	0	0
		8	0	0	0	0
		9	0	1	0	0
		10	0	0	0	0
		11	0	0	0	0
		12	0	0	0	0
		13	0	0	0	0
		14	1	0	0	0
		15	0	0	0	0
		16	0	0	0	0
		17	0	1	0	0
		18	1	2	0	1
		19	1	2	0	1
		20	1	0	0	1
		21	1	1	0	0
		22	1	0	0	0
		23	0	1	0	1

		24	1	1	0	0
		25	0	0	0	0
		26	1	0	0	0
		27	0	0	0	0
		28	1	1	0	0
		29	1	1	0	0
		30	0	1	0	0
		Total	11	14	3	7
		Stained:	19			
		None:	11			
CLO II-10 (pB31: pCAM 1300-35S46-GUS)	154	1	0	0	0	0
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	0	0	0	0
		7	0	0	0	0
		8	0	0	0	0
		9	0	0	0	0
		10	0	0	0	0
		11	0	0	0	0
		12	0	0	0	0
		13	0	0	0	0
		14	0	0	0	0
		15	0	0	0	0
		16	0	0	0	0
		17	0	0	0	0
		18	0	0	0	0
		19	0	0	0	0
		20	0	0	0	0
		21	0	0	0	0
		22	0	0	0	0
		23	0	0	0	0
		24	0	0	0	0
		25	0	0	0	0
		26	0	0	0	0
		27	0	0	0	0
		28	0	0	0	0
		29	0	0	0	0
		30	0	0	0	0
		31	0	0	0	0
		32	0	0	0	0
		33	0	0	0	0
		34	0	0	0	0
		35	0	0	0	0
		36	0	0	0	0
		37	0	0	0	0
		38	0	0	0	0
		39	0	0	0	0
		40	0	0	0	0
		Total	0	0	0	0

		Stained:	0			
		None:	40			
CLOII-12 (pB31: pCAM 1300-35S46-GUS)		1	0	0	1	1
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	0	0	0	0
		7	0	0	0	0
		8	0	0	1	1
		9	0	0	0	0
		10	0	0	0	0
		11	0	0	0	0
		12	0	0	0	0
		13	0	0	0	0
		14	0	0	0	0
		15	0	0	0	0
		16	0	0	0	0
		17	0	0	0	0
		18	0	0	0	0
		19	0	0	0	0
		20	0	0	0	0
		21	0	0	0	0
		22	0	0	0	1
		23	0	0	0	0
		24	0	0	0	0
		25	0	0	0	0
		26	0	0	0	0
		27	0	0	0	0
		28	0	0	0	0
		29	0	0	0	0
		30	0	0	0	0
		31	0	0	0	0
		32	0	0	0	0
		33	0	0	0	0
		34	0	0	0	0
		35	0	0	0	0
		36	0	0	0	0
		37	0	0	0	0
		Total	0	0	2	3
		Stained:	3			
		None:	34			
CLO III-4 (pB31: pCAM 1300-35S46-GUS)	440	1	1	1	0	0
		2	2	1	0	0
		3	1	0	0	0
		4	1	0	0	0
		5	1	0	0	0
		6	1	0	0	0
		7	0	0	0	0
		8	0	0	0	0
		9	1	0	0	0
		10	1	0	0	0

		11	1	0	0	1
		12	0	0	0	0
		13	1	1	0	0
		14	1	0	0	0
		15	1	0	0	0
		16	1	0	0	0
		17	0	1	0	1
		18	1	0	0	0
		19	0	0	0	0
		20	1	0	0	0
		21	1	0	0	1
		22	0	0	0	0
		23	0	0	0	0
		24	0	0	0	0
		25	0	0	0	0
		26	0	0	0	0
		Total	16	4	0	3
		Stained	17			
		None	9			
CLO III-17 (pB31: pCAM 1300-35S46-GUS)	443	1	1	1	0	0
		2	0	1	0	0
		3	0	0	0	0
		4	1	1	0	0
		5	1	0	0	0
		6	1	1	0	0
		7	1	1	0	0
		8	1	1	0	1
		9	1	1	0	0
		10	1	0	0	0
		11	0	0	0	0
		12	1	0	0	1
		13	1	1	0	0
		14	1	1	0	0
		15	1	0	0	0
		16	1	0	0	0
		17	1	0	0	0
		18	1	2	0	0
		19	1	0	0	0
		20	1	0	0	0
		21	1	1	0	0
		22	1	1	0	0
		23	1	0	0	0
		24	1	0	0	0
		25	0	0	0	0
		26	0	0	0	0
		27	1	0	0	0
		28	1	1	0	0
		29	1	0	0	0
		30	1	1	0	0
		Total	25	15	0	2
		Stained:	26			

		None:	4			
CLO III-22 (pB31: pCAM 1300-35S46-GUS)	425	1	0	0	0	0
		2	0	1	0	0
		3	0	0	1	0
		4	0	0	0	0
		5	1	0	0	0
		6	1	0	0	0
		7	1	0	0	0
		8	1	0	0	0
		9	1	0	0	0
		10	1	0	0	0
		11	0	0	0	0
		12	1	0	0	0
		13	0	0	0	0
		14	1	0	0	0
		15	1	0	0	0
		16	1	0	0	0
		17	1	0	0	0
		18	1	0	0	0
		19	1	1	0	0
		20	1	1	0	0
		21	2	0	0	0
		22	1	1	0	0
		23	1	1	0	0
		24	1	1	0	0
		25	1	1	0	0
		26	1	2	0	0
		27	1	1	0	0
		28	1	1	0	0
		29	1	0	0	1
		30	1	0	0	0
		31	0	0	0	0
		32	0	0	0	1
		33	1	0	0	0
		34	0	0	0	0
		35	0	0	0	0
		36	0	1	0	0
		Total	26	11	1	2
		Stained:	29			
		None:	7			
CLO III-27 (pB31: pCAM 1300-35S46-GUS)	128	1	1	0	0	0
		2	1	0	0	0
		3	1	1	1	0
		4	1	0	0	0
		5	1	1	0	0
		6	1	0	1	0
		7	1	2	0	0
		8	1	0	0	0
		9	1	0	0	0
		10	1	3	0	0
		11	1	0	0	0
		12	1	0	0	0

		13	1	0	0	0
		Total	13	4	2	0
		Stained:	13			
		None:	0			
CLOIII-50 (pB31: pCAM 1300-35S46-GUS)		1	0	0	0	0
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	0	0	0	0
		7	0	0	0	0
		8	0	0	0	0
		9	0	0	0	0
		10	0	0	0	0
		11	0	0	0	0
		Total	0	0	0	0
		Stained:	0			
		None:	11			
CLO III-52 (pB31: pCAM 1300-35S46-GUS)	446	1	0	0	0	0
		2	2	1	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	0	0	0	0
		7	0	0	0	0
		8	0	0	0	0
		9	1	0	0	0
		10	0	0	0	0
		11	0	0	0	0
		12	0	0	0	0
		13	1	0	0	0
		14	0	0	0	0
		15	0	0	0	0
		Total	3	1	0	0
		Stained:	3			
		None:	12			
CLO III-53 (pB31: pCAM 1300-35S46-GUS)	152	1	1	2	1	0
		2	3	3	2	2
		3	3	3	2	0
		4	3	1	1	1
		5	1	2	1	1
		6	0	2	1	0
		7	2	3	3	2
		8	2	1	1	2
		9	1	2	1	1
		10	2	2	1	2
		11	1	1	1	0
		12	3	3	3	0
		13	2	2	2	1
		14	1	1	0	0

		15	0	0	0	0
		Total	13	14	13	8
		Stained:	14			
		None:	1			
CLO III-55 (pB31: pCAM 1300-35S46-GUS)	432	1	1	0	0	0
		2	1	0	0	0
		3	1	0	0	0
		4	1	0	0	0
		5	1	0	0	0
		6	1	0		0
		7	1	1	0	0
		8	1	0	0	0
		9	1	0	0	0
		10	1	0	1	0
		11	1	0	0	0
		12	1	1	1	0
		13	1	0	0	0
		14	1	0	0	0
		15	1	0	0	0
		16	1	2	0	0
		17	1	0	0	0
		18	1	1	0	0
		19	1	1	0	0
		20	1	0	0	0
		21	1	0	0	0
		22	1	0	0	0
		23	0	0	0	0
		Total	22	5	2	0
		Stained:	22			
		None:	1			
CLO III-57 (pB31: pCAM 1300-35S46-GUS)	154	1	0	0	0	0
		2	0	0	0	0
		3	0	0	1	1
		4	1	1	1	1
		5	1	1	1	1
		6	0	1	1	1
		7	0	0	0	0
		8	0	0	1	1
		9	0	0	1	1
		10	0	2	1	1
		11	0	1	1	0
		12	0	1	1	1
		13	0	1	0	0
		14	0	1	0	0
		15	0	0	0	0
		16	0	0	0	0
		Total	2	8	9	8
		Stained:	11			
		None:	5			
CLO III-58 (pB31: pCAM 1300-35S46-GUS)	150	1	0	0	0	0

		2	0	0	0	0
		3	0	0	0	0
		Total	0	0	0	0
		Stained:	0			
		None:	3			
CLO III-63 (pB31: pCAM 1300-35S46-GUS)	154	1	0	0	0	0
		2	0	0	0	0
		3	0	0	3	1
		4	0	1	2	1
		5	0	0	1	1
		6	2	1	3	1
		7	1	1	2	0
		Total	2	3	5	4
		Stained:	5			
		None:	2			
CLO III-74 (pB31: pCAM 1300-35S46-GUS)	153	1	1	1	1	1
		2	1	1	1	1
		3	1	1	1	1
		4	0	1	1	1
		5	0	0	0	0
		6	0	0	0	0
		Total	3	4	4	4
		Stained:	4			
		None:	2			
CLO III-78 (pB31: pCAM 1300-35S46-GUS)		1	0	0	0	0
		2	0	0	0	0
		3	0	0	0	0
		4	0	0	0	0
		5	0	0	0	0
		6	0	0	0	0
		7	0	0	0	0
		8	0	0	0	0
		9	0	0	0	0
		10	0	0	0	0
		11	0	0	0	0
		12	0	0	0	0
		13	0	0	0	0
		14	0	0	0	0
		15	0	0	0	0
		16	0	0	0	0
		17	0	0	0	0
		18	0	0	0	0
		19	0	0	0	0
		20	0	0	0	0
		Total	0	0	0	0
		Stained:	0			
		None:	20			
CLO III-80 (pB31: pCAM 1300-35S46-GUS)	451	1	2	0	0	0
		2	0	0	0	0

	3	0	0	0	0
	4	0	0	0	0
	5	0	0	0	0
	6	0	0	0	0
	7	0	0	0	0
	8	0	0	0	0
	9	0	0	0	0
Total		1	0	0	0
Stained:		1			
None:		8			

Table A2.2: GUS staining in different tissues of individual *Arabidopsis thaliana* transgenic lines. The intensity was assessed visually using the following scale: 3 (strong), 2 (medium), 1 (week), and 0 (none), as shown in Figure 2.2A. Data represents scoring of tissues in pL1 vector for all sequences tested.

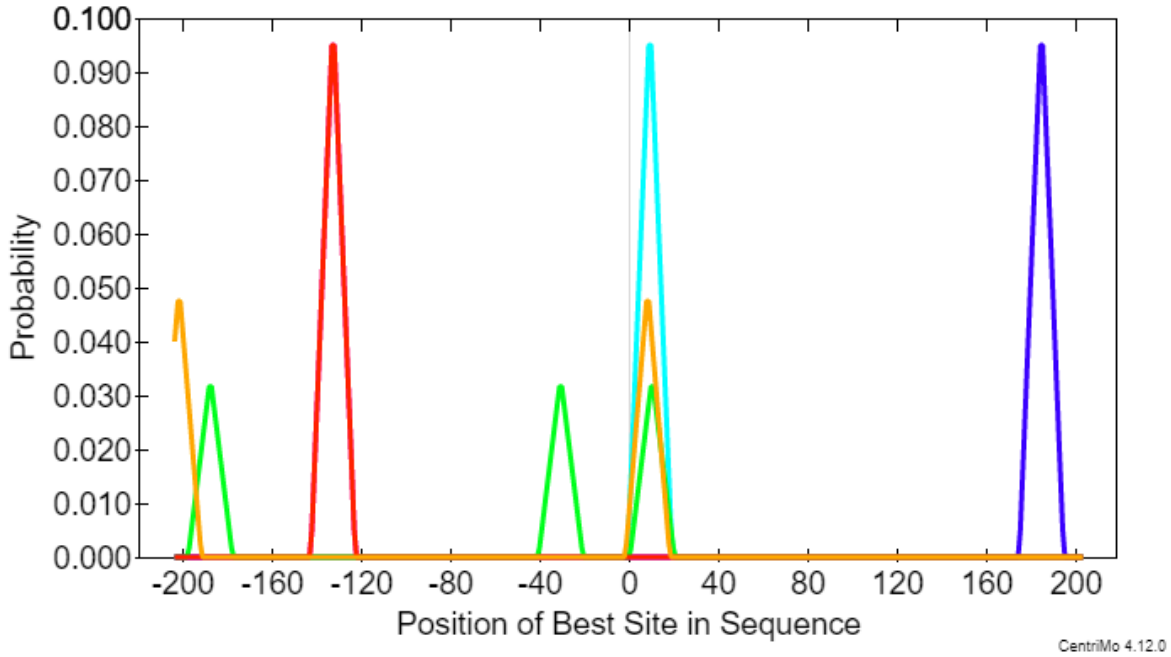
Construct	Size (bp)	Sample #	Flower	Leaf	Silique 1	Seed 2	Silique 2	Seed 2
pL1 control: ter35S-hptII-P35S←→Pnapin-GUS-ternos		1	3	4	4	0	3	1
		2	4	3	4	0	4	2
		3	0	1	0	1	0	0
		4	0	1	0	0	0	1
		5	3	4	3	2	4	1
		6	1	3	1	1	1	2
		7	0	0	0	0	0	0
		8	2	4	4	1	3	1
		9	0	1	0	2	0	1
		10	3	3	1	2	1	1
		11	4	4	4	1	4	1
		12	4	4	3	1	3	2
		13	3	4	4	1	3	0
		14	2	2	1	1	1	1
		15	3	4	4	1	4	1
		16	3	4	3	1	3	1
		17	4	4	4	1	2	1
		18	0	1	1	1	0	1
		19	3	3	2	1	2	2
		20	3	3	4	2	3	1
		21	4	4	4	2	3	1
		22	2	4	3	1	2	1
		23	1	2	2	2	2	3
		24	4	4	4	3	3	1
		25	0	1	0	0	0	1
		26	1	1	1	0	0	2
		27	0	0	0	0	0	2
		28	2	3	2	2	1	2
		29	2	3	2	2	4	1
		30	3	3	4	1	3	1
		31	0	3	1	0	0	2
		32	0	3	4	0	0	1
		33	0	0	0	1	0	0
		34	4	4	4	4	4	4
		35	2	4	1	0	1	0
		36	1	2	0	0	0	0
		37	4	4	4	0	4	0
		38	3	3	4	0	3	0
		39	3	3	4	0	2	0
		40	4	4	4	1	2	0
		41	1	2	4	1	1	0
		42	2	4	4	1	3	0
		43	1	2	1	0	1	0
		44	1	3	1	0	2	0

			45	1	3	1	0	1	0
			46	0	0	0	0	0	0
			47	0	1	0	2	0	0
			48	4	4	4	0	2	0
			49	3	3	2	0	1	0
			50	2	3	4	1	4	1
			51	1	1	0	0	0	2
			52	4	4	4	1	4	1
			53	1	4	4	1	3	2
			Total	41	49	42	41		
			Specific	2					
			Nonspecific	49					
			None	2					
InI-3 (pL1: napin-GUS)	pCAMBIA 1391	438	1	0	0	0	2	0	1
			2	0	0	0	1	0	2
			3	0	0	0	0	0	1
			4	0	0	0	2	0	2
			5	0	0	0	1	0	1
			6	0	0	0	1	0	1
			7	0	0	0	1	0	1
			8	0	0	0	1	0	1
			9	0	0	0	0	0	0
			10	0	0	0	0	0	1
			11	0	0	0	1	0	1
			12	0	0	0	1	0	2
			13	0	0	0	0	0	1
			14	0	0	0	2	0	1
			15	0	0	0	0	0	0
			16	0	0	1	2	1	2
			17	0	0	0	2	0	2
			18	0	0	0	2	0	1
			19	0	0	0	1	0	2
			20	0	0	1	1	0	2
			Total	0	0	2	15	1	18
			Specific:	16					
			Nonspecific:	2					
			None:	2					
InII-12pL1		427	1	0	0	0	0	0	1
			2	0	0	0	2	0	1
			3	0	0	0	1	0	2
			4	0	0	0	1	0	2
			5	0	2	0	1	0	2
			6	0	0	0	0	0	1
			7	0	0	0	0	0	1
			8	0	0	0	0	0	1
			9	0	0	0	1	0	1
			10	0	0	0	2	0	1
			11	0	0	0	1	0	1
			12	0	0	0	1	0	2
			13	0	0	0	0	0	1

		14	0	0	0	1	0	1
		15	0	0	0	0	0	1
		16	0	0	0	2	0	1
		17	0	0	0	2	0	0
		18	0	0	0	0	0	0
		19	0	0	0	3	0	3
		20	0	0	0	2	0	3
		21	0	0	0	0	0	0
		22	0	0	0	0	0	0
		23	0	0	0	0	0	0
		24	0	0	0	0	0	2
		25	0	0	0	0	0	2
		26	0	0	0	1	0	0
		27	0	0	0	0	0	2
		Total	0	1	0	14	0	21
		Specific	22					
		Nonspecific	1					
		None	4					
InIII-50 (pL1: pCAMBIA 1391 napin-GUS)	443	1	0	0	0	2	0	2
		2	0	0	0	1	0	2
		3	0	0	0	2	0	1
		4	0	0	0	2	0	1
		5	0	0	0	1	0	1
		6	0	0	0	1	0	1
		7	0	0	0	1	0	1
		8	0	0	0	1	0	1
		9	0	0	0	2	0	2
		10	0	0	0	1	0	2
		11	0	0	0	1	0	1
		12	0	0	0	2	0	2
		13	0	0	0	2	0	1
		Total	0	0	0	13	0	13
		Specific:	13					
		Nonspecific:	0					
		None:	0					
InIII-78_5' (pL1: pCAMBIA 1391 napin-GUS)	153	1	0	0	0	2	0	1
		2	0	0	0	2	0	2
		3	0	0	0	2	0	1
		4	0	0	0	0	0	0
		5	0	0	0	2	0	2
		6	0	0	0	1	0	2
		7	0	0	0	1	0	2
		8	0	0	0	0	0	2
		9	0	0	0	0	0	1
		10	0	0	0	0	0	1
		11	0	0	0	2	0	2
		12	0	0	0	2	0	2
		13	0	0	0	1	0	2

	14	0	0	0	3	0	2
	15	0	0	0	2	0	2
	16	0	0	2	2	2	2
Total		0	0	1	12	1	15

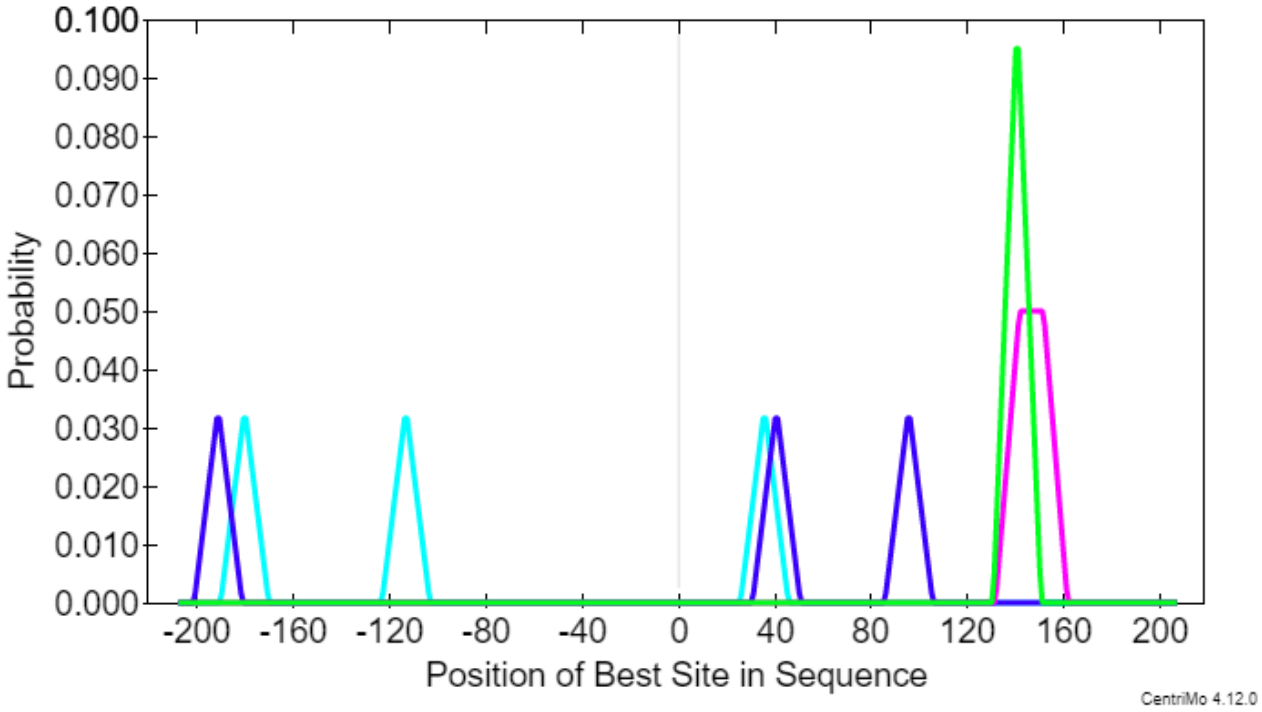
Specific:	14
Nonspecific:	1
None:	1



Legend:

Consensus	Max Probability Location
TCTCT	9.5
YRTTGCATACCY	185
CGATA	-132.5
GAGAA	-187.5
TATCG	-132.5
GAGAG	-201.5

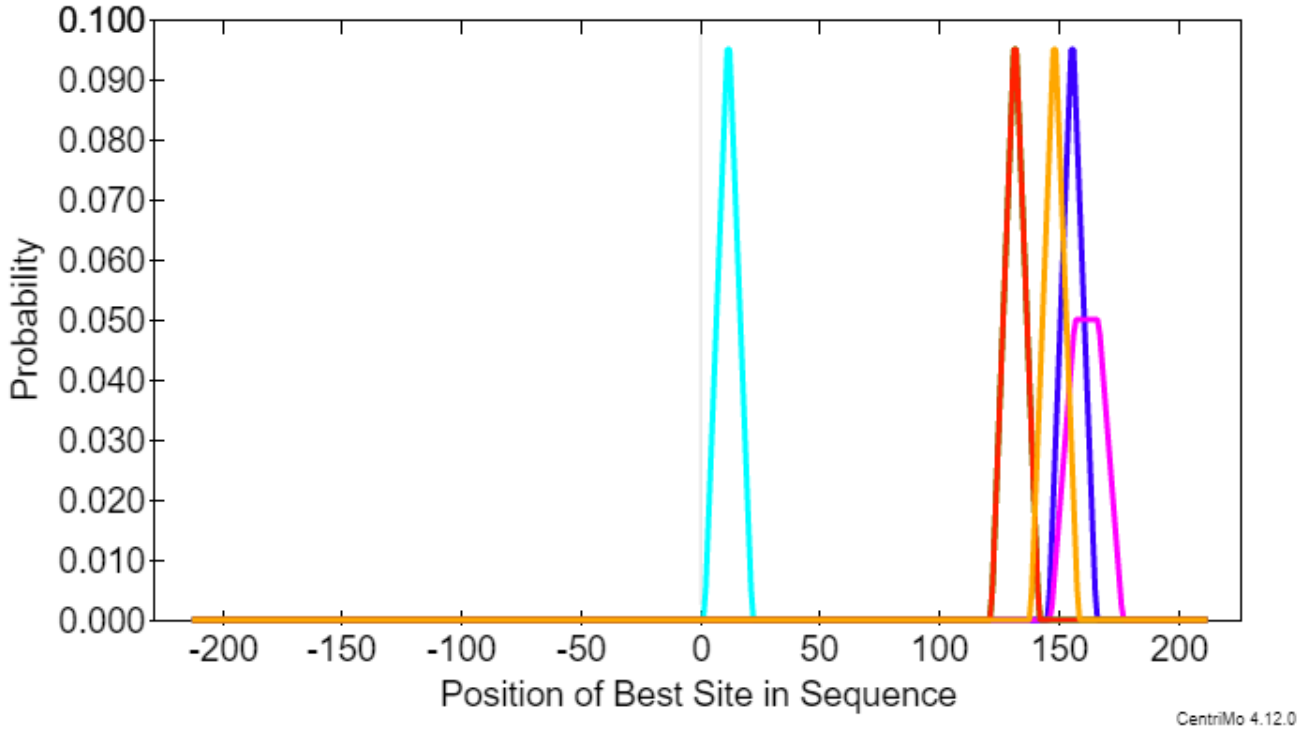
Figure A2.11: Motif probability graph of potential insulator protein binding sites on InI-3, produced by MemeSuite's CentriMo bioinformatics program.



Legend

	Consensus	Max Probability Location
	GAGAG	-180
	GAGAA	-191
	CTCTT	142
	TCTCT	141

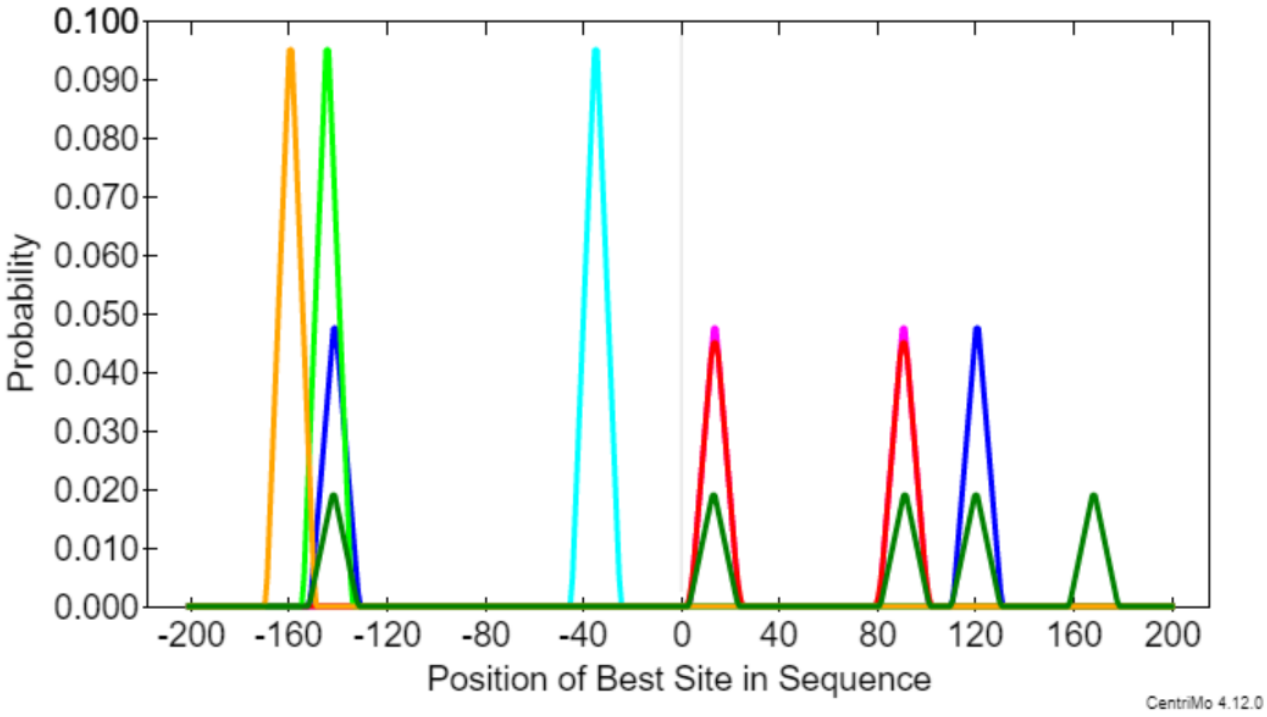
Figure A2.12: Motif probability graph of potential insulator protein binding sites on InII-12 produced by MemeSuite's CentriMo bioinformatics program.



Legend

	Consensus	Max Probability Location
	TGTTGCATACT	12
	TCTCT	156
	GAGAG	157
	CGATA	132
	TATCG	132
	AAGAGA	148.5

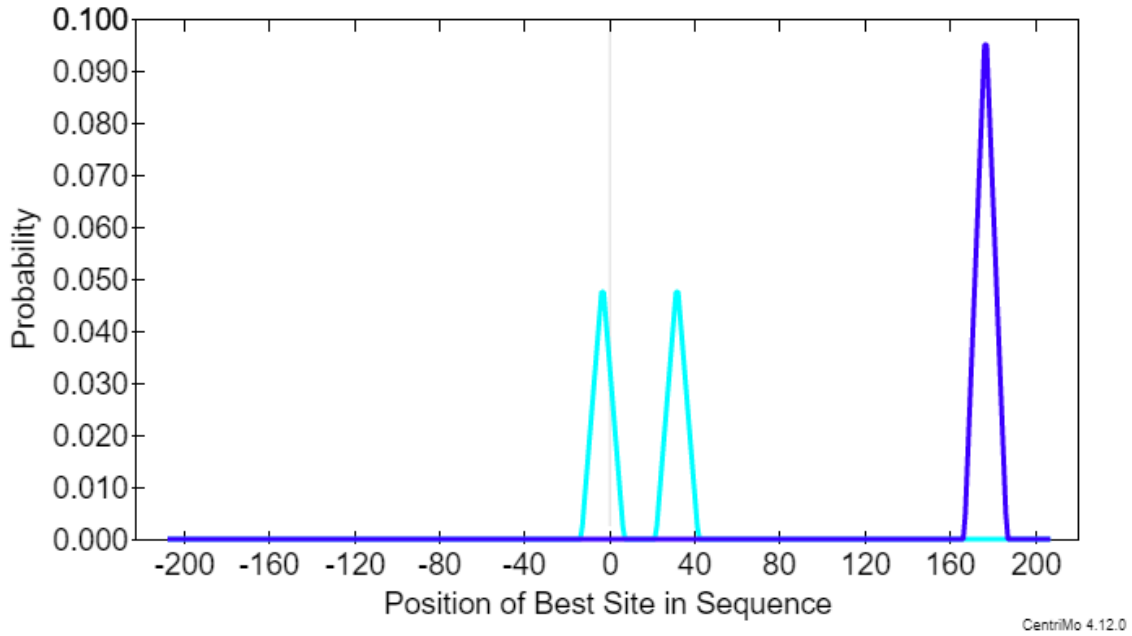
Figure A2.13: Motif probability graph of potential insulator protein binding sites on InIII-50 produced by MemeSuite's CentriMo bioinformatics program.



Legend

	Consensus	Max Probability Location
	TATCGATA	-34.5
	CTCTT	-141
	TCTCT	14
	GAGAG	13
	TATTGCCTACT	-144
	TGTTGCATACT	-159
	AAGAGA	-141.5

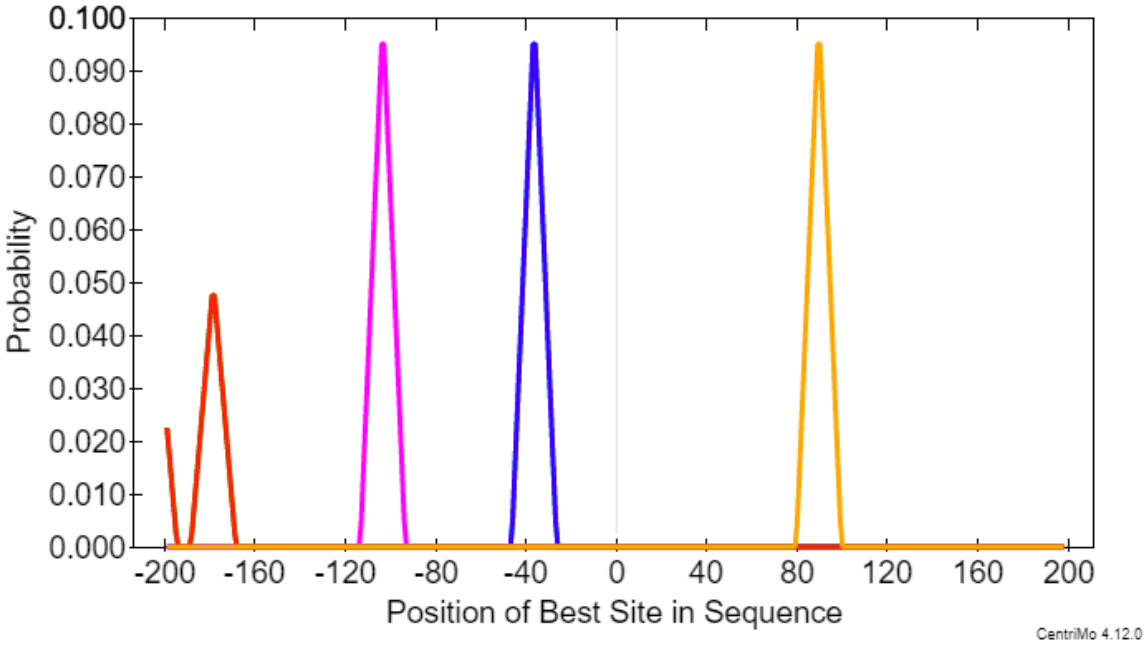
Figure A2.14: Motif probability graph of potential insulator protein binding sites on InIII-78 produced by MemeSuite's CentriMo bioinformatics program.



Enriched motifs (*E*-value ≤ 10 using the binomial test)

	Consensus	*E*-value	Max Probability	Max Probability Location
	TCTCT	7.0	0.5000	-3
	TTGCATACC	7.0	1.0000	177

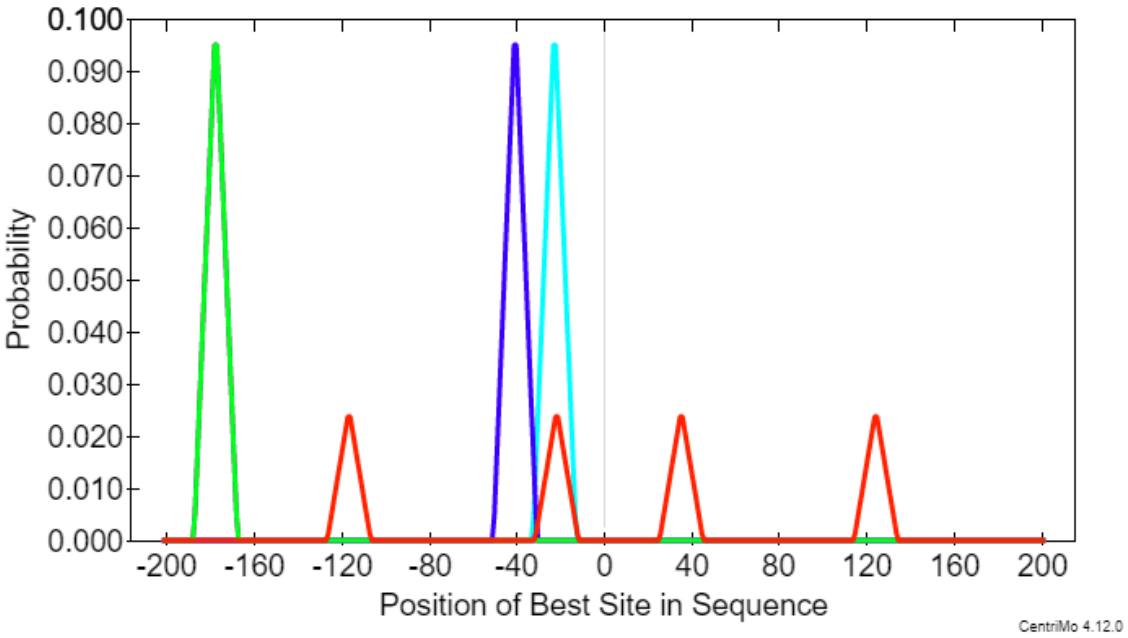
Figure A2.15: Motif probability graph of potential insulator protein binding sites on InIII-17 produced by MemeSuite's CentriMo bioinformatics program.



Enriched motifs (*E*-value ≤ 10 using the binomial test)

	Consensus	Max Probability	Max Probability Location
	CGATA	1.0000	-36
	TATCG	1.0000	-36
	TTGCATACC	1.0000	-103
	AAGAG	0.5000	-204
	CTCTT	0.5000	-204
	GAGAA	1.0000	90

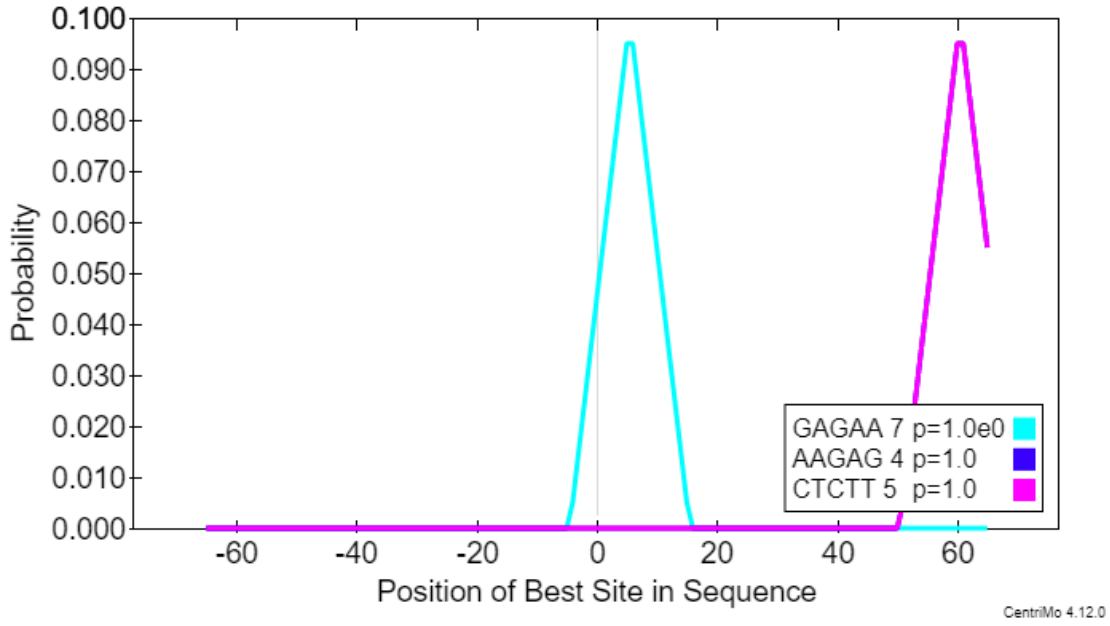
Figure A2.16: Motif probability graph of potential insulator protein binding sites on InIII-22 produced by MemeSuite's CentriMo bioinformatics program.



Enriched motifs (*E*-value ≤ 10 using the binomial test)

	Alt ID	Consensus	Max Probability	Max Probability Location
	TCTCT	TCTCT	1.0000	-22.5
	TTGCATACC	TTGCATACC	1.0000	-40.5
	AAGAG	AAGAG	1.0000	-177.5
	CTCTT	CTCTT	1.0000	-177.5
	GAGAA	GAGAA	0.2500	-116.5

Figure A2.17: Motif probability graph of potential insulator protein binding sites on InIII-55 produced by MemeSuite's CentriMo bioinformatics program.



Enriched motifs (*E*-value ≤ 10 using the binomial test)

	Consensus	Max Probability	Max Probability Location
	GAGAA	1.0000	5.5
	AAGAG	1.0000	60.5
	CTCTT	1.0000	60.5

Figure A2.18: Motif probability graph of potential insulator protein binding sites on InIII-57 produced by MemeSuite's CentriMo bioinformatics program.

Table A2.3: Potential plant *cis*-acting regulatory element binding motifs in InI-3 from PlantCARE online database. Search done on Oct 28, 2017.

Motif Name	Organism	Position	Strand	sequence	function
ABRE	*Hordeum vulgare*	38	+	GCAACGTGTC	cis-acting element involved in the abscisic acid responsiveness
ABRE	*Arabidopsis thaliana*	40	+	TACGTG	cis-acting element involved in the abscisic acid responsiveness
ACE	*Petroselinum crispum*	416	-	GCGACGTACC	cis-acting element involved in light responsiveness
Box-W1	*Petroselinum crispum*	310	-	TTGACC	fungal elicitor responsive element
CAAT-box	*Hordeum vulgare*	77	-	CAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Hordeum vulgare*	313	+	CAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Brassica rapa*	115	-	CAAAT	common cis-acting element in promoter and enhancer regions
G-Box	*Antirrhinum majus*	40	-	CACGTA	cis-acting regulatory element involved in light responsiveness
G-box	*Larix laricina*	38	-	GACACGTAGT	cis-acting regulatory element involved in light responsiveness
G-box	*Daucus carota*	40	+	TACGTG	cis-acting regulatory element involved in light responsiveness
GCN4_motif	*Oryza sativa*	316	-	CAAGCCA	cis-regulatory element involved in endosperm expression
HSE	*Lycopersicon esculentum*	199	+	CNNGAANNTTCNNG	cis-acting element involved in heat stress responsiveness
MBS	*Arabidopsis thaliana*	131, 181	+	TAACTG	MYB binding site involved in drought-inducibility
Sp1	*Oryza sativa*	24, 100	-	GGGCGG	light responsive element
Sp1	*Zea mays*	54, 53, 55, 333	+	CC(G/A)CCC	light responsive element
Sp1	*Zea mays*	249, 220	-	CC(G/A)CCC	light responsive element
TATA-box	*Lycopersicon esculentum*	48	+	TTTTA	core promoter element around -30 of transcription start
TATA-box	*Glycine max*	165	+	TAATA	core promoter element around -30 of transcription start
TCCC-motif	*Spinacia oleracea*	223	-	TCTCCCT	part of a light responsive element
W box	*Arabidopsis thaliana*	310	-	TTGACC	

Table A2.4: Potential plant cis-acting regulatory DNA binding motifs in InI-3 from New PLACE online database. Search done on Nov 13, 2017.

Motif Name	Location	Signal sequence	Site#
ACGTABREMOTIFAOSOSEM	40 (+)	TACGTGTC	S000281
ABREMOTIFAOSOSEM	40 (+)	TACGTGTC	S000299
ACGTATERD1	41 (-)	ACGT	S000415
ACGTABREMOTIFA2OSEM	41 (+)	ACGTGKC	S000394
ABRELATERD1	41 (+)	ACGTG	S000414
ACGTATERD1	41 (+)	ACGT	S000415
GADOWNAT	41 (+)	ACGTGTC	S000438
DOFCOREZM	47 (-)	AAAG	S000265
GT1CONSENSUS	49 (-)	GRWAAW	S000198
SORLIP1AT	73 (-)	GCCAC	S000482
CAATBOX1	78 (-)	CAAT	S000028
GATABOX	86 (-)	GATA	S000039
POLLEN1LELAT52	91 (-)	AGAAA	S000245
NTBBF1ARROLB	95 (-)	ACTTTA	S000273
TAAAGSTKST1	95 (+)	TAAAG	S000387
DOFCOREZM	96 (+)	AAAG	S000265
EBOXBNNAPA	115 (-)	CANNTG	S000144
MYCCONSENSUSAT	115 (-)	CANNTG	S000407
EBOXBNNAPA	115 (+)	CANNTG	S000144
MYCCONSENSUSAT	115 (+)	CANNTG	S000407
DOFCOREZM	130 (-)	AAAG	S000265
TAAAGSTKST1	130 (-)	TAAAG	S000387
MYBCORE	133 (-)	CNGTTR	S000176
MYB2AT	133 (+)	TAACTG	S000177
MYB2CONSENSUSAT	133 (+)	YAACKG	S000409
MYBCORE	136 (+)	CNGTTR	S000176
RAV1AAT	137 (-)	CAACA	S000314
SEBFCONSSTPR10A	143 (+)	YTGTCWC	S000391
ARFAT	144 (+)	TGTCTC	S000270
SURECOREATSULTR11	145 (-)	GAGAC	S000499
ARFAT	156 (-)	TGTCTC	S000270
SEBFCONSSTPR10A	156 (-)	YTGTCWC	S000391
SURECOREATSULTR11	156 (+)	GAGAC	S000499
CACTFTPPCA1	165 (-)	YACT	S000449
BOXIINTPATPB	169 (+)	ATAGAA	S000296
POLLEN1LELAT52	171 (+)	AGAAA	S000245
DOFCOREZM	173 (+)	AAAG	S000265
MYBCORE	184 (-)	CNGTTR	S000176
MYB2AT	184 (+)	TAACTG	S000177
MYB2CONSENSUSAT	184 (+)	YAACKG	S000409
CACTFTPPCA1	194 (-)	YACT	S000449
GTGANTG10	211 (-)	GTGA	S000378
RHERPATEXPA7	211 (+)	KCACGW	S000512
CGCGBOXAT	219 (-)	VCGCGB	S000501
CGCGBOXAT	219 (+)	VCGCGB	S000501
-10PEHVPSBD	232 (-)	TATTCT	S000392
P1BS	233 (-)	GNATATNC	S000459
P1BS	233 (+)	GNATATNC	S000459
ROOTMOTIFTAPOX1	234 (-)	ATATT	S000098
RHERPATEXPA7	239 (+)	KCACGW	S000512
QELEMENTZMZM13	247 (+)	AGGTCA	S000254
WBOXNTERF3	248 (-)	TGACY	S000457
WBOXNTCHN48	248 (-)	CTGACY	S000508
WRKY71OS	249 (-)	TGAC	S000447
GT1CONSENSUS	267 (+)	GRWAAW	S000198
GT1CONSENSUS	268 (+)	GRWAAW	S000198

GT1GMSCAM4 268 (+) GAAAAA S000453
DOFCOREZM 271 (+) AAAG S000265
LTRECOREATCOR15 274 (-) CCGAC S000153
PRECONSCRHSP70A 308 (-) SCGAYNRNNNNNNNNNNNNNNNHD S000506
QELEMENTZMZM13 314 (+) AGGTCA S000254
ELRECOREPCRP1 315 (-) TTGACC S000142
WBOXNTERF3 315 (-) TGACY S000457
WBOXATNPR1 316 (-) TTGAC S000390
WRKY71OS 316 (-) TGAC S000447
CAATBOX1 318 (+) CAAT S000028
CACTFTPPCA1 331 (+) YACT S000449
SEF3MOTIFGM 335 (+) AACCCA S000115
GTGANTG10 348 (-) GTGA S000378
GT1CONSENSUS 353 (-) GRWAAW S000198
GT1CONSENSUS 354 (-) GRWAAW S000198
E2FCONSENSUS 355 (+) WTTSSCSS S000476
CGCGBOXAT 371 (-) VCGCGB S000501
ABRERATCAL 371 (-) MACGYGB S000507
CGCGBOXAT 371 (+) VCGCGB S000501
IBOXCORE 376 (-) GATAA S000199
GATABOX 377 (-) GATA S000039
WBOXNTCHN48 380 (+) CTGACY S000508
WBOXHVISO1 381 (+) TGACT S000442
WRKY71OS 381 (+) TGAC S000447
WBOXNTERF3 381 (+) TGACY S000457
EECCRCAH1 382 (+) GANTTNC S000494
NODCON1GM 394 (-) AAAGAT S000461
OSE1ROOTNODULE 394 (-) AAAGAT S000467
DOFCOREZM 396 (-) AAAG S000265
TAAAGSTKST1 396 (-) TAAAG S000387
GT1CONSENSUS 397 (-) GRWAAW S000198
CGCGBOXAT 402 (-) VCGCGB S000501
CGCGBOXAT 402 (+) VCGCGB S000501
CURECORECR 423 (-) GTAC S000493
CURECORECR 423 (+) GTAC S000493
SEBFCONSSTPR10A 433 (+) YTGTCWC S000391
ARFAT 434 (+) TGTCTC S000270
SURECOREATSULTR11 435 (-) GAGAC S000499

Table A2.5: Potential eukaryotic transcription factor binding motifs in InI-3 from JASPAR Core online database. Search done on Oct 28, 2017.

Motif Name	Organism	Start	End	Strand	Predicted sequence	Function	JASPAR Matrix ID
SPI1	Homo sapiens	377	383	-	AGGAAGT	Tryptophan cluster factor	MA0080.2
Ubx	Drosophila melanogaster	165	168	+	TAAT	Homeo domain factors – HOX-related factors	MA0094.1
Deaf1	Drosophila melanogaster	209	214	-	TTCGTG	SAND domain factor	MA0185.1
Deaf1	Drosophila melanogaster	256	261	-	TTCGGG	SAND domain factor	MA0185.1
MSN2	Saccharomyces cerevisiae	57	61	-	AGGGG	C2H2 zinc finger factors	MA0341.1
MSN2	Saccharomyces cerevisiae	66	70	-	AGGGG	C2H2 zinc finger factors	MA0341.1
MSN4	Saccharomyces cerevisiae	57	61	-	AGGGG	C2H2 zinc finger factors	MA0342.1
MSN4	Saccharomyces cerevisiae	66	70	-	AGGGG	C2H2 zinc finger factors	MA0342.1
DOF2.4	Arabidopsis thaliana	265	271	+	AAAAAGT	C2H2 zinc finger factors	MA0982.1
DOF5.3	Arabidopsis thaliana	264	270	+	GAAAAAG	C2H2 zinc finger factors	MA1071.1

Table A2.6: Novel motif sequences in InI-3 found using WEEDER2.0 motif discovery tool. Search done on Nov 12, 13 2017.

Motif Sequence (WEEDER output)	Output: Arabidopsis - Tomtom (Motif Comparison Tool of known motifs)
AAAAAGTC	WRKY_tnt.WRKY50_col_a_m1 (WRKY50), WRKY_tnt.WRKY8_col_m1 (WRKY8), WRKY_tnt.WRKY8_colamp_a_m1 (WRKY8), WRKY_tnt.WRKY45_col_a_m1 (WRKY45), WRKY_tnt.WRKY17_colamp_a_m1 (WRKY17), WRKY_tnt.WRKY17_col_a_m1 (WRKY17), WRKY_tnt.WRKY43_col_a_m1 (WRKY43), WRKY_tnt.WRKY7_col_m1 (WRKY7), WRKY_tnt.WRKY43_colamp_a_m1 (WRKY43), WRKY_tnt.WRKY24_col_a_m1 (WRKY24), WRKY_tnt.WRKY40_colamp_a_m1 (WRKY40), WRKY_tnt.WRKY11_col_a_m1 (WRKY11), WRKY_tnt.WRKY18_colamp_a_m1 (WRKY18), WRKY_tnt.WRKY7_colamp_a_m1 (WRKY7), WRKY_tnt.WRKY40_col_m1 (WRKY40), WRKY_tnt.WRKY59_col_a_m1 (WRKY59), WRKY_tnt.WRKY25_colamp_a_m1 (WRKY25), WRKY_tnt.WRKY6_colamp_a_m1 (WRKY6), WRKY_tnt.WRKY47_colamp_a_m1 (WRKY47), WRKY_tnt.WRKY31_col_a_m1 (WRKY31), WRKY_tnt.WRKY45_colamp_a_m1 (WRKY45), WRKY_tnt.WRKY71_col_a_m1 (WRKY71), WRKY_tnt.WRKY21_col_m1 (WRKY21), WRKY_tnt.WRKY28_colamp_a_m1 (WRKY28), WRKY_tnt.WRKY29_col_a_m1 (WRKY29), WRKY_tnt.WRKY50_colamp_a_m1 (WRKY50), WRKY_tnt.WRKY15_colamp_a_m1 (WRKY15), WRKY_tnt.WRKY15_col_b_m1 (WRKY15), WRKY_tnt.WRKY3_col_a_m1 (WRKY3), WRKY_tnt.WRKY65_col_a_m1 (WRKY65), WRKY_tnt.WRKY65_colamp_a_m1 (WRKY65), WRKY_tnt.WRKY22_col_m1 (WRKY22), WRKY_tnt.WRKY22_colamp_a_m1 (WRKY22), WRKY_tnt.WRKY27_col_a_m1 (WRKY27), WRKY_tnt.WRKY75_colamp_a_m1 (WRKY75), WRKY_tnt.WRKY26_colamp_a_m1 (WRKY26), WRKY_tnt.WRKY6_col_a_m1 (WRKY6), WRKY_tnt.WRKY21_colamp_a_m1 (WRKY21), WRKY_tnt.WRKY70_col_m1 (WRKY70), WRKY_tnt.WRKY27_colamp_a_m1 (WRKY27), WRKY_tnt.WRKY75_col_a_m1 (WRKY75), WRKY_tnt.WRKY33_col_a_m1 (WRKY33), WRKY_tnt.WRKY14_col_a_m1 (WRKY14), WRKY_tnt.WRKY14_colamp_a_m1 (WRKY14), WRKY_tnt.WRKY24_colamp_a_m1 (WRKY24), WRKY_tnt.WRKY28_col_a_m1 (WRKY28), zfGRF_tnt.AT3G42860_col_a_m1 (AT3G42860), WRKY_tnt.WRKY26_col_a_m1 (WRKY26), WRKY_tnt.WRKY29_colamp_a_m1 (WRKY29)

AAAAAGTCGG	LOBAS2_tnt.LBD2_colamp_a_m1 (LBD2), MYBrelated_tnt.AT3G10580_colamp_a_m1 (AT3G10580), C3H_tnt.CDM1_colamp_a_m1 (CDM1), LOBAS2_tnt.LBD2_col_a_m1 (LBD2), WRKY_tnt.WRKY45_col_a_m1 (WRKY45), WRKY_tnt.WRKY50_col_a_m1 (WRKY50), WRKY_tnt.WRKY17_colamp_a_m1 (WRKY17), MYB_tnt.MS188_colamp_a_m1 (MS188), WRKY_tnt.WRKY17_col_a_m1 (WRKY17), WRKY_tnt.WRKY43_col_a_m1 (WRKY43), WRKY_tnt.WRKY40_colamp_a_m1 (WRKY40), WRKY_tnt.WRKY7_col_m1 (WRKY7), WRKY_tnt.WRKY18_colamp_a_m1 (WRKY18), WRKY_tnt.WRKY8_col_m1 (WRKY8), WRKY_tnt.WRKY8_colamp_a_m1 (WRKY8)
AAAAGT	WRKY_tnt.WRKY45_colamp_a_m1 (WRKY45), WRKY_tnt.WRKY71_col_a_m1 (WRKY71), WRKY_tnt.WRKY15_colamp_a_m1 (WRKY15), WRKY_tnt.WRKY45_col_a_m1 (WRKY45), WRKY_tnt.WRKY17_colamp_a_m1 (WRKY17), WRKY_tnt.WRKY22_col_m1 (WRKY22), WRKY_tnt.WRKY22_colamp_a_m1 (WRKY22), WRKY_tnt.WRKY27_col_a_m1 (WRKY27), WRKY_tnt.WRKY50_col_a_m1 (WRKY50), WRKY_tnt.WRKY75_colamp_a_m1 (WRKY75), WRKY_tnt.WRKY11_col_a_m1 (WRKY11), WRKY_tnt.WRKY7_col_m1 (WRKY7), WRKY_tnt.WRKY8_col_m1 (WRKY8), WRKY_tnt.WRKY8_colamp_a_m1 (WRKY8), C2C2dof_tnt.CDF3_colamp_a_m1 (CDF3), WRKY_tnt.WRKY21_col_m1 (WRKY21), WRKY_tnt.WRKY28_colamp_a_m1 (WRKY28), WRKY_tnt.WRKY65_colamp_a_m1 (WRKY65), WRKY_tnt.WRKY21_colamp_a_m1 (WRKY21), WRKY_tnt.WRKY70_col_m1 (WRKY70), WRKY_tnt.WRKY17_col_a_m1 (WRKY17), WRKY_tnt.WRKY43_col_a_m1 (WRKY43), WRKY_tnt.WRKY27_colamp_a_m1 (WRKY27), WRKY_tnt.WRKY75_col_a_m1 (WRKY75), C3H_tnt.AT5G63260_colamp_a_m1 (AT5G63260), WRKY_tnt.WRKY29_col_a_m1 (WRKY29), WRKY_tnt.WRKY50_colamp_a_m1 (WRKY50), Orphan_tnt.BBX31_col_a_m1 (BBX31), C2C2dof_tnt.COG1_colamp_a_m1 (COG1), WRKY_tnt.WRKY14_col_a_m1 (WRKY14), WRKY_tnt.WRKY14_colamp_a_m1 (WRKY14), WRKY_tnt.WRKY40_colamp_a_m1 (WRKY40), WRKY_tnt.WRKY24_colamp_a_m1 (WRKY24), WRKY_tnt.WRKY28_col_a_m1 (WRKY28), WRKY_tnt.WRKY15_col_b_m1 (WRKY15), WRKY_tnt.WRKY3_col_a_m1 (WRKY3), WRKY_tnt.WRKY65_col_a_m1 (WRKY65), C2C2dof_tnt.AT1G69570_col_a_m1 (AT1G69570), C2C2dof_tnt.OBP3_colamp_a_m1 (OBP3), WRKY_tnt.WRKY43_colamp_a_m1 (WRKY43), C2C2dof_tnt.dof24_col_a_m1 (dof24), C2C2dof_tnt.At4g38000_colamp_a_m1 (At4g38000), C2C2dof_tnt.AT1G69570_colamp_a_m1 (AT1G69570), C2C2dof_tnt.Adof1_col_a_m1 (Adof1), WRKY_tnt.WRKY26_col_a_m1 (WRKY26), WRKY_tnt.WRKY24_col_a_m1 (WRKY24), C2C2dof_tnt.At1g64620_colamp_a_m1 (At1g64620), C3H_tnt.AT5G63260_col_a_m1 (AT5G63260), C2C2dof_tnt.dof24_colamp_a_m1 (dof24), C2C2dof_tnt.AT2G28810_colamp_a_m1 (AT2G28810), C2C2dof_tnt.CDF3_col_a_m1 (CDF3), C2C2do

	f_tnt.OBP4_colamp_a_m1 (OBP4), WRKY_tnt.WRKY29_colamp_a_m1 (WRKY29), WRKY_tnt.WRKY18_colamp_a_m1 (WRKY18), C2C2 dof_tnt.DAG2_col_a_m1 (DAG2), WRKY_tnt.WRKY26_colamp_a_m1 (WRKY26), C2C2dof_tnt.AT5G66940_col_a_m1 (AT5G66940), C2C 2dof_tnt.Adof1_colamp_a_m1 (Adof1), C2C2dof_tnt.At5g62940_col_a_m1 (At5g62940), C2C2dof_tnt.COG1_col_a_m1 (COG1), WRKY_tnt. WRKY7_colamp_a_m1 (WRKY7), C2C2dof_tnt.AT2G28810_col_a_m1 (AT2G28810), C2C2dof_tnt.AT5G66940_colamp_a_m1 (AT5G66940), WRKY_tnt.WRKY40_col_m1 (WRKY40), WRKY_tnt.WRKY33_col_a_m1 (WRKY33), C2C2dof_tnt.AT5G02460_colamp_a_m1 (AT5G02460), C2C2dof_tnt.At3g45610_col_a_m1 (At3g45610), zfGRF_tnt.AT3G42860_col_a_m1 (AT3G42860), NAC_tnt.NTM1_colamp_a_m1 (NTM1), C2C2dof_tnt.AT3G52440_colamp_a_m1 (AT3G52440), C2C2dof_tnt.At1g64620_100ng20cy_b_m1 (At1g64620), WRKY_tnt.WRKY59_col_a_ m1 (WRKY59), C2C2dof_tnt.OBP3_col_a_m1 (OBP3), C2C2dof_tnt.DAG2_colamp_a_m1 (DAG2)
AGCGCG	CAMTA_tnt.CAMTA1_col_a_m1 (CAMTA1), FAR1_tnt.FAR1_col_a_m1 (FAR1), FAR1_tnt.FAR1_colamp_a_m1 (FAR1)

Table A2.7: Potential plant *cis*-acting regulatory element binding motifs in InII-12 from PlantCARE online database. Search done on Oct 28 2017.

Motif Name	Organism	Position	Strand	sequence	function
A-box	*Petroselinum crispum*	123	-	CCGTCC	cis-acting regulatory element
ABRE	*Arabidopsis thaliana*	104, 385	+	CACGTG, TACGTG	cis-acting element involved in the abscisic acid responsiveness
ACE	*Petroselinum hortense*	102	-	ACGTGGA	cis-acting element involved in light responsiveness
AE-box	*Arabidopsis thaliana*	304	-	AGAAACAA	part of a module for light response
CAAT-box	*Hordeum vulgare*	54, 322	-	CAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Brassica rapa*	188, 164	+	CAAAT	common cis-acting element in promoter and enhancer regions
CCGTCC-box	*Arabidopsis thaliana*	123	-	CCGTCC	cis-acting regulatory element related to meristem specific activation
G-Box	*Pisum sativum*	104	+	CACGTG	cis-acting regulatory element involved in light responsiveness
G-Box	*Triticum aestivum*	389	-	TCCACATGGCA	cis-acting regulatory element involved in light responsiveness
G-Box	*Antirrhinum majus*	385	-	CACGTA	cis-acting regulatory element involved in light responsiveness
G-box	*Brassica napus*	103	-	CACGTGG	cis-acting regulatory element involved in light responsiveness
G-box	*Daucus carota*	385	+	TACGTG	cis-acting regulatory element involved in light responsiveness
G-box	*Nicotiana plumbaginifolia*	335	-	CAGACGTGGCA	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	104	+	CACGTG	cis-acting regulatory element involved in light responsiveness
G-box	*Brassica oleracea*	384	-	TAACACGTAG	cis-acting regulatory element involved in light responsiveness
GT1-motif	*Solanum tuberosum*	394	-	AATCCACA	light responsive element
Sp1	*Zea mays*	74, 91, 86	+	CC(G/A)CCC	light responsive element
Sp1	*Zea mays*	227	-	CC(G/A)CCC	light responsive element
TATA-box	*Lycopersicon esculentum*	111, 375	+	TTTTA	core promoter element around -30 of transcription start
TCA-element	*Brassica oleracea*	250	+	GAGAAGAATA	cis-acting element involved in salicylic acid responsiveness
circadian	*Lycopersicon esculentum*	325	-	CAANNNNATC	cis-acting regulatory element involved in circadian control

Table A2.8: Potential plant cis-acting regulatory DNA binding motifs in InII-12 from New PLACE online database. Search done on Nov 13, 2017.

Motif Name	Location	Signal sequence	Site#
CURECORECR	9 (-)	GTAC	S000493
CURECORECR	9 (+)	GTAC	S000493
DOFCOREZM	13 (-)	AAAG	S000265
POLLEN1LELAT52	14 (-)	AGAAA	S000245
EECCRCAH1	39 (+)	GANTTNC	S000494
ROOTMOTIFTAPOX1	46 (+)	ATATT	S000098
CAATBOX1	48 (-)	CAAT	S000028
ACGTABOX	53 (-)	TACGTA	S000130
ACGTABOX	53 (+)	TACGTA	S000130
ACGTATERD1	54 (-)	ACGT	S000415
ACGTATERD1	54 (+)	ACGT	S000415
ARR1AT	65 (-)	NGATT	S000454
UPRMOTIFIIAT	83 (+)	CCNNNNNNNNNNNNCCACG	S000426
CAREOSREP1	90 (+)	CAACTC	S000421
IRO2OS	97 (-)	CACGTGG	S000505
ABRERATCAL	97 (-)	MACGYGB	S000507
CACGTGMOTIF	98 (-)	CACGTG	S000042
EBOXBNNAPA	98 (-)	CANNTG	S000144
MYCCONSENSUSAT	98 (-)	CANNTG	S000407
ABRELATERD1	98 (-)	ACGTG	S000414
CACGTGMOTIF	98 (+)	CACGTG	S000042
EBOXBNNAPA	98 (+)	CANNTG	S000144
MYCCONSENSUSAT	98 (+)	CANNTG	S000407
ABRERATCAL	98 (+)	MACGYGB	S000507
ACGTATERD1	99 (-)	ACGT	S000415
RHERPATEXPA7	99 (-)	KCACGW	S000512
ABRELATERD1	99 (+)	ACGTG	S000414
ACGTATERD1	99 (+)	ACGT	S000415
DOFCOREZM	104 (-)	AAAG	S000265
CACTFTPPCA1	110 (-)	YACT	S000449
CMSRE1IBSPOA	116 (+)	TGGACGG	S000511
PALBOXAPC	117 (-)	CCGTCC	S000137
SEBFCONSSTPR10A	126 (+)	YTGTCWC	S000391
ARFAT	127 (+)	TGTCTC	S000270
SURECOREATSULTR11	128 (-)	GAGAC	S000499
ARFAT	139 (-)	TGTCTC	S000270
SEBFCONSSTPR10A	139 (-)	YTGTCWC	S000391
SURECOREATSULTR11	139 (+)	GAGAC	S000499
GTGANTG10	153 (+)	GTGA	S000378
WRKY71OS	154 (+)	TGAC	S000447
WBOXNTERF3	154 (+)	TGACY	S000457
EBOXBNNAPA	175 (-)	CANNTG	S000144
MYCCONSENSUSAT	175 (-)	CANNTG	S000407
EBOXBNNAPA	175 (+)	CANNTG	S000144
MYCCONSENSUSAT	175 (+)	CANNTG	S000407
EBOXBNNAPA	182 (-)	CANNTG	S000144
MYCCONSENSUSAT	182 (-)	CANNTG	S000407
EBOXBNNAPA	182 (+)	CANNTG	S000144
MYCCONSENSUSAT	182 (+)	CANNTG	S000407
SV40COREENHAN	190 (+)	GTGGWWHG	S000123

MYBCORE 209 (+) CNGTTR S000176
MYBPZM 210 (-) CCWACC S000179
-10PEHVPSBD 248 (-) TATTCT S000392
POLASIG3 250 (+) AATAAT S000088
ARR1AT 253 (-) NGATT S000454
CGCGBOXAT 263 (-) VCGCGB S000501
CGCGBOXAT 263 (+) VCGCGB S000501
CURECORECR 269 (-) GTAC S000493
CURECORECR 269 (+) GTAC S000493
POLLEN1LELAT52 301 (-) AGAAA S000245
GTGANTG10 310 (-) GTGA S000378
ROOTMOTIFTAPOX1 314 (+) ATATT S000098
CAATBOX1 316 (-) CAAT S000028
ARR1AT 318 (+) NGATT S000454
CIACADIANLELHC 319 (-) CAANNNNATC S000252
EECCRCAH1 319 (+) GANTTNC S000494
GT1CONSENSUS 320 (-) GRWAAW S000198
POLLEN1LELAT52 322 (-) AGAAA S000245
SORLIP1AT 330 (+) GCCAC S000482
GATABOX 345 (-) GATA S000039
NODCON2GM 348 (+) CTCTT S000462
OSE2ROOTNODULE 348 (+) CTCTT S000468
DOFCOREZM 350 (-) AAAG S000265
GTGANTG10 353 (-) GTGA S000378
CACTFTPPCA1 356 (+) YACT S000449
NODCON2GM 358 (+) CTCTT S000462
OSE2ROOTNODULE 358 (+) CTCTT S000468
PYRIMIDINEBOXOSRAMY1A 367 (+) CCTTTT S000259
DOFCOREZM 368 (-) AAAG S000265
ACGTATERD1 380 (-) ACGT S000415
RHERPATEXPA7 380 (-) KCACGW S000512
ABRELATERD1 380 (+) ACGTG S000414
ACGTATERD1 380 (+) ACGT S000415
ARR1AT 391 (+) NGATT S000454
EECCRCAH1 392 (+) GANTTNC S000494
GT1CONSENSUS 393 (-) GRWAAW S000198
SEBFCONSSTPR10A 409 (+) YTGTCWC S000391
ARFAT 410 (+) TGTCTC S000270
SURECOREATSULTR11 411 (-) GAGAC S000499

Table A2.9: Potential eukaryotic transcription factor binding motifs in InII-12 from JASPAR Core online database. Search done on Oct 28, 2017.

Name	Organism	Score	Start	End	Strand	Predicted sequence	Function	Matrix ID
Arnt	Mus musculus	10.3511	104	109	+	CACGTG	Basic helix-loop-helix factors (bHLH) - PAS domain factors	MA0004.1
Arnt	Mus musculus	10.3511	104	109	-	CACGTG	Basic helix-loop-helix factors (bHLH) - PAS domain factors	MA0004.1
Ubx	Drosophila melanogaster	6.91669	258	261	+	TAAT	Homeo domain factors	MA0094.1
Mycn	Mus musculus	10.5328	104	109	+	CACGTG	Basic helix-loop-helix factors (bHLH)	MA0104.1
Mycn	Mus musculus	10.5328	104	109	-	CACGTG	Basic helix-loop-helix factors (bHLH)	MA0104.1
ZNF354C		8.91578	395	400	-	ATCCAC		MA0130.1
FEV		12.0467	399	406	-	CAGGAAAT		MA0156.1
bap		11.9861	113	119	+	TTAAGTG		MA0211.1
ADR1		10.2598	207	213	-	ACCCCAC		MA0268.1
HAP2	Saccharomyces cerevisiae	8.68725	162	166	-	TTGGT	Heteromeric CCAAT-binding factors	MA0313.1
MSN2	Saccharomyces cerevisiae	8.79054	77	81	-	AGGGG	C6 zinc cluster factors	MA0341.1
MSN2	Saccharomyces cerevisiae	8.79054	244	248	+	AGGGG	C6 zinc cluster factors	MA0341.1
MSN2	Saccharomyces cerevisiae	8.79054	312	316	-	AGGGG	C6 zinc cluster factors	MA0341.1
MSN4	Saccharomyces cerevisiae	9.22701	77	81	-	AGGGG	C6 zinc cluster factors	MA0342.1
MSN4	Saccharomyces cerevisiae	9.22701	244	248	+	AGGGG	C6 zinc cluster factors	MA0342.1
MSN4	Saccharomyces cerevisiae	9.22701	312	316	-	AGGGG	C6 zinc cluster factors	MA0342.1
YAP5		9.61465	39	44	-	AAGCAT		MA0417.1
ISL2		12.0118	113	120	-	GCACTTAA		MA0914.1
OsI_08196		14.5411	190	197	-	GGGCCCAC		MA1050.1
TCP15		15.0766	190	197	-	GGGCCCAC		MA1062.1
TCP20		13.6507	188	197	-	GGGCCCACCA		MA1065.1
TCP23		15.171	190	197	-	GGGCCCAC		MA1066.1

Table A2.10: Novel motif sequences in InII-12 found using WEEDER2.0 motif discovery tool. Search done on Nov 12, 13 2017.

Motif Sequence (WEEDER output)	Output: Arabidopsis - Tomtom (Motif Comparison Tool of known motifs)
AAATAG	No motif matches
AAATAGAC	No motif matches
AAATAGACCG	No motif matches
ACAGCT	Tcf21_DBD, MSC_full, UP00258_1 (Tgif2_3451.1), TGIF1_DBD, UP00122_1 (Tgif1_2342.2), MA0521.1 (Tcf12), MYF6_full, TFAP4_full, TFAP4_DBD, TGIF2_DBD
ACGTGC	BZR_tnt.At1g78700_col_a_m1 (At1g78700), BZR_tnt.At4g18890_colamp_a_m1 (At4g18890), bHLH_tnt.bHLH69_col_a_m1 (bHLH69), BZR_tnt.At4g18890_col_a_m1 (At4g18890), BZR_tnt.At4g36780_col_a_m1 (At4g36780), BZR_tnt.At1g78700_colamp_a_m1 (At1g78700), NAC_tnt.ANAC047_colamp_a_d1 (ANAC047), bHLH_tnt.bHLH31_col_m1 (bHLH31), bHLH_tnt.bHLH77_col_a_m1 (bHLH77), bHLH_tnt.bHLH74_col_a_m1 (bHLH74)
ACTTGT	No motif matches
AGGAGA	No motif matches
AGGTGG	No motif matches

Table A2.11: Potential plant *cis*-acting regulatory element binding motifs in InIII-50 from PlantCARE online database. Search done on Oct 28 2017.

Motif Name	Organism	Position	Strand	sequence	function
A-box	*Petroselinum crispum*	134	+	CCGTCC	cis-acting regulatory element
ABRE	*Oryza sativa*	324	+	GCCGCGTGGC	cis-acting element involved in the abscisic acid responsiveness
ABRE	*Arabidopsis thaliana*	417	+	ACGTGGC	cis-acting element involved in the abscisic acid responsiveness
ABRE	*Arabidopsis thaliana*	416	-	CACGTG	cis-acting element involved in the abscisic acid responsiveness
CAAT-box	*Glycine max*	4	+	CAATT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Hordeum vulgare*	6	-	CAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Glycine max*	5	-	CAATT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Brassica rapa*	121	-	CAAAT	common cis-acting element in promoter and enhancer regions
CCGTCC-box	*Arabidopsis thaliana*	134	+	CCGTCC	cis-acting regulatory element related to meristem specific activation
CE3	*Oryza sativa*	43	-	GACGCGTGTC	cis-acting element involved in ABA and VP1 responsiveness
G-Box	*Pisum sativum*	416	-	CACGTG	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	324, 413	-	GCCACGTGGA	cis-acting regulatory element involved in light responsiveness
G-box	*Arabidopsis thaliana*	413	-	GCCACGTGGTA	cis-acting regulatory element involved in light responsiveness
GARE-motif	*Brassica oleracea*	385	-	AAACAGA	gibberellin-responsive element
I-box	*Flaveria trinervia*	366	-	GATATGG	part of a light responsive element
MBS	*Zea mays*	193	+	CGGTCA	MYB Binding Site
MNF1	*Zea mays*	125	+	GTGCCC(A/T)(A/T)	light responsive element
Sp1	*Zea mays*	130	+	CC(G/A)CCC	light responsive element
box II	*Petroselinum hortense*	324	+	TCCACGTGGC	part of a light responsive element
box S	*Arabidopsis thaliana*	186	-	AGCCACC	
circadian	*Lycopersicon esculentum*	363	+	CAANNNNATC	cis-acting regulatory element involved in circadian control

Table A2.12: Potential plant cis-acting regulatory DNA binding motifs in InIII-50 from New PLACE online database. Search done on Nov 13, 2017.

Motif Name	Location	Signal sequence	Site#
SORLIP2AT	8 (-)	GGGCC	S000483
CGCGBOXAT	23 (-)	VCGCGB	S000501
CGCGBOXAT	23 (+)	VCGCGB	S000501
VOZATVPP	25 (-)	GCGTNNNNNNNACGC	S000456
VOZATVPP	25 (+)	GCGTNNNNNNNACGC	S000456
SV40COREENHAN	30 (-)	GTGGWWHG	S000123
CURECORECR	31 (-)	GTAC	S000493
CURECORECR	31 (+)	GTAC	S000493
CGCGBOXAT	45 (-)	VCGCGB	S000501
CGCGBOXAT	45 (+)	VCGCGB	S000501
CURECORECR	49 (-)	GTAC	S000493
CURECORECR	49 (+)	GTAC	S000493
DPBFCOREDCDC3	51 (+)	ACACNNG	S000292
RAV1AAT	60 (-)	CAACA	S000314
CACTFTPPCA1	66 (+)	YACT	S000449
GRAZMRAB28	70 (-)	CATGCCGCC	S000220
GCCCORE	70 (-)	GCCGCC	S000430
RYREPEATLEGUMINBOX	75 (+)	CATGCAY	S000100
RYREPEATBNNAPA	75 (+)	CATGCA	S000264
CACTFTPPCA1	79 (+)	YACT	S000449
DOFCOREZM	85 (-)	AAAG	S000265
TAAAGSTKST1	85 (-)	TAAAG	S000387
POLASIG1	86 (-)	AATAAA	S000080
TATABOX5	87 (+)	TTATTT	S000203
RHERPATEXPA7	92 (-)	KCACGW	S000512
EBOXBNNAPA	97 (-)	CANNTG	S000144
MYCCONSENSUSAT	97 (-)	CANNTG	S000407
EBOXBNNAPA	97 (+)	CANNTG	S000144
MYCCONSENSUSAT	97 (+)	CANNTG	S000407
PALBOXAPC	111 (+)	CCGTCC	S000137
ARR1AT	123 (-)	NGATT	S000454
SEBFCONSSTPR10A	130 (+)	YTGTCWC	S000391
ARFAT	131 (+)	TGTCTC	S000270
SURECOREATSULTR11	132 (-)	GAGAC	S000499
ARFAT	143 (-)	TGTCTC	S000270
SEBFCONSSTPR10A	143 (-)	YTGTCWC	S000391
SURECOREATSULTR11	143 (+)	GAGAC	S000499
CACTFTPPCA1	149 (-)	YACT	S000449
MYBCORE	151 (-)	CNGTTR	S000176
MYB2CONSENSUSAT	151 (+)	YAACKG	S000409
MYBCOREATCYCB1	152 (+)	AACGG	S000502
SORLIP1AT	164 (-)	GCCAC	S000482
WBOXNTERF3	171 (-)	TGACY	S000457
WRKY71OS	172 (-)	TGAC	S000447
GTGANTG10	173 (-)	GTGA	S000378
CGACGOSAMY3	196 (+)	CGACG	S000205
CGCGBOXAT	198 (-)	VCGCGB	S000501
CGCGBOXAT	198 (+)	VCGCGB	S000501
POLLEN1LELAT52	204 (+)	AGAAA	S000245
-300CORE	211 (-)	TGTAAAG	S000001

DOFCOREZM 211 (-) AAAG S000265
TAAAGSTKST1 211 (-) TAAAG S000387
DPBFCOREDCDC3 215 (+) ACACNNG S000292
GT1CONSENSUS 220 (+) GRWAAW S000198
ROOTMOTIFTAPOX1 223 (-) ATATT S000098
PRECONSCRHSP70A 228 (+) SCGAYNRNNNNNNNNNNNNNNHD S000506
SORLIP1AT 234 (-) GCCAC S000482
CAREOSREP1 239 (-) CAACTC S000421
CACTFTPPCA1 252 (-) YACT S000449
CURECORECR 253 (-) GTAC S000493
CURECORECR 253 (+) GTAC S000493
DOFCOREZM 257 (+) AAAG S000265
SORLIP2AT 260 (-) GGGCC S000483
TBOXATGAPB 266 (-) ACTTTG S000383
DOFCOREZM 267 (+) AAAG S000265
LTRECOREATCOR15 270 (-) CCGAC S000153
DRECRTCOREAT 270 (-) RCCGAC S000418
CBFHV 270 (-) RYCGAC S000497
SURECOREATSULTR11 298 (-) GAGAC S000499
CGCGBOXAT 302 (-) VCGCGB S000501
ABRERATCAL 302 (-) MACGYGB S000507
CGCGBOXAT 302 (+) VCGCGB S000501
SORLIP1AT 306 (-) GCCAC S000482
CACTFTPPCA1 313 (+) YACT S000449
DOFCOREZM 315 (-) AAAG S000265
GT1CONSENSUS 316 (-) GRWAAW S000198
HEXAMERATH4 321 (+) CCGTCG S000146
CGACGOSAMY3 322 (-) CGACG S000205
CACTFTPPCA1 327 (+) YACT S000449
CGACGOSAMY3 332 (+) CGACG S000205
CIACADIANLELHC 340 (+) CAANNNNATC S000252
GATABOX 346 (-) GATA S000039
SORLIP2AT 350 (+) GGGCC S000483
GT1CONSENSUS 366 (-) GRWAAW S000198
IBOXCORE 367 (-) GATAA S000199
GATABOX 368 (-) GATA S000039
INRNTPSADB 373 (+) YTCANTYY S000395
GTGANTG10 374 (-) GTGA S000378
CACTFTPPCA1 375 (+) YACT S000449
BOXIINTPATPB 387 (-) ATAGAA S000296
GATABOX 390 (-) GATA S000039
GTGANTG10 392 (-) GTGA S000378
RHERPATEXPA7 392 (+) KCACGW S000512
CACGTGMOTIF 393 (-) CACGTG S000042
EBOXBNNAPA 393 (-) CANNTG S000144
MYCCONSENSUSAT 393 (-) CANNTG S000407
ABRELATERD1 393 (-) ACGTG S000414
CACGTGMOTIF 393 (+) CACGTG S000042
EMBP1TAEM 393 (+) CACGTGGC S000119
EBOXBNNAPA 393 (+) CANNTG S000144
ABREATCONSENSUS 393 (+) YACGTGGC S000406
MYCCONSENSUSAT 393 (+) CANNTG S000407
IRO2OS 393 (+) CACGTGG S000505
ABRERATCAL 393 (+) MACGYGB S000507
ACGTATERD1 394 (-) ACGT S000415
BOXIIPCCHS 394 (+) ACGTGGC S000229

ACGTABREMOTIFA2OSEM	394 (+)	ACGTGKC	S000394
ABRELATERD1	394 (+)	ACGTG	S000414
ACGTATERD1	394 (+)	ACGT	S000415
SORLIP1AT	396 (-)	GCCAC	S000482
CGCGBOXAT	405 (-)	VCGCGB	S000501
CGCGBOXAT	405 (+)	VCGCGB	S000501
SEBFCONSSTPR10A	425 (+)	YTGTCWC	S000391
ARFAT	426 (+)	TGTCTC	S000270
SURECOREATSULTR11	427 (-)	GAGAC	S000499

Table A2.13: Potential eukaryotic transcription factor binding motifs in InIII-50 from JASPAR Core online database. Search done on Oct 28, 2017.

Name	Organism	Start	End	Strand	Predicted sequence	Function	Matrix ID
Arnt	Mus musculus	399	404	+	CACGTG	Basic helix-loop-helix factors (bHLH) - PAS domain factors	MA0004.1
Arnt	Mus musculus	399	404	-	CACGTG	Basic helix-loop-helix factors (bHLH) - PAS domain factors	MA0004.1
Dof3	Zea mays	89	94	-	AAAGCG	C2H2 zinc finger factors - Dof-type	MA0021.1
Dof3	Zea mays	215	220	-	AAAGCG	C2H2 zinc finger factors - Dof-type	MA0021.1
SPI1	Homo sapiens	382	388	-	AGGAAGT	Tryptophan cluster factors - Ets-related factors	MA0080.2
SPIB	Homo sapiens	384	390	-	AGAGGAA	Tryptophan cluster factors - Ets-related factors	MA0081.1
Ubx	Drosophila melanogaster	128	131	+	TAAT	Homeo domain factors	MA0094.1
Mycn	Mus musculus	399	404	+	CACGTG	Basic helix-loop-helix factors (bHLH)	MA0104.1
Mycn	Mus musculus	399	404	-	CACGTG	Basic helix-loop-helix factors (bHLH)	MA0104.1
Mycn	Mus musculus	399	406	-	GCCACGTG	Basic helix-loop-helix factors (bHLH)	MA0104.3
Deaf1	Drosophila melanogaster	97	102	+	TTCGTG	SAND domain factors	MA0185.1
Optix	Drosophila melanogaster	396	400	-	TGATA	Homeo domain factors – HD-sine factors	MA0199.1
FZF1	Saccharomyces cerevisiae	395	400	+	CTATCA	C2H2 zinc finger factors – factors with multiple dispersed zinc fingers	MA0298.1
HAL9	Saccharomyces cerevisiae	163	167	-	CGGAA	C6 zinc cluster factors	MA0311.1
MSN2	Saccharomyces cerevisiae	111	115	-	AGGGG	C2H2 zinc finger factors	MA0341.1
MSN4	Saccharomyces cerevisiae	111	115	-	AGGGG	C2H2 zinc finger factors	MA0342.1
TYE7	Saccharomyces cerevisiae	398	404	-	CACGTGA	Basic helix-loop-helix factors (bHLH)	MA0409.1
PIF4	Arabidopsis thaliana	399	406	+	CACGTGGC	Basic helix-loop-helix factors (bHLH)	MA0561.1

Table A2.14: Novel motif sequences in InIII-50 found using WEEDER2.0 motif discovery tool. Search done on Nov 12, 13 2017.

Motif Sequence (WEEDER output)	Output: Arabidopsis - Tomtom (Motif Comparison Tool of known motifs)
AAACGC	No motif matches
AAACGCTT	No motif matches
AAACGCTTTA	No motif matches
ACGCAC	No motif matches
AGGAGG	C2H2_tnt.TF3A_col_a_m1 (TF3A)
AGGATG	No motif matches

Table A2.15: Potential plant *cis*-acting regulatory element binding motifs in InIII-78 from PlantCARE online database. Search done on Oct 28 2017.

Motif Name	Organism	Position	Strand	sequence	function
A-box	*Petroselinum crispum*	393	+	CCGTCC	cis-acting regulatory element
ABRE	*Arabidopsis thaliana*	370	-	TACGTG	cis-acting element involved in the abscisic acid responsiveness
CAAT-box	*Hordeum vulgare*	50	+	CAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Brassica rapa*	175	+	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT-box	*Brassica rapa*	81	-	CAAAT	common cis-acting element in promoter and enhancer regions
CCAAT-box	*Hordeum vulgare*	121	-	CAACGG	MYBHv1 binding site
CCGTCC-box	*Arabidopsis thaliana*	393	+	CCGTCC	cis-acting regulatory element related to meristem specific activation
CGTCA-motif	*Hordeum vulgare*	411	-	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
G-Box	*Antirrhinum majus*	370	+	CACGTA	cis-acting regulatory element involved in light responsiveness
G-box	*Zea mays*	196	-	CACGAC	cis-acting regulatory element involved in light responsiveness
G-box	*Zea mays*	412	-	CACGTC	cis-acting regulatory element involved in light responsiveness
G-box	*Daucus carota*	370	-	TACGTG	cis-acting regulatory element involved in light responsiveness
GC-motif	*Zea mays*	102	+	CCCCCG	enhancer-like element involved in anoxic specific inducibility
GC-motif	*Zea mays*	257	-	CCCCCG	enhancer-like element involved in anoxic specific inducibility
GT1-motif	*Arabidopsis thaliana*	261	+	GGTTAA	light responsive element
Sp1	*Zea mays*	300	+	CC(G/A)CCC	light responsive element
TCCACCT-motif	*Petroselinum hortense*	296	+	TCCACCT	
TCCC-motif	*Spinacia oleracea*	225	-	TCTCCCT	part of a light responsive element
TCT-motif	*Arabidopsis thaliana*	337	+	TCTTAC	part of a light responsive element
TGACG-motif	*Hordeum vulgare*	411	+	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness

Table A2.16: Potential plant cis-acting regulatory DNA binding motifs in InIII-78 from New PLACE online database. Search done on Nov 13, 2017.

Motif Name	Location	Signal sequence	Site#
RBCSCONSENSUS	16 (-)	AATCCAA	S000127
ARR1AT	18 (+)	NGATT	S000454
RYREPEATBNNAPA	24 (+)	CATGCA	S000264
CAATBOX1	44 (+)	CAAT	S000028
ARR1AT	45 (-)	NGATT	S000454
CACTFTPPCA1	63 (-)	YACT	S000449
NODCON2GM	68 (+)	CTCTT	S000462
OSE2ROOTNODULE	68 (+)	CTCTT	S000468
EBOXBNNAPA	74 (-)	CANNTG	S000144
MYCCONSENSUSAT	74 (-)	CANNTG	S000407
EBOXBNNAPA	74 (+)	CANNTG	S000144
MYCCONSENSUSAT	74 (+)	CANNTG	S000407
RAV1AAT	78 (-)	CAACA	S000314
SURECOREATSULTR11	92 (-)	GAGAC	S000499
MYB2CONSENSUSAT	115 (-)	YAACKG	S000409
MYBCOREATCYCB1	115 (-)	AACGG	S000502
MYBCORE	115 (+)	CNGTTR	S000176
SEBFCONSSTPR10A	124 (+)	YTGTCWC	S000391
ARFAT	125 (+)	TGTCTC	S000270
SURECOREATSULTR11	126 (-)	GAGAC	S000499
ARFAT	137 (-)	TGTCTC	S000270
SEBFCONSSTPR10A	137 (-)	YTGTCWC	S000391
SURECOREATSULTR11	137 (+)	GAGAC	S000499
SURECOREATSULTR11	164 (+)	GAGAC	S000499
TATABOX5	170 (-)	TTATTT	S000203
ARR1AT	179 (-)	NGATT	S000454
GTGANTG10	181 (-)	GTGA	S000378
ACGTCBOX	187 (-)	GACGTC	S000131
ACGTCBOX	187 (+)	GACGTC	S000131
ACGTATERD1	188 (-)	ACGT	S000415
ACGTATERD1	188 (+)	ACGT	S000415
CGACGOSAMY3	189 (-)	CGACG	S000205
RHERPATEXPA7	191 (-)	KCACGW	S000512
CGACGOSAMY3	202 (-)	CGACG	S000205
SURECOREATSULTR11	214 (-)	GAGAC	S000499
MYBCORE	232 (-)	CNGTTR	S000176
CGCGBOXAT	248 (-)	VCGCGB	S000501
CGCGBOXAT	248 (+)	VCGCGB	S000501
GT1CORE	255 (+)	GGTTAA	S000125
E2FCONSENSUS	311 (+)	WTTSSCSS	S000476
BS1EGCCR	316 (-)	AGCGGG	S000352
BOXCPSAS1	320 (+)	CTCCCAC	S000226
CACTFTPPCA1	324 (+)	YACT	S000449
NODCON2GM	330 (+)	CTCTT	S000462
OSE2ROOTNODULE	330 (+)	CTCTT	S000468
BS1EGCCR	336 (-)	AGCGGG	S000352
CURECORECR	361 (-)	GTAC	S000493
CURECORECR	361 (+)	GTAC	S000493
ABRELATERD1	364 (-)	ACGTG	S000414
ACGTATERD1	365 (-)	ACGT	S000415

ACGTATERD1	365 (+) ACGT	S000415
MYB1AT	368 (+) WAACCA	S000408
DOFCOREZM	380 (-) AAAG	S000265
E2FCONSENSUS	381 (+) WTTSSCSS	S000476
PALBOXAPC	387 (+) CCGTCC	S000137
GATABOX	397 (-) GATA	S000039
MYBST1	397 (-) GGATA	S000180
REBETALGLHCB21	397 (-) CGGATA	S000363
CGCGBOXAT	400 (-) VCGCGB	S000501
ABRERATCAL	400 (-) MACGYGB	S000507
CGCGBOXAT	400 (+) VCGCGB	S000501
GTGANTG10	404 (+) GTGA	S000378
HEXMOTIFTAH3H4	405 (-) ACGTCA	S000053
ASF1MOTIFCAMV	405 (+) TGACG	S000024
TGACGTVMAMY	405 (+) TGACGT	S000377
WRKY71OS	405 (+) TGAC	S000447
ACGTATERD1	407 (-) ACGT	S000415
RHERPATEXPA7	407 (-) KCACGW	S000512
ABRELATERD1	407 (+) ACGTG	S000414
ACGTATERD1	407 (+) ACGT	S000415
GTGANTG10	409 (+) GTGA	S000378
SEBFCONSSTPR10A	415 (+) YTGTCWC	S000391
ARFAT	416 (+) TGTCTC	S000270
SURECOREATSULTR11	417 (-) GAGAC	S000499

Table A2.17: Potential eukaryotic transcription factor binding motifs in InIII-78 from JASPAR Core online database. Search done on Oct 28, 2017

Name	Organism	Start	End	Strand	Predicted sequence	Function	Matrix ID
GATA2	Homo sapiens	403	407	-	GGATA	GATA-type zinc fingers	MA0036.1
ECM22	Saccharomyces cerevisiae	42	48	+	CTCCGGA	C6 zinc cluster factors	MA0292.1
HAP2	Saccharomyces cerevisiae	173	177	-	TTGGT	Heteromeric CCAAT-binding factors	MA0313.1
MSN2	Saccharomyces cerevisiae	165	169	+	AGGGG	C6 zinc cluster factors	MA0341.1
MSN4	Saccharomyces cerevisiae	165	169	+	AGGGG	C6 zinc cluster factors	MA0342.1
SIP4	Saccharomyces cerevisiae	42	48	+	CTCCGGA	C6 zinc cluster factors	MA0380.1
SUT1	Saccharomyces cerevisiae	255	261	+	CGCGGGG	C6 zinc cluster factors	MA0399.1
NAC055	Arabidopsis thaliana	369	376	+	ACACGTAA		MA0937.1

Table A2.18: Novel motif sequences in InIII-78 found using WEEDER2.0 motif discovery tool. Search done on Nov 12, 13 2017.

Motif Sequence (WEEDER output)	Output: Arabidopsis - Tomtom (Motif Comparison Tool of known motifs)
AAATAACG	MYB_tnt.MYB56_col_a_m1 (MYB56), MYB_tnt.MYB56_colamp_a_m1 (MYB56), MYB_tnt.MYB105_colamp_a_m1 (MYB105)
AAATAACGAA	MYB_tnt.MYB56_col_a_m1 (MYB56), MYB_tnt.MYB56_colamp_a_m1 (MYB56)
AAATCA	No motif matches

Table A2.19: Results for Repressor Sequence Search from bioinformatics data

Repressor Type/Gene	Insulators (No. of sites)	Protein Motif	Consensus Sequence	Reference	Databases
EAR (ERF-associated amphiphilic repression)	InI-3 (3) InII-10 (1) InII-12 (1) InIII-50 (1)	WBOXNTERF3	TGACY	Ohta *et al.,* 2000	Plant CARE
R/KLFGV	InI-3 (3) InI-6 (3) InII-10 (1) InII-12 (3) InIII-50 (3) InIII-78 (3)	ARFAT	TGTCTC	Ikeda *et al.,* 2009	Plant CARE
R/KLFGV	InI-3 (1) InIII-50 (1) InIII-78 (1)	RAV1AAT	CAACA	Ikeda *et al.,* 2009	Plant CARE
R/KLFGV	InI-3 (2) InI-6 (1)	MYB2AT	TAACTG	Kawase *et al.,* 2017	Plant CARE
R/KLFGV	InI-3 (2) InI-6 (1) InIII-50 (1)	MYB2CONSENSUSAT	YAACKG	Kawase *et al.,* 2017	Plant CARE

	InIII-78 (1)			
WRKY gene	InI-3 (3+2)	WRKY71OS	TGAC	Plant CARE
	InI-6 (1)		AAAAAGTC	WEEDER2.0
	InII-12 (1)		AAAAGT	
	InIII-50 (1)			
	InIII-78 (1)			

Appendix – Chapter 3

DNA Analysis of *Arabidopsis thaliana* transformed with UAS_5'end sequence

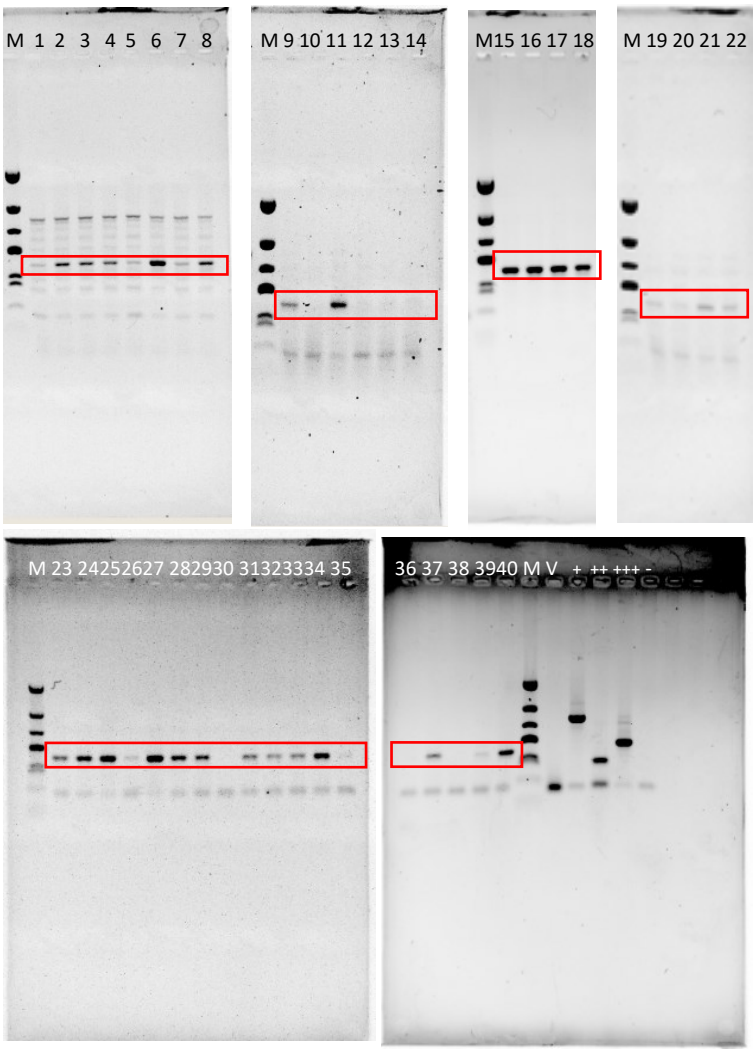


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone UAS_5'end in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate UAS_5'end insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 319bp.
M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-40: Transgenic plant DNA containing UAS_5'endpL1 inserts (319bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
++ : I-3Δ3 sequenced control (312bp)
+++ : UASrpg sequenced control (416bp)
- : Negative control containing water in replace of DNA

Alignment of "UAS_5'end" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_UAS_5'end    #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
>UAS_5'end_S6B      #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
>UAS_5'end_S7B      #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
>UAS_5'end_S8B      #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
>UAS_5'end_S9B      #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
>UAS_5'end_S12B     #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
                                      ..................................................
 UAS_5'end          #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
 Consensus

>Input_UAS_5'end    #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
>UAS_5'end_S6B      #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
>UAS_5'end_S7B      #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
>UAS_5'end_S8B      #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
>UAS_5'end_S9B      #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
>UAS_5'end_S12B     #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
                                      ..................................................
 UAS_5'end          #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA


>Input_UAS_5'end    #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
>UAS_5'end_S6B      #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
>UAS_5'end_S7B      #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
>UAS_5'end_S8B      #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
>UAS_5'end_S9B      #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
>UAS_5'end_S12B     #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
                                      .................................................
 UAS_5'end          #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with UAS_mR1 sequence

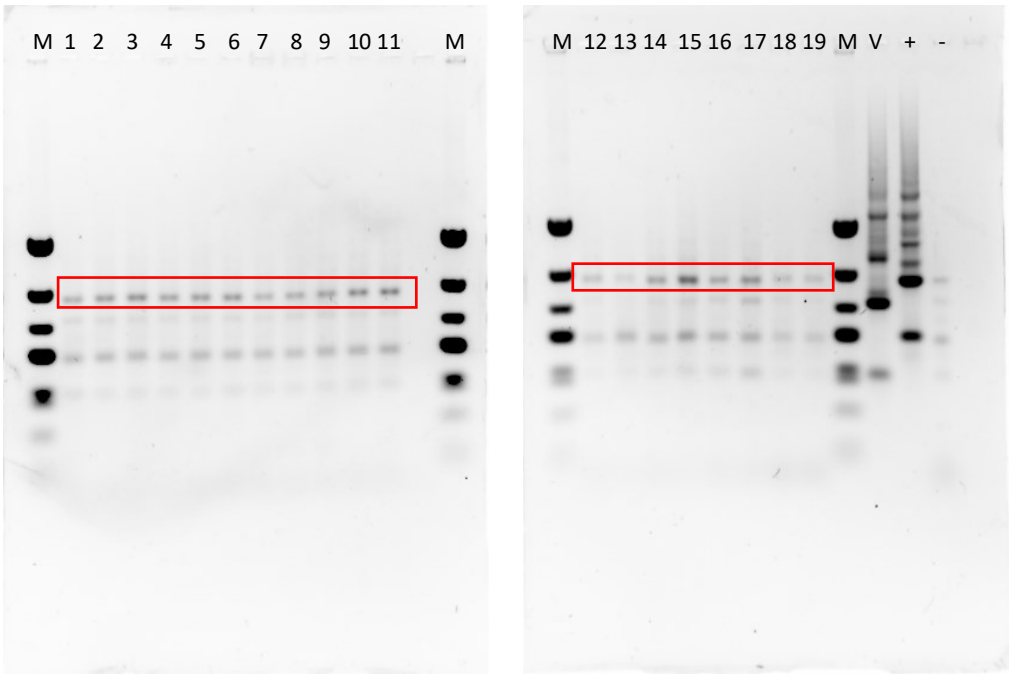


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone UAS_mR1 in pL1 with 1381F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate UAS_mR1 insulator are shown. Amplification of DNA used 1381F and NapinSeqR primers to produce a band with an expected size of 649bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80

1-19 : Transgenic plant DNA containing UAS_mR1 pL1 inserts (649bp)

V : pL1 control vector (no insert, 500bp)

+ : UASrpg sequenced control (746bp)

- : Negative control containing water in replace of DNA

Alignment of "UAS_mR1" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_UAS_mR1      #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR1_S3         #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR1_S4-58      #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR1_S6         #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR1_S7         #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR1_S10-89     #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
                                      ..................................................
 UAS_mR1            #1                AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
 Consensus

>Input_UAS_mR1      #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGAACATACA TCCCCATGTA
<UAS_mR1_S3         #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGAACATACA TCCCCATGTA
<UAS_mR1_S4-58      #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGAACATACA TCCCCATGTA
<UAS_mR1_S6         #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGAACATACA TCCCCATGTA
<UAS_mR1_S7         #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGAACATACA TCCCCATGTA
<UAS_mR1_S10-89     #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGAACATACA TCCCCATGTA
                                      ..................................................
 UAS_mR1            #51               GGGCATGATG TGACTGTCGC CCGTACATTT AGAACATACA TCCCCATGTA


>Input_UAS_mR1      #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR1_S3         #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR1_S4-58      #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR1_S6         #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR1_S7         #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR1_S10-89     #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
                                      .................................................
 UAS_mR1            #101              TAATCATTTG CATCCATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with UAS_mR2 sequence

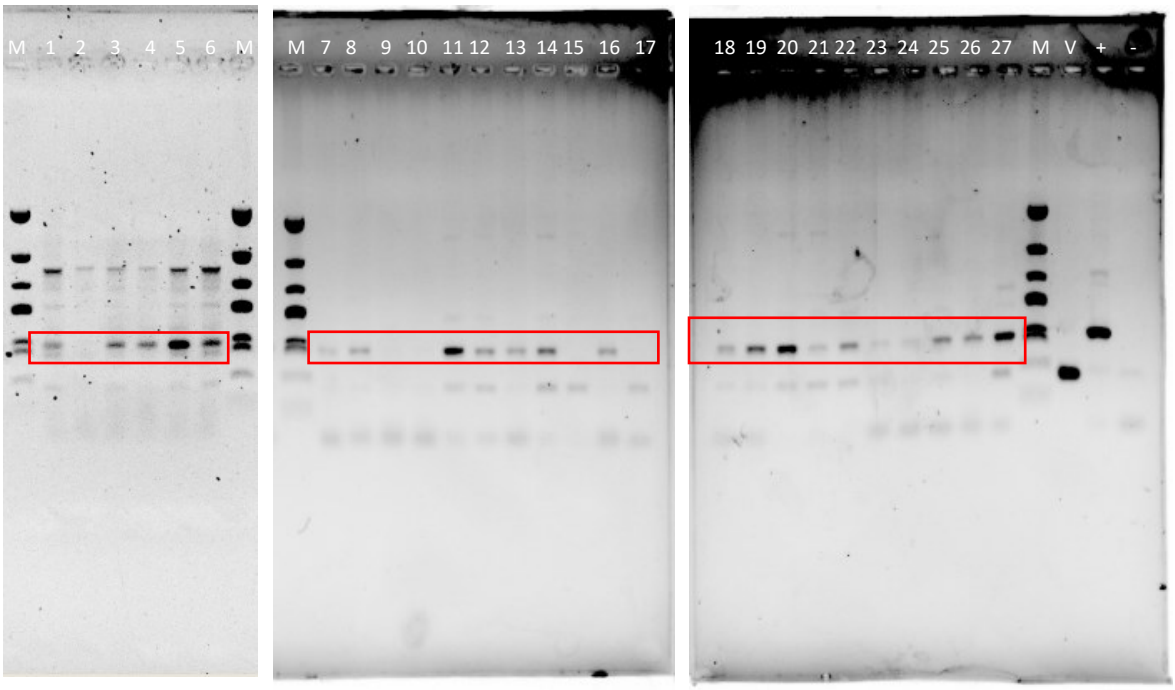


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone UAS_mR2 in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate UAS_mR2 insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 319bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80

1-27 : Transgenic plant DNA containing UAS_mR2 pL1 inserts (319bp)

V : pL1 control vector (no insert, 170bp)

+ : UASrpg sequenced control (416bp)

- : Negative control containing water in replace of DNA

Alignment of "UAS_mR2" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_UAS_mR2      #1              AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR2_S5         #1              AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR2_S9         #1              AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR2_S12        #1              AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR2_S13        #1              AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
<UAS_mR2_S14        #1              AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
                                    ..................................................
 UAS_mR2            #1              AAGCTTGGCC CAGAATACCC TCCTTGACAG TCTTGACGTG CGCAGCTCAG
 Consensus

>Input_UAS_mR2      #51             GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
<UAS_mR2_S5         #51             GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
<UAS_mR2_S9         #51             GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
<UAS_mR2_S12        #51             GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
<UAS_mR2_S13        #51             GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
<UAS_mR2_S14        #51             GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA
                                    ..................................................
 UAS_mR2            #51             GGGCATGATG TGACTGTCGC CCGTACATTT AGCCCATACA TCCCCATGTA


>Input_UAS_mR2      #101            TAATCATTTG CATAAATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR2_S5         #101            TAATCATTTG CATAAATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR2_S9         #101            TAATCATTTG CATAAATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR2_S12        #101            TAATCATTTG CATAAATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR2_S13        #101            TAATCATTTG CATAAATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
<UAS_mR2_S14        #101            TAATCATTTG CATAAATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
                                    .................................................
 UAS_mR2            #101            TAATCATTTG CATAAATACA TTTTGATGGC CGCACGGCGC GAACTGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with UAS_3'end sequence

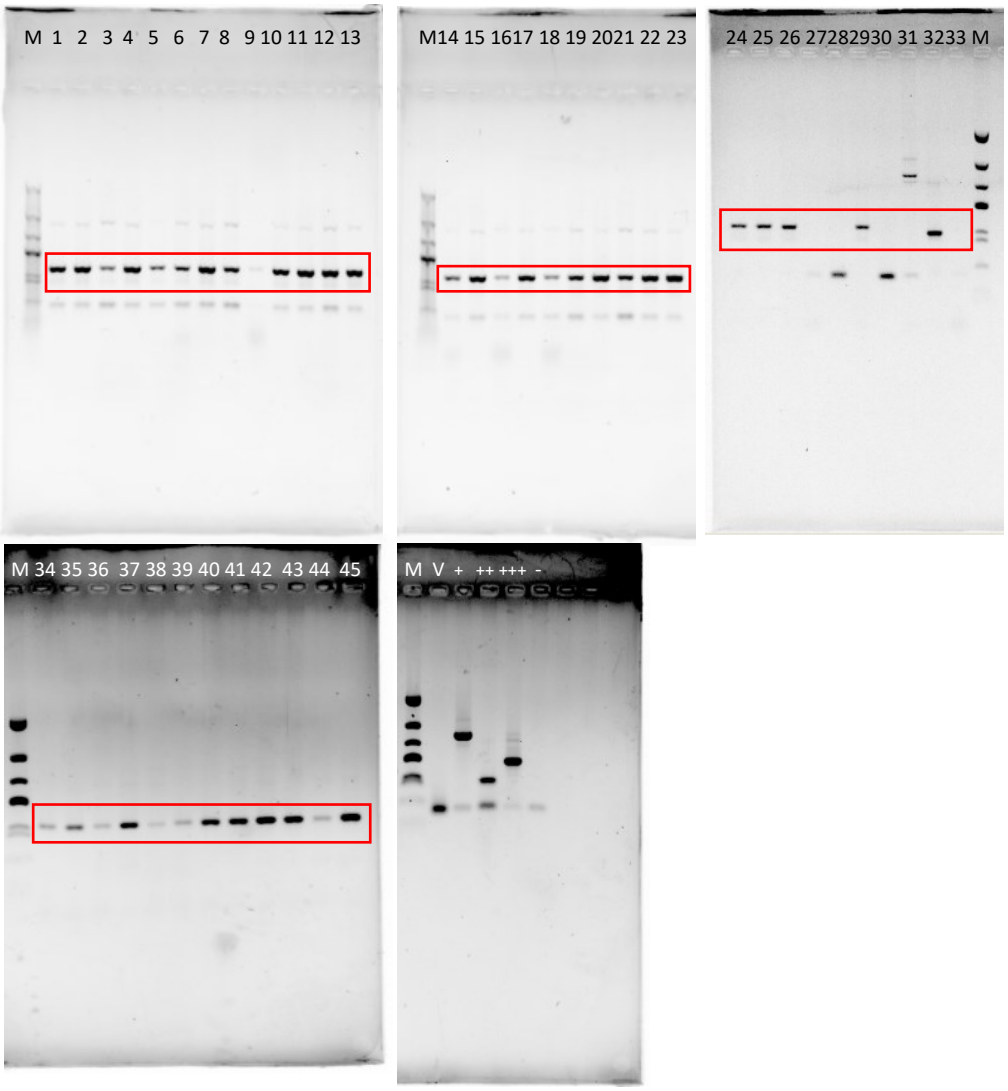


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone UAS_3'end in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate UAS_3'end insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 279bp.
M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-45 : Transgenic plant DNA containing UAS_3'endpL1 inserts (279bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
++ : I-3Δ3 sequenced control (312bp)
+++ : UASrpg sequenced control (416bp)
- : Negative control containing water in replace of DNA

Alignment of "UAS_3'end" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_UAS_3'end    #1                AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_3'end_S4-18    #1                AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_3'end_S10-18   #1                AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_3'end_S4-23    #1                AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_3'end_S7-23    #1                AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_3'end_S12-23   #1                AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
                                      ..................................................
 UAS_3'end          #1                AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC


>Input_UAS_3'end    #51               GCTCCCCTCA CAGACGCGTT GAATTCTCCC CACGGCGCGC CCCTGTAGAG
<UAS_3'end_S4-18    #51               GCTCCCCTCA CAGACGCGTT GAATTCTCCC CACGGCGCGC CCCTGTAGAG
<UAS_3'end_S10-18   #51               GCTCCCCTCA CAGACGCGTT GAATTCTCCC CACGGCGCGC CCCTGTAGAG
<UAS_3'end_S4-23    #51               GCTCCCCTCA CAGACGCGTT GAATTCTCCC CACGGCGCGC CCCTGTAGAG
<UAS_3'end_S7-23    #51               GCTCCCCTCA CAGACGCGTT GAATTCTCCC CACGGCGCGC CCCTGTAGAG
<UAS_3'end_S12-23   #51               GCTCCCCTCA CAGACGCGTT GAATTCTCCC CACGGCGCGC CCCTGTAGAG
                                      ..................................................
 UAS_3'end          #51               GCTCCCCTCA CAGACGCGTT GAATTCTCCC CACGGCGCGC CCCTGTAGAG


>Input_UAS_3'end    #101              AAACTGCAG
<UAS_3'end_S4-18    #101              AAACTGCAG
<UAS_3'end_S10-18   #101              AAACTGCAG
<UAS_3'end_S4-23    #101              AAACTGCAG
<UAS_3'end_S7-23    #101              AAACTGCAG
<UAS_3'end_S12-23   #101              AAACTGCAG
                                      .........
 UAS_3'end          #101              AAACTGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with UAS_ ΔSu(Hw) sequence

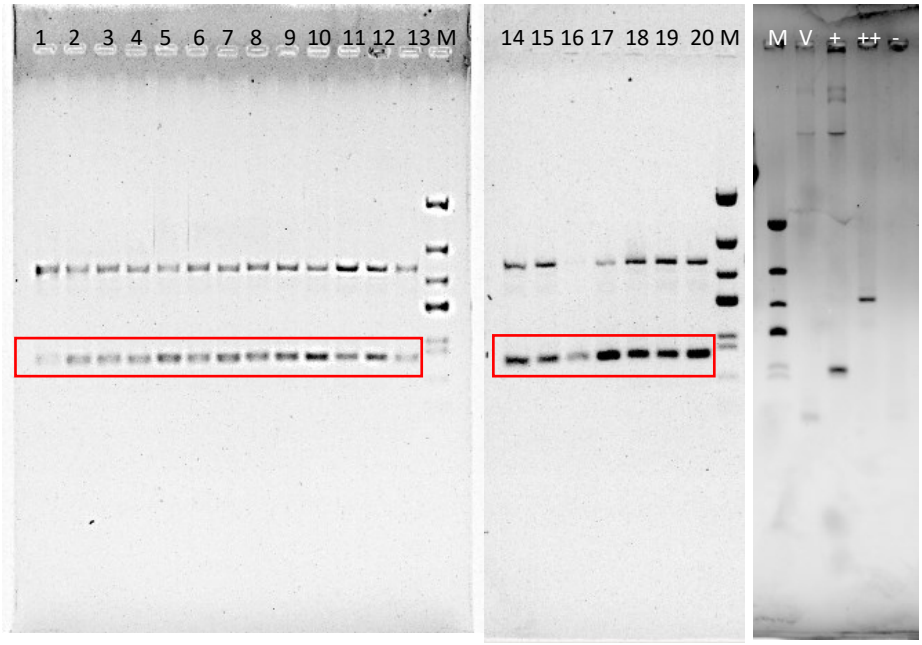


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone UAS_ ΔSu(Hw) in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate UAS_ ΔSu(Hw) insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 270bp.
M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-20 : Transgenic plant DNA containing UAS_ ΔSu(Hw)pL1 inserts (270bp)
V : pL1 control vector (no insert, 170bp)
+ : UASrpg sequenced control (416bp)
++ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "UAS_ ΔSu(Hw)" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_UAS_ deltaSu.#1                AAGCTTCGAA AAATTACGGC TCCTCCTGCG AGCAGGGAAA CGCTCCCCTC
<UAS_deltaSu(Hw)_S2 #1                AAGCTTCGAA AAATTACGGC TCCTCCTGCG AGCAGGGAAA CGCTCCCCTC
<UAS_deltaSu(Hw)_S4 #1                AAGCTTCGAA AAATTACGGC TCCTCCTGCG AGCAGGGAAA CGCTCCCCTC
<UAS_deltaSu(Hw)_S5 #1                AAGCTTCGAA AAATTACGGC TCCTCCTGCG AGCAGGGAAA CGCTCCCCTC
<UAS_deltaSu(Hw)_S9 #1                AAGCTTCGAA AAATTACGGC TCCTCCTGCG AGCAGGGAAA CGCTCCCCTC
<UAS_deltaSu(Hw)_S10#1                AAGCTTCGAA AAATTACGGC TCCTCCTGCG AGCAGGGAAA CGCTCCCCTC
                                      ..................................................
 UAS_ deltaSu(Hw)   #1                AAGCTTCGAA AAATTACGGC TCCTCCTGCG AGCAGGGAAA CGCTCCCCTC
 Consensus

>Input_UAS_ deltaSu.#51               ACAGACGCGT TGAATTCTCC CCACGGCGCG CCCCTGTAGA GAAACTGCAG
<UAS_deltaSu(Hw)_S2 #51               ACAGACGCGT TGAATTCTCC CCACGGCGCG CCCCTGTAGA GAAACTGCAG
<UAS_deltaSu(Hw)_S4 #51               ACAGACGCGT TGAATTCTCC CCACGGCGCG CCCCTGTAGA GAAACTGCAG
<UAS_deltaSu(Hw)_S5 #51               ACAGACGCGT TGAATTCTCC CCACGGCGCG CCCCTGTAGA GAAACTGCAG
<UAS_deltaSu(Hw)_S9 #51               ACAGACGCGT TGAATTCTCC CCACGGCGCG CCCCTGTAGA GAAACTGCAG
<UAS_deltaSu(Hw)_S10#51               ACAGACGCGT TGAATTCTCC CCACGGCGCG CCCCTGTAGA GAAACTGCAG
                                      ..................................................
 UAS_ deltaSu(Hw)   #51               ACAGACGCGT TGAATTCTCC CCACGGCGCG CCCCTGTAGA GAAACTGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with UAS_ ΔCTC sequence

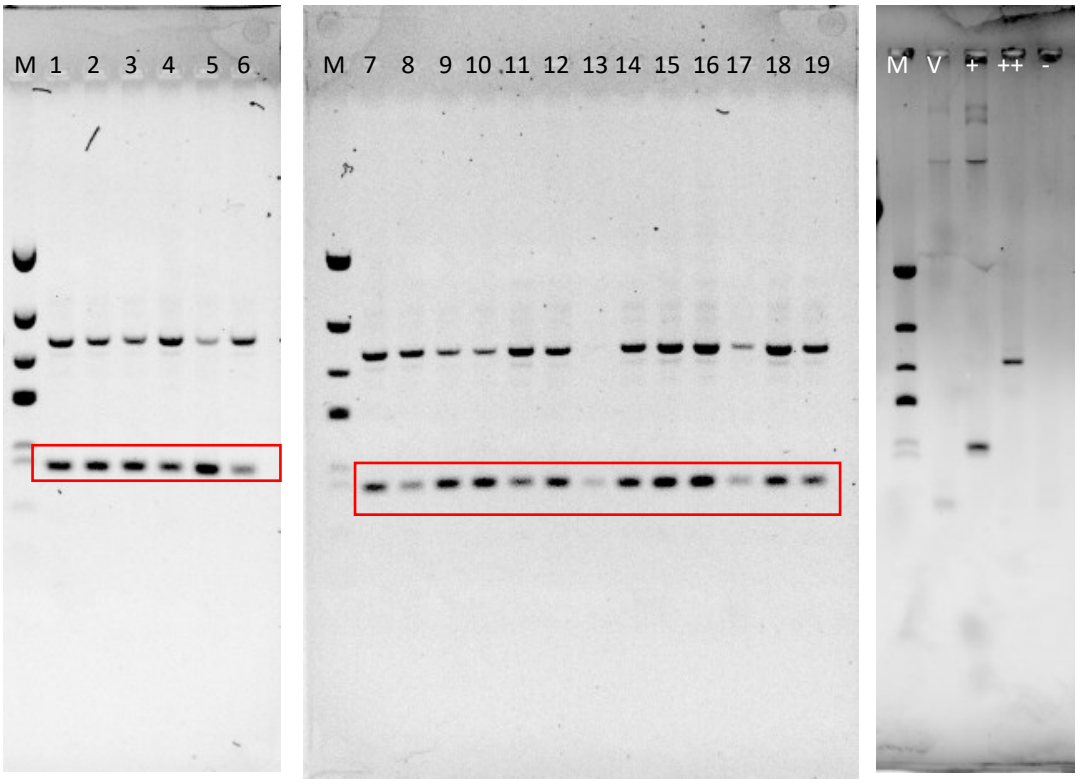


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone UAS_ ΔCTC in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate UAS_ ΔCTC insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 274bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-19 : Transgenic plant DNA containing UAS_ ΔCTCpL1 inserts (274bp)
V : pL1 control vector (no insert, 170bp)
+ : UASrpg sequenced control (416bp)
++ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "UAS_ ΔCTC" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_UAS_deltaCTC #1              AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_deltaCTC_S2    #1              AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_deltaCTC_S3    #1              AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_deltaCTC_S5    #1              AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_deltaCTC_S7    #1              AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
<UAS_deltaCTC_S9    #1              AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC
                                    ..................................................
 AT2_CTC            #1              AAGCTTCGAA AAATTACGGC TCCTCGCTGC AGACCTGCGA GCAGGGAAAC


>Input_UAS_deltaCTC #51             GCTCACAGAC GCGTTGAATT CTCCCCACGG CGCGCCCCTG TAGAGAAACT
<UAS_deltaCTC_S2    #51             GCTCACAGAC GCGTTGAATT CTCCCCACGG CGCGCCCCTG TAGAGAAACT
<UAS_deltaCTC_S3    #51             GCTCACAGAC GCGTTGAATT CTCCCCACGG CGCGCCCCTG TAGAGAAACT
<UAS_deltaCTC_S5    #51             GCTCACAGAC GCGTTGAATT CTCCCCACGG CGCGCCCCTG TAGAGAAACT
<UAS_deltaCTC_S7    #51             GCTCACAGAC GCGTTGAATT CTCCCCACGG CGCGCCCCTG TAGAGAAACT
<UAS_deltaCTC_S9    #51             GCTCACAGAC GCGTTGAATT CTCCCCACGG CGCGCCCCTG TAGAGAAACT
                                    ..................................................
 AT2_CTC            #51             GCTCACAGAC GCGTTGAATT CTCCCCACGG CGCGCCCCTG TAGAGAAACT


>Input_UAS_deltaCTC #101            GCAG
<UAS_deltaCTC_S2    #101            GCAG
<UAS_deltaCTC_S3    #101            GCAG
<UAS_deltaCTC_S5    #101            GCAG
<UAS_deltaCTC_S7    #101            GCAG
<UAS_deltaCTC_S9    #101            GCAG
                                    ....
 AT2_CTC            #101            GCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with BEAD1c_5'end sequence

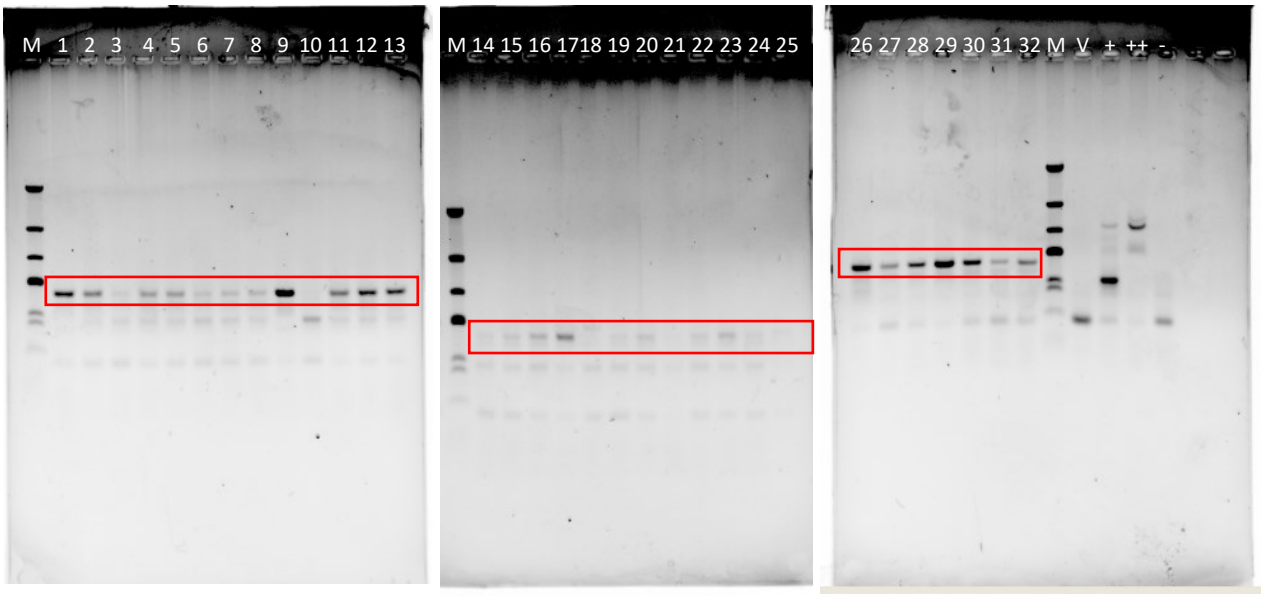


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone BEAD1c_5'end in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate BEAD1c_5'end insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 435bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-32 : Transgenic plant DNA containing BEAD1c_5'end pL1 inserts (435bp)
V : pL1 control vector (no insert, 170bp)
+ : UASrpg sequenced control (416bp)
++ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "BEAD1c_5′end" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_BEAD1c_5'end #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
<BEAD1c_5'end_S21   #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
<BEAD1c_5'end_S24   #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
>BEAD1c_5'end_S28   >#1>                   AGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
<BEAD1c_5'end_S29   #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
<BEAD1c_5'end_S31   #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
                                    ..................................................
 BEAD1c_5'end       #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
 Consensus

>Input_BEAD1c_5'end #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
<BEAD1c_5'end_S21   #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
<BEAD1c_5'end_S24   #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
>BEAD1c_5'end_S28   #44             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
<BEAD1c_5'end_S29   #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
<BEAD1c_5'end_S31   #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
                                    ..................................................
 BEAD1c_5'end       #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA


>Input_BEAD1c_5'end #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
<BEAD1c_5'end_S21   #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
<BEAD1c_5'end_S24   #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
>BEAD1c_5'end_S28   #94             TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
<BEAD1c_5'end_S29   #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
<BEAD1c_5'end_S31   #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
                                    ..................................................
 BEAD1c_5'end       #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC


>Input_BEAD1c_5'end #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
<BEAD1c_5'end_S21   #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
<BEAD1c_5'end_S24   #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
>BEAD1c_5'end_S28   #144            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
<BEAD1c_5'end_S29   #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
<BEAD1c_5'end_S31   #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
```

```
                                                  ..................................................
 BEAD1c_5'end       #151                          ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT


>Input_BEAD1c_5'end #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
<BEAD1c_5'end_S21   #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
<BEAD1c_5'end_S24   #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
>BEAD1c_5'end_S28   #194                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
<BEAD1c_5'end_S29   #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
<BEAD1c_5'end_S31   #201                          GGAAGAGGGA TGTTGAGGGC CCYGGGGGGG CCNCGKCGGS GWATTGGCTG
                                                  ..................................................
 BEAD1c_5'end       #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
                                                                        +    **    +* +   +  +


>Input_BEAD1c_5'end #251                          CCCAGGCCTC TGCAG
<BEAD1c_5'end_S21   #251                          CCCAGGGGTC GGAAT
<BEAD1c_5'end_S24   #251                          CCCAGGCNGC GGAAT
>BEAD1c_5'end_S28   #244                          CCCAGGCNGC GGAAT
<BEAD1c_5'end_S29   #251                          CCCAGGNNGC GGAAT
<BEAD1c_5'end_S31   #251                          CCCAG
                                                  ..................................................
 BEAD1c_5'end       #251                          CCCAGGCSGC GGAAT
                                                        ***  * * *
```

DNA Analysis of *Arabidopsis thaliana* transformed with BEAD1c_3'end sequence

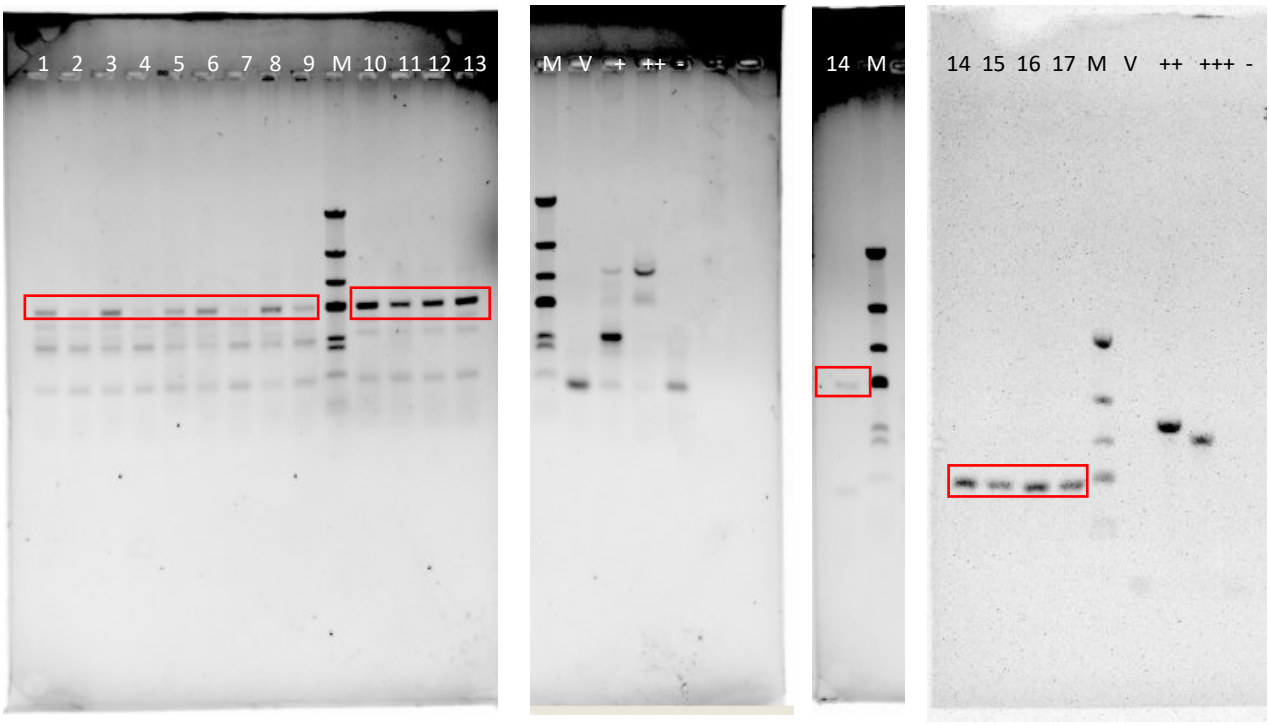


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone BEAD1c_3'end in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate BEAD1c_3'end insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 471bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-17 : Transgenic plant DNA containing BEAD1c_3'end pL1 inserts (471bp)
V : pL1 control vector (no insert, 170bp)
+ : UASrpg sequenced control (416bp)
++ : BEAD1c sequenced control (718bp)
+++ : III-7pL1 sequenced control (628bp)
- : Negative control containing water in replace of DNA

DNA Analysis of *Arabidopsis thaliana* transformed with BEAD1c_3'end sequence (continued)

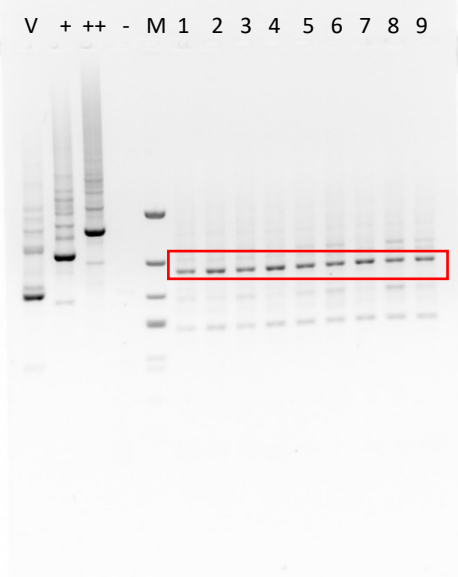


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone BEAD1c_3'end in pL1 with 1381F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate BEAD1c_3'end insulator are shown. Amplification of DNA used 1381F and NapinSeqR primers to produce a band with an expected size of 801bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-9 : Transgenic plant DNA containing BEAD1c_3'end pL1 inserts (801bp)
V : pL1 control vector (no insert, 500bp)
+ : UASrpg sequenced control (746bp)
++ : BEAD1c sequenced control (1,048bp)
- : Negative control containing water in replace of DNA

Alignment of "BEAD1c_3′end" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_BEAD1c_3'end #1              AAGCTTAGGC CTGCACTGCC GCCTGCCGGC AGGGGTCCAG TCCACGAGAC
<BEAD1c_3'end_S1    >#1>                 TAGGC CTGCACTGCC GCCTGCCGGC AGGGGTCCAG TCCACGAGAC
<BEAD1c_3'end_S2    #1             AAGCTTAGGC CTGCACTGCC GCCTGCCGGC AGGGGTCCAG TCCACGAGAC
<BEAD1c_3'end_S3    #1             AAGCTTAGGC CTGCACTGCC GCCTGCCGGC AGGGGTCCAG TCCACGAGAC
<BEAD1c_3'end_S5    #1             AAGCTTAGGC CTGCACTGCC GCCTGCCGGC AGGGGTCCAG TCCACGAGAC
>BEAD1c_3'end_S6    >#1>                           GCC GCCTGCCGGC AGGGGTCCAG TCCACGAGAC
                                   ..................................................
 BEAD1c_3'end       #1             AAGCTTAGGC CTGCACTGCC GCCTGCCGGC AGGGGTCCAG TCCACGAGAC


>Input_BEAD1c_3'end #51             CCAGCTCCCT GCTGGCGGAA GTCCATTTCA GAGCTTCCGG TTCTCCCAAG
<BEAD1c_3'end_S1    #46             CCAGCTCCCT GCTGGCGGAA GTCCATTTCA GAGCTTCCGG TTCTCCCAAG
<BEAD1c_3'end_S2    #51             CCAGCTCCCT GCTGGCGGAA GTCCATTTCA GAGCTTCCGG TTCTCCCAAG
<BEAD1c_3'end_S3    #51             CCAGCTCCCT GCTGGCGGAA GTCCATTTCA GAGCTTCCGG TTCTCCCAAG
<BEAD1c_3'end_S5    #51             CCAGCTCCCT GCTGGCGGAA GTCCATTTCA GAGCTTCCGG TTCTCCCAAG
>BEAD1c_3'end_S6    #34             CCAGCTCCCT GCTGGCGGAA GTCCATTTCA GAGCTTCCGG TTCTCCCAAG
                                   ..................................................
 BEAD1c_3'end       #51             CCAGCTCCCT GCTGGCGGAA GTCCATTTCA GAGCTTCCGG TTCTCCCAAG


>Input_BEAD1c_3'end #101            TCCAAGGATT ATGCTCACTC CCCACCCACA GTCTCTTAGT GTCTGTCCCT
<BEAD1c_3'end_S1    #96             TCCAAGGATT ATGCTCACTC CCCACCCACA GTCTCTTAGT GTCTGTCCCT
<BEAD1c_3'end_S2    #101            TCCAAGGATT ATGCTCACTC CCCACCCACA GTCTCTTAGT GTCTGTCCCT
<BEAD1c_3'end_S3    #101            TCCAAGGATT ATGCTCACTC CCCACCCACA GTCTCTTAGT GTCTGTCCCT
<BEAD1c_3'end_S5    #101            TCCAAGGATT ATGCTCACTC CCCACCCACA GTCTCTTAGT GTCTGTCCCT
>BEAD1c_3'end_S6    #84             TCCAAGGATT ATGCTCACTC CCCACCCACA GTCTCTTAGT GTCTGTCCCT
                                   ..................................................
 BEAD1c_3'end       #101            TCCAAGGATT ATGCTCACTC CCCACCCACA GTCTCTTAGT GTCTGTCCCT


>Input_BEAD1c_3'end #151            GCTCTAAAGA TGTTGTCTGG GCTTGATATT AATATGAGAG CTGACTGTTC
<BEAD1c_3'end_S1    #146            GCTCTAAAGA TGTTGTCTGG GCTTGATATT AATATGAGAG CTGACTGTTC
<BEAD1c_3'end_S2    #151            GCTCTAAAGA TGTTGTCTGG GCTTGATATT AATATGAGAG CTGACTGTTC
```

```
<BEAD1c_3'end_S3    #151            GCTCTAAAGA TGTTGTCTGG GCTTGATATT AATATGAGAG CTGACTGTTC
<BEAD1c_3'end_S5    #151            GCTCTAAAGA TGTTGTCTGG GCTTGATATT AATATGAGAG CTGACTGTTC
>BEAD1c_3'end_S6    #134            GCTCTAAAGA TGTTGTCTGG GCTTGATATT AATATGAGAG CTGACTGTTC
                                    ..................................................
 BEAD1c_3'end       #151            GCTCTAAAGA TGTTGTCTGG GCTTGATATT AATATGAGAG CTGACTGTTC


>Input_BEAD1c_3'end #201            CCTTCCTGAT CTAGACCATA ACCATCTTCA AGTTAAATTG CTCCTCCTCT
<BEAD1c_3'end_S1    #196            CCTTCCTGAT CTAGACCATA ACCATCTTCA AGTTAAATTG CTCCTCCTCT
<BEAD1c_3'end_S2    #201            CCTTCCTGAT CTAGACCATA ACCATCTTCA AGTTAAATTG CTCCTCCTCT
<BEAD1c_3'end_S3    #201            CCTTCCTGAT CTAGACCATA ACCATCTTCA AGTTAAATTG CTCCTCCTCT
<BEAD1c_3'end_S5    #201            CCTTCCTGAT CTAGACCATA ACCATCTTCA AGTTAAATTG CTCCTCCTCT
>BEAD1c_3'end_S6    #184            CCTTCCTGAT CTAGACCATA ACCATCTTCA AGTTAAATTG CTCCTCCTCT
                                    ..................................................
 BEAD1c_3'end       #201            CCTTCCTGAT CTAGACCATA ACCATCTTCA AGTTAAATTG CTCCTCCTCT


>Input_BEAD1c_3'end #251            TCTAACTGCC CAACCTCACC CACGTCTGAC CATACCCAAG CACAGCTGCA
<BEAD1c_3'end_S1    #246            TCTAACTGCC CAACCTCACC CACGTCTGAC CATACCCAAG CACAGCTGCA
<BEAD1c_3'end_S2    #251            TCTAACTGCC CAACCTCACC CACGTCTGAC CATACCCAAG CACAGCTGCA
<BEAD1c_3'end_S3    #251            TCTAACTGCC CAACCTCACC CACGTCTGAC CATACCCAAG CACAGCTGCA
<BEAD1c_3'end_S5    #251            TCTAACTGCC CAACCTCACC CACGTCTGAC CATACCCAAG CACAGCTGCA
>BEAD1c_3'end_S6    #234            TCTAACTGCC CAACCTCACC CACGTCTGAC CATACCCAAG CACAGCTGCA
                                    ..................................................
 BEAD1c_3'end       #251            TCTAACTGCC CAACCTCACC CACGTCTGAC CATACCCAAG CACAGCTGCA


>Input_BEAD1c_3'end #301            G
<BEAD1c_3'end_S1    #296            G
<BEAD1c_3'end_S2    #301            G
<BEAD1c_3'end_S3    #301            G
<BEAD1c_3'end_S5    #301            G
>BEAD1c_3'end_S6    #284            G
                                    ..................................................
 BEAD1c_3'end       #301            G
```

DNA Analysis of *Arabidopsis thaliana* transformed with ΔBEADA sequence

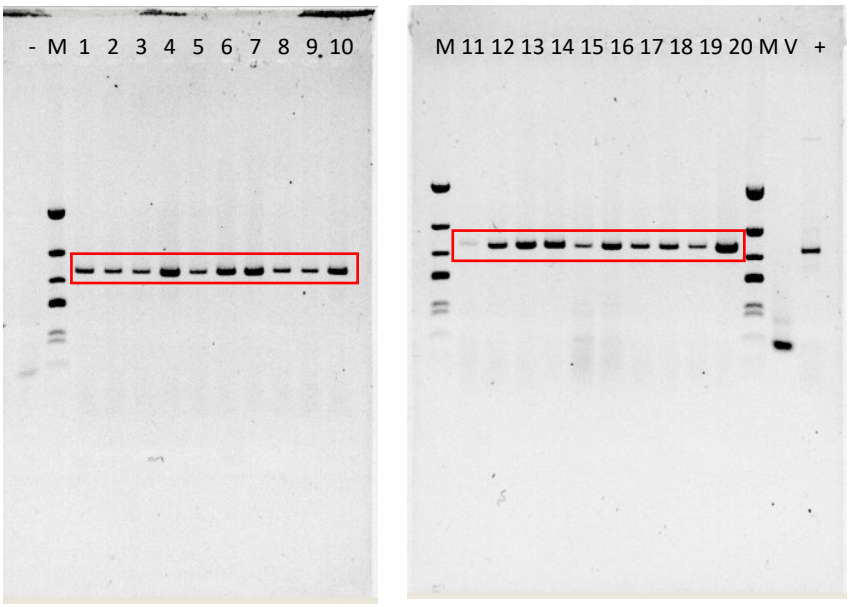


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone ΔBEADA in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate ΔBEADA insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 679bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-20 : Transgenic plant DNA containing ΔBEADA pL1 inserts (679bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "ΔBEADA" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_deltaBEADA   #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
<deltaBEADA_S1      #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
<deltaBEADA_S2      #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
<deltaBEADA_S4      #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
<deltaBEADA_S5      #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
<deltaBEADA_S10     #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT
                                    ..................................................
 BEADA_del          #1              AAGCTTCAGT AATACGGGTA GCTGGGACAT GCCATATTTG GAACACATTT


>Input_deltaBEADA   #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
<deltaBEADA_S1      #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
<deltaBEADA_S2      #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
<deltaBEADA_S4      #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
<deltaBEADA_S5      #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
<deltaBEADA_S10     #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA
                                    ..................................................
 BEADA_del          #51             ATACTAAAAA AGTATTCATT GTTTATCTGA AATTCAAATT CCACTGGGCA


>Input_deltaBEADA   #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
<deltaBEADA_S1      #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
<deltaBEADA_S2      #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
<deltaBEADA_S4      #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
<deltaBEADA_S5      #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
<deltaBEADA_S10     #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC
                                    ..................................................
 BEADA_del          #101            TCCTGTGTTT TATCTGGCAA TGCTAGGCAT GCAGAATACC AAAAGTAAGC


>Input_deltaBEADA   #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
<deltaBEADA_S1      #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
<deltaBEADA_S2      #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
<deltaBEADA_S4      #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
<deltaBEADA_S5      #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
<deltaBEADA_S10     #151            ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT
```

```
                                                  ..................................................
 BEADA_del          #151                          ACCAGGCAGG CCAGAGTCCC ACCATGAGCA TCTTCAGGGC CCCTGGATGT


>Input_deltaBEADA   #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
<deltaBEADA_S1      #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
<deltaBEADA_S2      #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
<deltaBEADA_S4      #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
<deltaBEADA_S5      #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
<deltaBEADA_S10     #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG
                                                  ..................................................
 BEADA_del          #201                          GGAAGAGGGA TGTTGAGGGC CCAGGGGCTG CCTTGCCGGT GCATTGGCTG


>Input_deltaBEADA   #251                          CACGAGACCC AGCTCCCTGC TGGCGGAAGT CCATTTCAGA GCTTCCGGTT
<deltaBEADA_S1      #251                          CACGAGACCC AGCTCCCTGC TGGCGGAAGT CCATTTCAGA GCTTCCGGTT
<deltaBEADA_S2      #251                          CACGAGACCC AGCTCCCTGC TGGCGGAAGT CCATTTCAGA GCTTCCGGTT
<deltaBEADA_S4      #251                          CACGAGACCC AGCTCCCTGC TGGCGGAAGT CCATTTCAGA GCTTCCGGTT
<deltaBEADA_S5      #251                          CACGAGACCC AGCTCCCTGC TGGCGGAAGT CCATTTCAGA GCTTCCGGTT
<deltaBEADA_S10     #251                          CACGAGACCC AGCTCCCTGC TGGCGGAAGT CCATTTCAGA GCTTCCGGTT
                                                  ..................................................
 BEADA_del          #251                          CACGAGACCC AGCTCCCTGC TGGCGGAAGT CCATTTCAGA GCTTCCGGTT


>Input_deltaBEADA   #301                          CTCCCAAGTC CAAGGATTAT GCTCACTCCC CACCCACAGT CTCTTAGTGT
<deltaBEADA_S1      #301                          CTCCCAAGTC CAAGGATTAT GCTCACTCCC CACCCACAGT CTCTTAGTGT
<deltaBEADA_S2      #301                          CTCCCAAGTC CAAGGATTAT GCTCACTCCC CACCCACAGT CTCTTAGTGT
<deltaBEADA_S4      #301                          CTCCCAAGTC CAAGGATTAT GCTCACTCCC CACCCACAGT CTCTTAGTGT
<deltaBEADA_S5      #301                          CTCCCAAGTC CAAGGATTAT GCTCACTCCC CACCCACAGT CTCTTAGTGT
<deltaBEADA_S10     #301                          CTCCCAAGTC CAAGGATTAT GCTCACTCCC CACCCACAGT CTCTTAGTGT
                                                  ..................................................
 BEADA_del          #301                          CTCCCAAGTC CAAGGATTAT GCTCACTCCC CACCCACAGT CTCTTAGTGT


>Input_deltaBEADA   #351                          CTGTCCCTGC TCTAAAGATG TTGTCTGGGC TTGATATTAA TATGAGAGCT
<deltaBEADA_S1      #351                          CTGTCCCTGC TCTAAAGATG TTGTCTGGGC TTGATATTAA TATGAGAGCT
<deltaBEADA_S2      #351                          CTGTCCCTGC TCTAAAGATG TTGTCTGGGC TTGATATTAA TATGAGAGCT
<deltaBEADA_S4      #351                          CTGTCCCTGC TCTAAAGATG TTGTCTGGGC TTGATATTAA TATGAGAGCT
<deltaBEADA_S5      #351                          CTGTCCCTGC TCTAAAGATG TTGTCTGGGC TTGATATTAA TATGAGAGCT
```

```
<deltaBEADA_S10     #351              CTGTCCCTGC TCTAAAGATG TTGTCTGGGC TTGATATTAA TATGAGAGCT
                                      ..................................................
 BEADA_del          #351              CTGTCCCTGC TCTAAAGATG TTGTCTGGGC TTGATATTAA TATGAGAGCT


>Input_deltaBEADA   #401              GACTGTTCCC TTCCTGATCT AGACCATAAC CATCTTCAAG TTAAATTGCT
<deltaBEADA_S1      #401              GACTGTTCCC TTCCTGATCT AGACCATAAC CATCTTCAAG TTAAATTGCT
<deltaBEADA_S2      #401              GACTGTTCCC TTCCTGATCT AGACCATAAC CATCTTCAAG TTAAATTGCT
<deltaBEADA_S4      #401              GACTGTTCCC TTCCTGATCT AGACCATAAC CATCTTCAAG TTAAATTGCT
<deltaBEADA_S5      #401              GACTGTTCCC TTCCTGATCT AGACCATAAC CATCTTCAAG TTAAATTGCT
<deltaBEADA_S10     #401              GACTGTTCCC TTCCTGATCT AGACCATAAC CATCTTCAAG TTAAATTGCT
                                      ..................................................
 BEADA_del          #401              GACTGTTCCC TTCCTGATCT AGACCATAAC CATCTTCAAG TTAAATTGCT


>Input_deltaBEADA   #451              CCTCCTCTTC TAACTGCCCA ACCTCACCCA CGTCTGACCA TACCCAAGCA
<deltaBEADA_S1      #451              CCTCCTCTTC TAACTGCCCA ACCTCACCCA CGTCTGACCA TACCCAAGCA
<deltaBEADA_S2      #451              CCTCCTCTTC TAACTGCCCA ACCTCACCCA CGTCTGACCA TACCCAAGCA
<deltaBEADA_S4      #451              CCTCCTCTTC TAACTGCCCA ACCTCACCCA CGTCTGACCA TACCCAAGCA
<deltaBEADA_S5      #451              CCTCCTCTTC TAACTGCCCA ACCTCACCCA CGTCTGACCA TACCCAAGCA
<deltaBEADA_S10     #451              CCTCCTCTTC TAACTGCCCA ACCTCACCCA CGTCTGACCA TACCCAAGCA
                                      ..................................................
 BEADA_del          #451              CCTCCTCTTC TAACTGCCCA ACCTCACCCA CGTCTGACCA TACCCAAGCA


>Input_deltaBEADA   #501              CAGCTGCAG
<deltaBEADA_S1      #501              CAGCTGCAG
<deltaBEADA_S2      #501              CAGCTGCAG
<deltaBEADA_S4      #501              CAGCTGCAG
<deltaBEADA_S5      #501              CAGCTGCAG
<deltaBEADA_S10     #501              CAGCTGCAG
                                      .........
 BEADA_del          #501              CAGCTGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with gypsy sequence

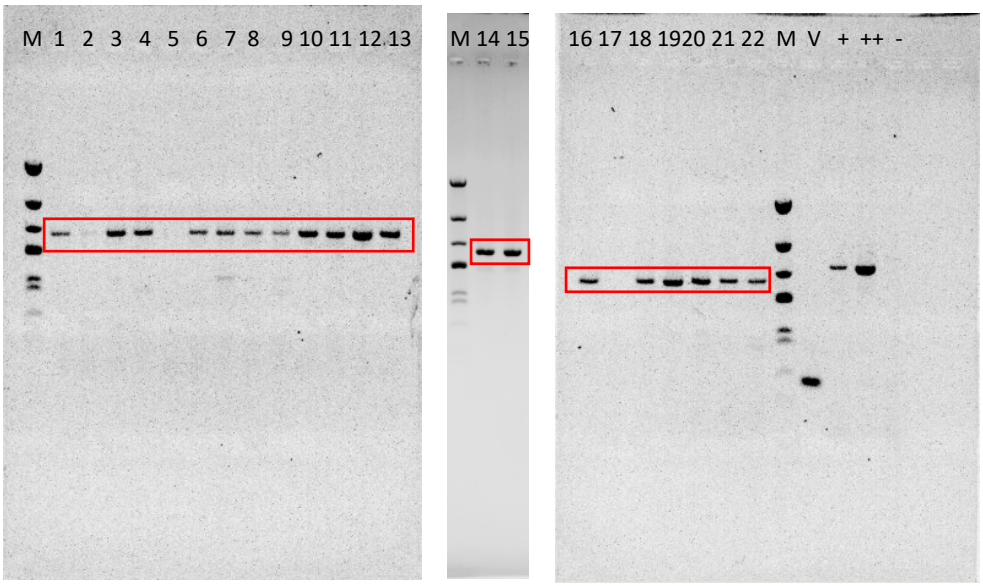


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone gypsy in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate gypsy insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 522bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-22 : Transgenic plant DNA containing gypsy pL1 inserts (522bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
++ : III-7pL1 sequenced control (628bp)
- : Negative control containing water in replace of DNA

Alignment of "gypsy" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_gypsy        #1              AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTGCAT
<gypsy_S1B          #1              AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTGCAT
<gypsy_S9B          #1              AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTGCAT
<gypsy_S10B         #1              AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTGCAT
<gypsy_S11B         #1              AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTGCAT
<gypsy_S12B         #1              AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTGCAT
                                    ..................................................
 gypsy              #1              AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTGCAT
 Consensus

>Input_gypsy        #51             ATTTTCGGCA AAGTAAAATT TTGTTGCATA CCTTATCAAA AAATAAGTGC
<gypsy_S1B          #51             ATTTTCGGCA AAGTAAAATT TTGTTGCATA CCTTATCAAA AAATAAGTGC
<gypsy_S9B          #51             ATTTTCGGCA AAGTAAAATT TTGTTGCATA CCTTATCAAA AAATAAGTGC
<gypsy_S10B         #51             ATTTTCGGCA AAGTAAAATT TTGTTGCATA CCTTATCAAA AAATAAGTGC
<gypsy_S11B         #51             ATTTTCGGCA AAGTAAAATT TTGTTGCATA CCTTATCAAA AAATAAGTGC
<gypsy_S12B         #51             ATTTTCGGCA AAGTAAAATT TTGTTGCATA CCTTATCAAA AAATAAGTGC
                                    ..................................................
 gypsy              #51             ATTTTCGGCA AAGTAAAATT TTGTTGCATA CCTTATCAAA AAATAAGTGC


>Input_gypsy        #101            TGCATACTTT TTAGAGAAAC CAAATAATTT TTTATTGCAT ACCCGTTTTT
<gypsy_S1B          #101            TGCATACTTT TTAGAGAAAC CAAATAATTT TTTATTGCAT ACCCGTTTTT
<gypsy_S9B          #101            TGCATACTTT TTAGAGAAAC CAAATAATTT TTTATTGCAT ACCCGTTTTT
<gypsy_S10B         #101            TGCATACTTT TTAGAGAAAC CAAATAATTT TTTATTGCAT ACCCGTTTTT
<gypsy_S11B         #101            TGCATACTTT TTAGAGAAAC CAAATAATTT TTTATTGCAT ACCCGTTTTT
<gypsy_S12B         #101            TGCATACTTT TTAGAGAAAC CAAATAATTT TTTATTGCAT ACCCGTTTTT
                                    ..................................................
 gypsy              #101            TGCATACTTT TTAGAGAAAC CAAATAATTT TTTATTGCAT ACCCGTTTTT


>Input_gypsy        #151            AATAAAATAC ATTGCATACC CTCTTTTAAT AAAAAATATT GCATACTTTG
<gypsy_S1B          #151            AATAAAATAC ATTGCATACC CTCTTTTAAT AAAAAATATT GCATACTTTG
<gypsy_S9B          #151            AATAAAATAC ATTGCATACC CTCTTTTAAT AAAAAATATT GCATACTTTG
<gypsy_S10B         #151            AATAAAATAC ATTGCATACC CTCTTTTAAT AAAAAATATT GCATACTTTG
<gypsy_S11B         #151            AATAAAATAC ATTGCATACC CTCTTTTAAT AAAAAATATT GCATACTTTG
<gypsy_S12B         #151            AATAAAATAC ATTGCATACC CTCTTTTAAT AAAAAATATT GCATACTTTG
```

```
                                    ..................................................
 gypsy              #151            AATAAAATAC ATTGCATACC CTCTTTTAAT AAAAAATATT GCATACTTTG


>Input_gypsy        #201            ACGAAACAAA TTTTCGTTGC ATACCCAATA AAAGATTATT ATATTGCATA
<gypsy_S1B          #201            ACGAAACAAA TTTTCGTTGC ATACCCAATA AAAGATTATT ATATTGCATA
<gypsy_S9B          #201            ACGAAACAAA TTTTCGTTGC ATACCCAATA AAAGATTATT ATATTGCATA
<gypsy_S10B         #201            ACGAAACAAA TTTTCGTTGC ATACCCAATA AAAGATTATT ATATTGCATA
<gypsy_S11B         #201            ACGAAACAAA TTTTCGTTGC ATACCCAATA AAAGATTATT ATATTGCATA
<gypsy_S12B         #201            ACGAAACAAA TTTTCGTTGC ATACCCAATA AAAGATTATT ATATTGCATA
                                    ..................................................
 gypsy              #201            ACGAAACAAA TTTTCGTTGC ATACCCAATA AAAGATTATT ATATTGCATA


>Input_gypsy        #251            CCCGTTTTTA ATAAAATACA TTGCATACCC TCTTTTAATA AAAAATATTG
<gypsy_S1B          #251            CCCGTTTTTA ATAAAATACA TTGCATACCC TCTTTTAATA AAAAATATTG
<gypsy_S9B          #251            CCCGTTTTTA ATAAAATACA TTGCATACCC TCTTTTAATA AAAAATATTG
<gypsy_S10B         #251            CCCGTTTTTA ATAAAATACA TTGCATACCC TCTTTTAATA AAAAATATTG
<gypsy_S11B         #251            CCCGTTTTTA ATAAAATACA TTGCATACCC TCTTTTAATA AAAAATATTG
<gypsy_S12B         #251            CCCGTTTTTA ATAAAATACA TTGCATACCC TCTTTTAATA AAAAATATTG
                                    ..................................................
 gypsy              #251            CCCGTTTTTA ATAAAATACA TTGCATACCC TCTTTTAATA AAAAATATTG


>Input_gypsy        #301            CATACGTTGA CGAAACAAAT TTTCGTTGCA TACCCAATAA AAGATTATTA
<gypsy_S1B          #301            CATACGTTGA CGAAACAAAT TTTCGTTGCA TACCCAATAA AAGATTATTA
<gypsy_S9B          #301            CATACGTTGA CGAAACAAAT TTTCGTTGCA TACCCAATAA AAGATTATTA
<gypsy_S10B         #301            CATACGTTGA CGAAACAAAT TTTCGTTGCA TACCCAATAA AAGATTATTA
<gypsy_S11B         #301            CATACGTTGA CGAAACAAAT TTTCGTTGCA TACCCAATAA AAGATTATTA
<gypsy_S12B         #301            CATACGTTGA CGAAACAAAT TTTCGTTGCA TACCCAATAA AAGATTATTA
                                    ..................................................
 gypsy              #301            CATACGTTGA CGAAACAAAT TTTCGTTGCA TACCCAATAA AAGATTATTA


>Input_gypsy        #351            TATTGCATAC CTTTTCTTGC CATACCATTT AGCCGATCAA TTCTGCAG
<gypsy_S1B          #351            TATTGCATAC CTTTTCTTGC CATACCATTT AGCCGATCAA TTCTGCAG
<gypsy_S9B          #351            TATTGCATAC CTTTTCTTGC CATACCATTT AGCCGATCAA TTCTGCAG
<gypsy_S10B         #351            TATTGCATAC CTTTTCTTGC CATACCATTT AGCCGATCAA TTCTGCAG
<gypsy_S11B         #351            TATTGCATAC CTTTTCTTGC CATACCATTT AGCCGATCAA TTCTGCAG
```

```
<gypsy_S12B         #351                TATTGCATAC CTTTTCTTGC CATACCATTT AGCCGATCAA TTCTGCAG
                                        ..................................................
 gypsy              #351                TATTGCATAC CTTTTCTTGC CATACCATTT AGCCGATCAA TTCTGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with gypsy_ ΔSu(Hw) sequence

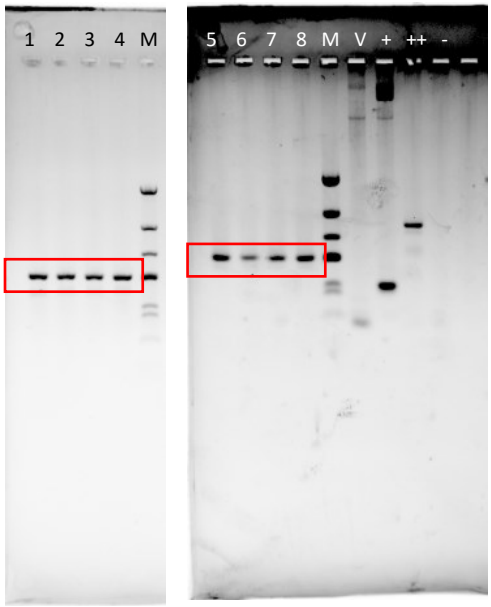


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone gypsy_ ΔSu(Hw) in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate gypsy_ ΔSu(Hw) insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 438bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-8 : Transgenic plant DNA containing gypsy_ ΔSu(Hw) pL1 inserts (438bp)
V : pL1 control vector (no insert, 170bp)
+ : UASrpg sequenced control (746bp)
++ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "gypsy_deltaSu(Hw)" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_gypsy_deltaS.#1          AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTTTTC
<gypsy_deltaSu(Hw)_.#1          AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTTTTC
<gypsy_deltaSu(Hw)_.#1          AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTTTTC
<gypsy_deltaSu(Hw)_.#1          AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTTTTC
<gypsy_deltaSu(Hw)_.#1          AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTTTTC
<gypsy_deltaSu(Hw)_.#1          AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTTTTC
                                ..................................................
 gypsy_deltaSu(Hw)  #1          AAGCTTCACG TAATAAGTGT GCGTTGAATT TATTCGCAAA AACATTTTTC


>Input_gypsy_deltaS.#51         GGCAAAGTAA AATTTTGTCC TTATCAAAAA ATAAGTGCCT TTTTAGAGAA
<gypsy_deltaSu(Hw)_.#51         GGCAAAGTAA AATTTTGTCC TTATCAAAAA ATAAGTGCCT TTTTAGAGAA
<gypsy_deltaSu(Hw)_.#51         GGCAAAGTAA AATTTTGTCC TTATCAAAAA ATAAGTGCCT TTTTAGAGAA
<gypsy_deltaSu(Hw)_.#51         GGCAAAGTAA AATTTTGTCC TTATCAAAAA ATAAGTGCCT TTTTAGAGAA
<gypsy_deltaSu(Hw)_.#51         GGCAAAGTAA AATTTTGTCC TTATCAAAAA ATAAGTGCCT TTTTAGAGAA
<gypsy_deltaSu(Hw)_.#51         GGCAAAGTAA AATTTTGTCC TTATCAAAAA ATAAGTGCCT TTTTAGAGAA
                                ..................................................
 gypsy_deltaSu(Hw)  #51         GGCAAAGTAA AATTTTGTCC TTATCAAAAA ATAAGTGCCT TTTTAGAGAA


>Input_gypsy_deltaS.#101        ACCAAATAAT TTTTTATCCC GTTTTTAATA AAATACATCC CTCTTTTAAT
<gypsy_deltaSu(Hw)_.#101        ACCAAATAAT TTTTTATCCC GTTTTTAATA AAATACATCC CTCTTTTAAT
<gypsy_deltaSu(Hw)_.#101        ACCAAATAAT TTTTTATCCC GTTTTTAATA AAATACATCC CTCTTTTAAT
<gypsy_deltaSu(Hw)_.#101        ACCAAATAAT TTTTTATCCC GTTTTTAATA AAATACATCC CTCTTTTAAT
<gypsy_deltaSu(Hw)_.#101        ACCAAATAAT TTTTTATCCC GTTTTTAATA AAATACATCC CTCTTTTAAT
<gypsy_deltaSu(Hw)_.#101        ACCAAATAAT TTTTTATCCC GTTTTTAATA AAATACATCC CTCTTTTAAT
                                ..................................................
 gypsy_deltaSu(Hw)  #101        ACCAAATAAT TTTTTATCCC GTTTTTAATA AAATACATCC CTCTTTTAAT


>Input_gypsy_deltaS.#151        AAAAAATATC TTTGACGAAA CAAATTTTCG TCCCAATAAA AGATTATTAT
<gypsy_deltaSu(Hw)_.#151        AAAAAATATC TTTGACGAAA CAAATTTTCG TCCCAATAAA AGATTATTAT
<gypsy_deltaSu(Hw)_.#151        AAAAAATATC TTTGACGAAA CAAATTTTCG TCCCAATAAA AGATTATTAT
<gypsy_deltaSu(Hw)_.#151        AAAAAATATC TTTGACGAAA CAAATTTTCG TCCCAATAAA AGATTATTAT
<gypsy_deltaSu(Hw)_.#151        AAAAAATATC TTTGACGAAA CAAATTTTCG TCCCAATAAA AGATTATTAT
<gypsy_deltaSu(Hw)_.#151        AAAAAATATC TTTGACGAAA CAAATTTTCG TCCCAATAAA AGATTATTAT
```

```
                                  ..................................................
 gypsy_deltaSu(Hw)  #151          AAAAAATATC TTTGACGAAA CAAATTTTCG TCCCAATAAA AGATTATTAT


>Input_gypsy_deltaS.#201          ATCCCGTTTT TAATAAAATA CATCCCTCTT TTAATAAAAA ATATCGTTGA
<gypsy_deltaSu(Hw)_.#201          ATCCCGTTTT TAATAAAATA CATCCCTCTT TTAATAAAAA ATATCGTTGA
<gypsy_deltaSu(Hw)_.#201          ATCCCGTTTT TAATAAAATA CATCCCTCTT TTAATAAAAA ATATCGTTGA
<gypsy_deltaSu(Hw)_.#201          ATCCCGTTTT TAATAAAATA CATCCCTCTT TTAATAAAAA ATATCGTTGA
<gypsy_deltaSu(Hw)_.#201          ATCCCGTTTT TAATAAAATA CATCCCTCTT TTAATAAAAA ATATCGTTGA
<gypsy_deltaSu(Hw)_.#201          ATCCCGTTTT TAATAAAATA CATCCCTCTT TTAATAAAAA ATATCGTTGA
                                  ..................................................
 gypsy_deltaSu(Hw)  #201          ATCCCGTTTT TAATAAAATA CATCCCTCTT TTAATAAAAA ATATCGTTGA


>Input_gypsy_deltaS.#251          CGAAACAAAT TTTCGTCCCA ATAAAAGATT ATTATATCCT TTTCTTGCCA
<gypsy_deltaSu(Hw)_.#251          CGAAACAAAT TTTCGTCCCA ATAAAAGATT ATTATATCCT TTTCTTGCCA
<gypsy_deltaSu(Hw)_.#251          CGAAACAAAT TTTCGTCCCA ATAAAAGATT ATTATATCCT TTTCTTGCCA
<gypsy_deltaSu(Hw)_.#251          CGAAACAAAT TTTCGTCCCA ATAAAAGATT ATTATATCCT TTTCTTGCCA
<gypsy_deltaSu(Hw)_.#251          CGAAACAAAT TTTCGTCCCA ATAAAAGATT ATTATATCCT TTTCTTGCCA
<gypsy_deltaSu(Hw)_.#251          CGAAACAAAT TTTCGTCCCA ATAAAAGATT ATTATATCCT TTTCTTGCCA
                                  ..................................................
 gypsy_deltaSu(Hw)  #251          CGAAACAAAT TTTCGTCCCA ATAAAAGATT ATTATATCCT TTTCTTGCCA


>Input_gypsy_deltaS.#301          TACCATTTAG CCGATCAATT CTGCAG
<gypsy_deltaSu(Hw)_.#301          TACCATTTAG CCGATCAATT CTGCAG
<gypsy_deltaSu(Hw)_.#301          TACCATTTAG CCGATCAATT CTGCAG
<gypsy_deltaSu(Hw)_.#301          TACCATTTAG CCGATCAATT CTGCAG
<gypsy_deltaSu(Hw)_.#301          TACCATTTAG CCGATCAATT CTGCAG
<gypsy_deltaSu(Hw)_.#301          TACCATTTAG CCGATCAATT CTGCAG
                                  ..................................................
 gypsy_deltaSu(Hw)  #301          TACCATTTAG CCGATCAATT CTGCAG
```

DNA Analysis of *Arabidopsis thaliana* transformed with UASrpgINV sequence

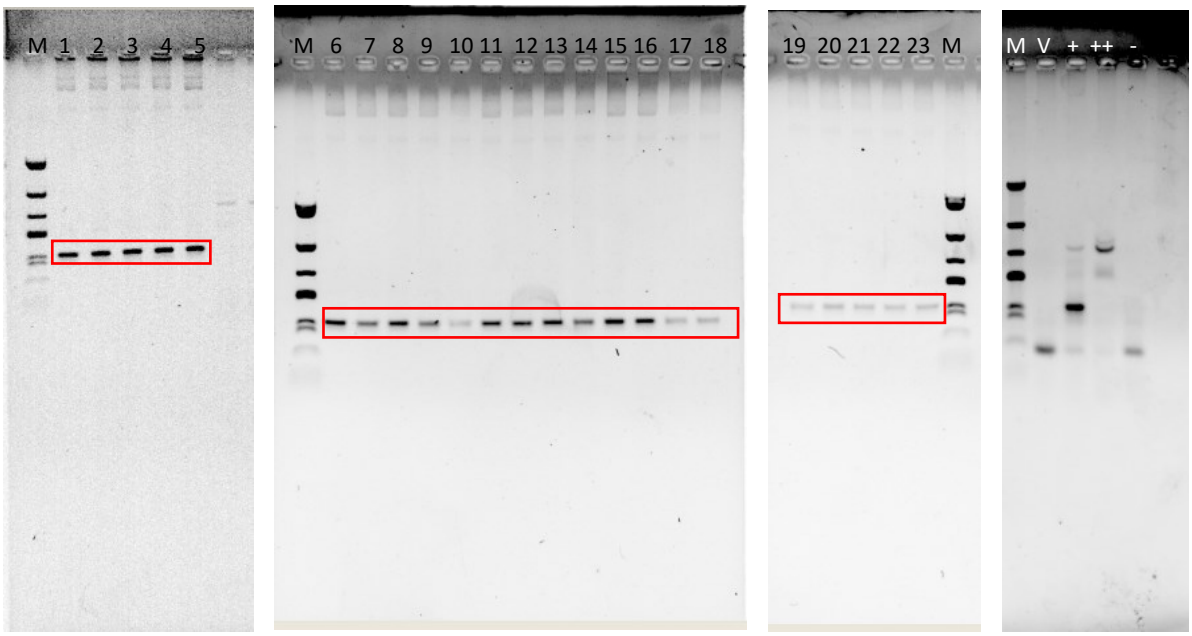


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone UASrpgINV in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate UASrpgINV insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 416bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-23 : Transgenic plant DNA containing UASrpgINV pL1 inserts (416bp)
V : pL1 control vector (no insert, 170bp)
+ : UASrpg sequenced control (416bp)
++ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "UASrpgINV" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_UASrpgINV    #1              AAGCTTTTTC TCTACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
<UASrpgINV_S1B      #1              AAGCTTTTTC TCTACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
<UASrpgINV_S7B      #1              AAGCTTTTTC TCTACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
<UASrpgINV_S8B      #1              AAGCTTTTTC TCTACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
<UASrpgINV_S9       #1              AAGCTTTTTC TCTACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
<UASrpgINV_S15B     #1              AAGCTTTTTC TCTACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
                                    ..................................................
 UAS_INV            #1              AAGCTTTTTC TCTACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT


>Input_UASrpgINV    #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGAGGAG CCGTAATTTT
<UASrpgINV_S1B      #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGAGGAG CCGTAATTTT
<UASrpgINV_S7B      #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGAGGAG CCGTAATTTT
<UASrpgINV_S8B      #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGAGGAG CCGTAATTTT
<UASrpgINV_S9       #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGAGGAG CCGTAATTTT
<UASrpgINV_S15B     #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGAGGAG CCGTAATTTT
                                    ..................................................
 UAS_INV            #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGAGGAG CCGTAATTTT


>Input_UASrpgINV    #101            TCGTTCGCGC CGTGCGGCCA TCAAAATGTA TGGATGCAAA TGATTATACA
<UASrpgINV_S1B      #101            TCGTTCGCGC CGTGCGGCCA TCAAAATGTA TGGATGCAAA TGATTATACA
<UASrpgINV_S7B      #101            TCGTTCGCGC CGTGCGGCCA TCAAAATGTA TGGATGCAAA TGATTATACA
<UASrpgINV_S8B      #101            TCGTTCGCGC CGTGCGGCCA TCAAAATGTA TGGATGCAAA TGATTATACA
<UASrpgINV_S9       #101            TCGTTCGCGC CGTGCGGCCA TCAAAATGTA TGGATGCAAA TGATTATACA
<UASrpgINV_S15B     #101            TCGTTCGCGC CGTGCGGCCA TCAAAATGTA TGGATGCAAA TGATTATACA
                                    ..................................................
 UAS_INV            #101            TCGTTCGCGC CGTGCGGCCA TCAAAATGTA TGGATGCAAA TGATTATACA


>Input_UASrpgINV    #151            TGGGGATGTA TGGGCTAAAT GTACGGGCGA CAGTCACATC ATGCCCCTGA
<UASrpgINV_S1B      #151            TGGGGATGTA TGGGCTAAAT GTACGGGCGA CAGTCACATC ATGCCCCTGA
<UASrpgINV_S7B      #151            TGGGGATGTA TGGGCTTTAT GTACGGGCGA CAGTCTCATC ATGCCCCNGA
<UASrpgINV_S8B      #151            TGGGGATGTA TGGGCTAAAT GTACGGGCGA CAGTCACATC ATGCCCCTGA
<UASrpgINV_S9       #151            TGGGGATGTA TGGGCTAAAT GTACGGGCGA CAGTCACATC ATGCCCCTGA
<UASrpgINV_S15B     #151            TGGGGAAGTA TGGGCTAAAT MTATGGGCGA CAGTCGCATC ACGCCCCCAA
```

```
                                    ..................................................
 UAS_INV            #151            TGGGGATGTA TGGGCTAAAT GTACGGGCGA CAGTCACATC ATGCCCCTGA
                                          *          **   +  *            *      *     **

>Input_UASrpgINV    #201            GCTGCGCACG TCAAGACTGT CAAGGAGGGT ATTCTGGGCC CTGCAG
<UASrpgINV_S1B      #201            GCTGCGCACG TCAAGACTGT CAAGGAGGGT ATTCTGGGCC CTGCAG
<UASrpgINV_S7B      #201            GCTGCACACG TTGAGCCTGS CGGGGAGGGT ATTCTGGGCC CTGCAG
<UASrpgINV_S8B      #201            GCTGCGCACG TCAAGACTGT CAAGGAGGGT ATTCTGGGCC CTGCAG
<UASrpgINV_S9       #201            GCTGCGCACG TCAAGACTGT CAAGGAGGGT ATTCTGGGCC CTGCAG
<UASrpgINV_S15B     #201            GCTGCACACG TTGAGACTGG CGCGGAGGCT ATTCTGGGCC CTGCAG
                                    ..................................................
 UAS_INV            #201            GCTGCGCACG TCAAGACTGT CAAGGAGGGT ATTCTGGGCC CTGCAG
                                         *      **  *   *  **     *
```

DNA Analysis of *Arabidopsis thaliana* transformed with UAS_5'INV sequence

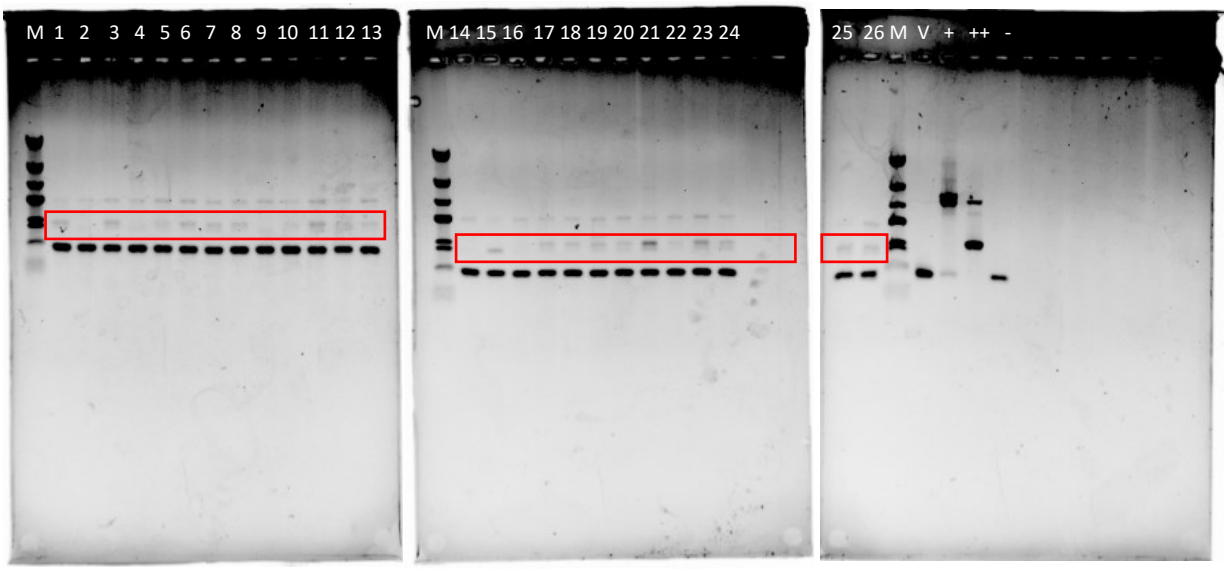


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone UAS_5'INV in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate UAS_5'INV insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 319bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-26 : Transgenic plant DNA containing UAS_5'INV pL1 inserts (319bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
++ : UASrpg sequenced control (416bp)
- : Negative control containing water in replace of DNA

Alignment of "UAS_5′INV" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_UAS_5'INV    #1                AAGCTTTTCG CGCCGTGCGG CCATCAAAAT GTATGGATGC AAATGATTAT
>UAS_5'INV_S11      #1                AAGCTTTTCG CGCCGTGCGG CCATCAAAAT GTATGGATGC AAATGATTAT
>UAS_5'INV_S10      #1                AAGCTTTTCG CGCCGTGCGG CCATCAAAAT GTATGGATGC AAATGATTAT
>UAS_5'INV_S8       #1                AAGCTTTTCG CGCCGTGCGG CCATCAAAAT GTATGGATGC AAATGATTAT
>UAS_5'INV_S5       #1                AAGCTTTTCG CGCCGTGCGG CCATCAAAAT GTATGGATGC AAATGATTAT
<UAS_5'INV_S1       #1                AAGCTTTTCG CGCCGTGCGG CCATCAAAAT GTATGGATGC AAATGATTAT
                                      ..................................................
 UAS_5'INV          #1                AAGCTTTTCG CGCCGTGCGG CCATCAAAAT GTATGGATGC AAATGATTAT


>Input_UAS_5'INV    #51               ACATGGGGAT GTATGGGCTA AATGTACGGG CGACAGTCAC ATCATGCCCC
>UAS_5'INV_S11      #51               ACATGGGGAT GTATGGGCTA AATGTACGGG CGACAGTCAC ATCATGCCCC
>UAS_5'INV_S10      #51               ACATGGGGAT GTATGGGCTA AATGTACGGG CGACAGTCAC ATCATGCCCC
>UAS_5'INV_S8       #51               ACATGGGGAT GTATGGGCTA AATGTACGGG CGACAGTCAC ATCATGCCCC
>UAS_5'INV_S5       #51               ACATGGGGAT GTATGGGCTA AATGTACGGG CGACAGTCAC ATCATGCCCC
<UAS_5'INV_S1       #51               ACATGGGGAT GTATGGGCTA AATGTACGGG CGACAGTCAC ATCATGCCCC
                                      ..................................................
 UAS_5'INV          #51               ACATGGGGAT GTATGGGCTA AATGTACGGG CGACAGTCAC ATCATGCCCC


>Input_UAS_5'INV    #101              TGAGCTGCGC ACGTCAAGAC TGTCAAGGAG GGTATTCTGG GCCCTGCAG
>UAS_5'INV_S11      #101              TGAGCTGCGC ACGTCAAGAC TGTCAAGGAG GGTATTCTGG GCCCTGCAG
>UAS_5'INV_S10      #101              TGAGCTGCGC ACGTCAAGAC TGTCAAGGAG GGTATTCTGG GCCCTGCAG
>UAS_5'INV_S8       #101              TGAGCTGCGC ACGTCAAGAC TGTCAAGGAG GGTATTCTGG GCCCTGCAG
>UAS_5'INV_S5       #101              TGAGCTGCGC ACGTCAAGAC TGTCAAGGAG GGTATTCTGG GCCCTGCAG
<UAS_5'INV_S1       #101              TGAGCTGCGC ATGTTGGGCC TGGCGCGCAG GCTATTCAGG GCCCTGCAG
                                      ..................................................
 UAS_5'INV          #101              TGAGCTGCGC ACGTCAAGAC TGTCAAGGAG GGTATTCTGG GCCCTGCAG
                                         *  *** *    * ** *    *     *
```

DNA Analysis of *Arabidopsis thaliana* transformed with UAS_3'INV sequence

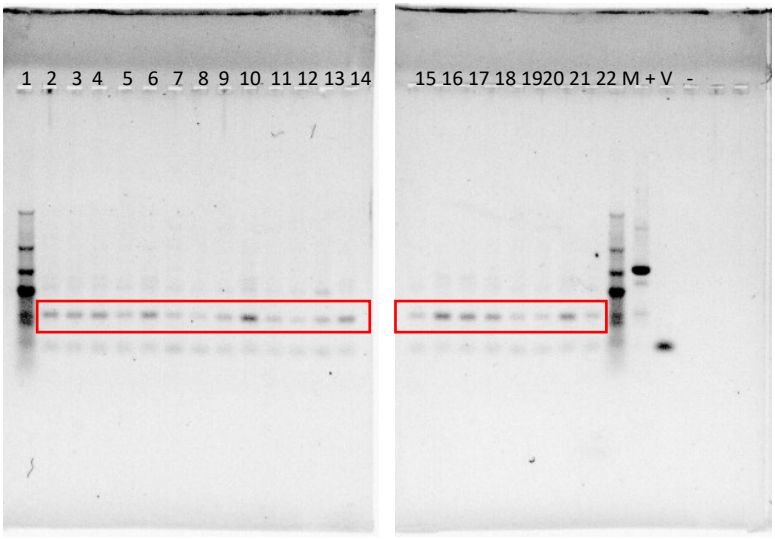


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone UAS_3'INV in pL1 with pL1F and NapinSeqRev primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate UAS_3'INV insulator are shown. Amplification of DNA used pL1F and NapinSeqR primers to produce a band with an expected size of 279bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-26 : Transgenic plant DNA containing UAS_3'INV pL1 inserts (279bp)
V : pL1 control vector (no insert, 170bp)
+ : BEAD1c sequenced control (718bp)
- : Negative control containing water in replace of DNA

Alignment of "UAS_3'INV" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_UAS_3'INV    #1              AAGCTTTTTC TCTACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
>UAS_3'INV_S9B      #1              AAGCTTTTGC TATACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
>UAS_3'INV_S10      #1              AACCTTTTGC TATACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
>UAS_3'INV_S16      #1              AACCTTTTGC TATACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
>UAS_3'INV_S20      #1              AACCTTTTTC TATACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
                                    ..................................................
 UAS_3'INV          #1              AACCTTTTGC TATACAGGGG CGCGCCGTGG GGAGAATTCA ACGCGTCTGT
                                      *     *   *

>Input_UAS_3'INV    #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGAGGAG CCGTAATTTT
>UAS_3'INV_S9B      #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGACGGG CCCTTATTTT
>UAS_3'INV_S10      #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGACGGG CCGAAATTTT
>UAS_3'INV_S16      #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGACGGG CCCCTATTTT
>UAS_3'INV_S20      #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGACAGG CCCCAATTTT
                                    ..................................................
 UAS_3'INV          #51             GAGGGGAGCG TTTCCCTGCT CGCAGGTCTG CAGCGACGGG CCCYAATTTT
                                                                      ***    ***

>Input_UAS_3'INV    #101            TCGCTGCAG
>UAS_3'INV_S9B      #101            TCGCTGCAGC GG
>UAS_3'INV_S10      #101            TCACTGCAG
>UAS_3'INV_S16      #101            TCTCTGCAGA G
>UAS_3'INV_S20      #101            TCACTGCAGA GG
                                    ..................................................
 UAS_3'INV          #101            TCRCTGCAGA GG
                                      *      *
```

DNA Analysis of *Arabidopsis thaliana* transformed with BEAD1c_INV sequence

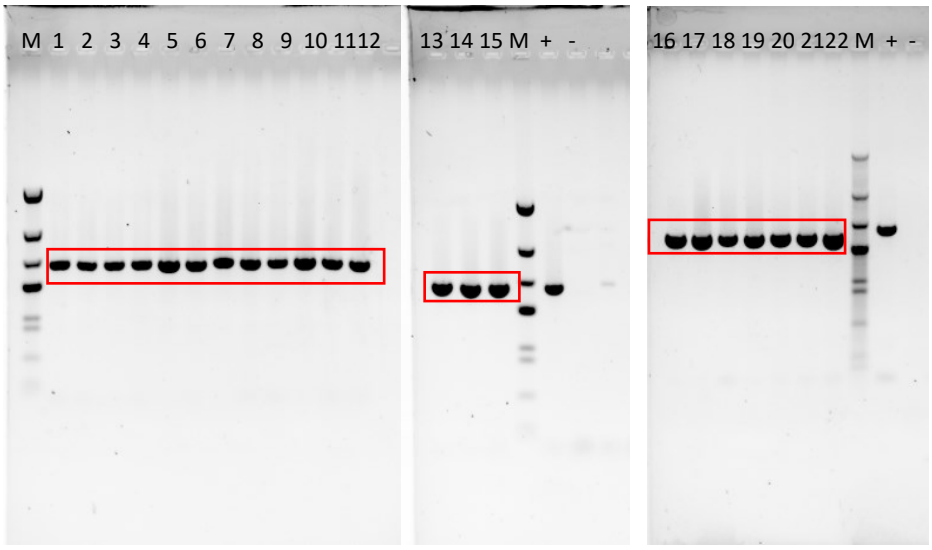


Figure A3. : Example of agarose gel electrophoresis showing PCR amplification of clone BEAD1c_INV in pL1 with BEAD specific primers. Each gel contained 14 wells, therefore samples and controls were continued on separate gels, all with the same PCR components and conditions. Transgenic *Arabidopsis thaliana* DNA samples containing candidate BEAD1c_INV insulator are shown. Amplification of DNA used BEAD specific primers to produce a band with an expected size of 548bp.

M : DNA ladder with sizes (in bp) 1353, 815, 587, (458, 449, 434), 298, 267, 174, 102, 80
1-22 : Transgenic plant DNA containing BEAD1c_INV pL1 inserts (548bp)
+ : BEAD1c sequenced control (548bp)
- : Negative control containing water in replace of DNA

Alignment of "BEAD1cINV" samples and consensus sequence output from Sequencher 4.10.1

```
>Input_BEAD1cINV    #1              AAGCTTCTGT GCTTGGGTAT GGTCAGACGT GGGTGAGGTT GGGCAGTTAG
<BEAD1cINV_S1       #1              AAGCTTCTGT GCTTGGGTAT GGTCAGACGT GGGTGAGGTT GGGCAGTTAG
<BEAD1cINV_S2       #1              AAGCTTCTGT GCTTGGGTAT GGTCAGACGT GGGTGAGGTT GGGCAGTTAG
<BEAD1cINV_S3       #1              AAGCTTCTGT GCTTGGGTAT GGTCAGACGT GGGTGAGGTT GGGCAGTTAG
<BEAD1cINV_S9       #1              AAGCTTCTGT GCTTGGGTAT GGTCAGACGT GGGTGAGGTT GGGCAGTTAG
<BEAD1cINV_S10      #1              AAGCTTCTGT GCTTGGGTAT GGTCAGACGT GGGTGAGGTT GGGCAGTTAG
                                    ..................................................
 BEAD1cINV          #1              AAGCTTCTGT GCTTGGGTAT GGTCAGACGT GGGTGAGGTT GGGCAGTTAG
 Consensus

>Input_BEAD1cINV    #51             AAGAGGAGGA GCAATTTAAC TTGAAGATGG TTATGGTCTA GATCAGGAAG
<BEAD1cINV_S1       #51             AAGAGGAGGA GCAATTTAAC TTGAAGATGG TTATGGTCTA GATCAGGAAG
<BEAD1cINV_S2       #51             AAGAGGAGGA GCAATTTAAC TTGAAGATGG TTATGGTCTA GATCAGGAAG
<BEAD1cINV_S3       #51             AAGAGGAGGA GCAATTTAAC TTGAAGATGG TTATGGTCTA GATCAGGAAG
<BEAD1cINV_S9       #51             AAGAGGAGGA GCAATTTAAC TTGAAGATGG TTATGGTCTA GATCAGGAAG
<BEAD1cINV_S10      #51             AAGAGGAGGA GCAATTTAAC TTGAAGATGG TTATGGTCTA GATCAGGAAG
                                    ..................................................
 BEAD1cINV          #51             AAGAGGAGGA GCAATTTAAC TTGAAGATGG TTATGGTCTA GATCAGGAAG


>Input_BEAD1cINV    #101            GGAACAGTCA GCTCTCATAT TAATATCAAG CCCAGACAAC ATCTTTAGAG
<BEAD1cINV_S1       #101            GGAACAGTCA GCTCTCATAT TAATATCAAG CCCAGACAAC ATCTTTAGAG
<BEAD1cINV_S2       #101            GGAACAGTCA GCTCTCATAT TAATATCAAG CCCAGACAAC ATCTTTAGAG
<BEAD1cINV_S3       #101            GGAACAGTCA GCTCTCATAT TAATATCAAG CCCAGACAAC ATCTTTAGAG
<BEAD1cINV_S9       #101            GGAACAGTCA GCTCTCATAT TAATATCAAG CCCAGACAAC ATCTTTAGAG
<BEAD1cINV_S10      #101            GGAACAGTCA GCTCTCATAT TAATATCAAG CCCAGACAAC ATCTTTAGAG
                                    ..................................................
 BEAD1cINV          #101            GGAACAGTCA GCTCTCATAT TAATATCAAG CCCAGACAAC ATCTTTAGAG


>Input_BEAD1cINV    #151            CAGGGACAGA CACTAAGAGA CTGTGGGTGG GGAGTGAGCA TAATCCTTGG
<BEAD1cINV_S1       #151            CAGGGACAGA CACTAAGAGA CTGTGGGTGG GGAGTGAGCA TAATCCTTGG
<BEAD1cINV_S2       #151            CAGGGACAGA CACTAAGAGA CTGTGGGTGG GGAGTGAGCA TAATCCTTGG
<BEAD1cINV_S3       #151            CAGGGACAGA CACTAAGAGA CTGTGGGTGG GGAGTGAGCA TAATCCTTGG
<BEAD1cINV_S9       #151            CAGGGACAGA CACTAAGAGA CTGTGGGTGG GGAGTGAGCA TAATCCTTGG
<BEAD1cINV_S10      #151            CAGGGACAGA CACTAAGAGA CTGTGGGTGG GGAGTGAGCA TAATCCTTGG
```

```
                                    ..................................................
 BEAD1cINV          #151            CAGGGACAGA CACTAAGAGA CTGTGGGTGG GGAGTGAGCA TAATCCTTGG


>Input_BEAD1cINV    #201            ACTTGGGAGA ACCGGAAGCT CTGAAATGGA CTTCCGCCAG CAGGGAGCTG
<BEAD1cINV_S1       #201            ACTTGGGAGA ACCGGAAGCT CTGAAATGGA CTTCCGCCAG CAGGGAGCTG
<BEAD1cINV_S2       #201            ACTTGGGAGA ACCGGAAGCT CTGAAATGGA CTTCCGCCAG CAGGGAGCTG
<BEAD1cINV_S3       #201            ACTTGGGAGA ACCGGAAGCT CTGAAATGGA CTTCCGCCAG CAGGGAGCTG
<BEAD1cINV_S9       #201            ACTTGGGAGA ACCGGAAGCT CTGAAATGGA CTTCCGCCAG CAGGGAGCTG
<BEAD1cINV_S10      #201            ACTTGGGAGA ACCGGAAGCT CTGAAATGGA CTTCCGCCAG CAGGGAGCTG
                                    ..................................................
 BEAD1cINV          #201            ACTTGGGAGA ACCGGAAGCT CTGAAATGGA CTTCCGCCAG CAGGGAGCTG


>Input_BEAD1cINV    #251            GGTCTCGTGG ACTGGACCCC TGCCGGCAGG CGGCAGTGCA GGCCTGGGCA
<BEAD1cINV_S1       #251            GGTCTCGTGG ACTGGACCCC TGCCGGCAGG CGGCAGTGCA GGCCTGGGCA
<BEAD1cINV_S2       #251            GGTCTCGTGG ACTGGACCCC TGCCGGCAGG CGGCAGTGCA GGCCTGGGCA
<BEAD1cINV_S3       #251            GGTCTCGTGG ACTGGACCCC TGCCGGCAGG CGGCAGTGCA GGCCTGGGCA
<BEAD1cINV_S9       #251            GGTCTCGTGG ACTGGACCCC TGCCGGCAGG CGGCAGTGCA GGCCTGGGCA
<BEAD1cINV_S10      #251            GGTCTCGTGG ACTGGACCCC TGCCGGCAGG CGGCAGTGCA GGCCTGGGCA
                                    ..................................................
 BEAD1cINV          #251            GGTCTCGTGG ACTGGACCCC TGCCGGCAGG CGGCAGTGCA GGCCTGGGCA


>Input_BEAD1cINV    #301            GCCAATGCAC CGGCAAGGCA GCCCCTGGGC CCTCAACATC CCTCTTCCAC
<BEAD1cINV_S1       #301            GCCAATGCAC CGGCAAGGCA GCCCCTGGGC CCTCAACATC CCTCTTCCAC
<BEAD1cINV_S2       #301            GCCAATGCAC CGGCAAGGCA GCCCCTGGGC CCTCAACATC CCTCTTCCAC
<BEAD1cINV_S3       #301            GCCAATGCAC CGGCAAGGCA GCCCCTGGGC CCTCAACATC CCTCTTCCAC
<BEAD1cINV_S9       #301            GCCAATGCAC CGGCAAGGCA GCCCCTGGGC CCTCAACATC CCTCTTCCAC
<BEAD1cINV_S10      #301            GCCAATGCAC CGGCAAGGCA GCCCCTGGGC CCTCAACATC CCTCTTCCAC
                                    ..................................................
 BEAD1cINV          #301            GCCAATGCAC CGGCAAGGCA GCCCCTGGGC CCTCAACATC CCTCTTCCAC


>Input_BEAD1cINV    #351            ATCCAGGGGC CCTGAAGATG CTCATGGTGG GACTCTGGCC TGCCTGGTGC
<BEAD1cINV_S1       #351            ATCCAGGGGC CCTGAAGATG CTCATGGTGG GACTCTGGCC TGCCTGGTGC
<BEAD1cINV_S2       #351            ATCCAGGGGC CCTGAAGATG CTCATGGTGG GACTCTGGCC TGCCTGGTGC
<BEAD1cINV_S3       #351            ATCCAGGGGC CCTGAAGATG CTCATGGTGG GACTCTGGCC TGCCTGGTGC
<BEAD1cINV_S9       #351            ATCCAGGGGC CCTGAAGATG CTCATGGTGG GACTCTGGCC TGCCTGGTGC
```

```
<BEAD1cINV_S10      #351              ATCCAGGGGC CCTGAAGATG CTCATGGTGG GACTCTGGCC TGCCTGGTGC
                                      ..................................................
 BEAD1cINV          #351              ATCCAGGGGC CCTGAAGATG CTCATGGTGG GACTCTGGCC TGCCTGGTGC


>Input_BEAD1cINV    #401              TTACTTTTGG TATTCTGCAT GCCTAGCATT GCCAGATAAA ACACAGGATG
<BEAD1cINV_S1       #401              TTACTTTTGG TATTCTGCAT GCCTAGCATT GCCAGATAAA ACACAGGATG
<BEAD1cINV_S2       #401              TTACTTTTGG TATTCTGCAT GCCTAGCATT GCCAGATAAA ACACAGGATG
<BEAD1cINV_S3       #401              TTACTTTTGG TATTCTGCAT GCCTAGCATT GCCAGATAAA ACACAGGATG
<BEAD1cINV_S9       #401              TTACTTTTGG TATTCTGCAT GCCTAGCATT GCCAGATAAA ACACAGGATG
<BEAD1cINV_S10      #401              TTACTTTTGG TATTCTGCAT GCCTAGCATT GCCAGATAAA ACACAGGATG
                                      ..................................................
 BEAD1cINV          #401              TTACTTTTGG TATTCTGCAT GCCTAGCATT GCCAGATAAA ACACAGGATG


>Input_BEAD1cINV    #451              CCCAGTGGAA TTTGAATTTC AGATAAACAA TGAATACTTT TTTAGTATAA
<BEAD1cINV_S1       #451              CCCAGTGGAA TTTGAATTTC AGATAAACAA TGAATACTTT TTTAGTATAA
<BEAD1cINV_S2       #451              CCCAGTGGAA TTTGAATTTC AGATAAACAA TGAATACTTT TTTAGTATAA
<BEAD1cINV_S3       #451              CCCAGTGGAA TTTGAATTTC AGATAAACAA TGAATACTTT TTTAGTATAA
<BEAD1cINV_S9       #451              CCCAGTGGAA TTTGAATTTC AGATAAACAA TGAATACTTT TTTAGTATAA
<BEAD1cINV_S10      #451              CCCAGTGGAA TTTGAATTTC AGATAAACAA TGAATACTTT TTTAGTATAA
                                      ..................................................
 BEAD1cINV          #451              CCCAGTGGAA TTTGAATTTC AGATAAACAA TGAATACTTT TTTAGTATAA


>Input_BEAD1cINV    #501              ATGTGTTCCA AATATGGCAT GTCCCAGCTA CCCGTATTAC TGCTGCAG
<BEAD1cINV_S1       #501              ATGTGTTCCA AATATGGCAT GTCCCAGCTA CCCGTATTAC TGCTGCAG
<BEAD1cINV_S2       #501              ATGTGTTCCA AATATGGCAT GTCCCAGCTA CCCGTATTAC TGCTGCAG
<BEAD1cINV_S3       #501              ATGTGTTCCA AATATGGCAT GTCCCAGCTA CCCGTATTAC TGCTGCAG
<BEAD1cINV_S9       #501              ATGTGTTCCA AATATGGCAT GTCCCAGCTA CCCGTATTAC TGCTGCAG
<BEAD1cINV_S10      #501              ATGTGTTCCA AATATGGCAT GTCCCAGCTA CCCGTATTAC TGCTGCAG
                                      ..................................................
 BEAD1cINV          #501              ATGTGTTCCA AATATGGCAT GTCCCAGCTA CCCGTATTAC TGCTGCAG
```

Table A3.1 GUS staining in different tissues of individual *Arabidopsis thaliana* transgenic lines. The intensity was assessed visually using the following scale: 3 (strong), 2 (medium), 1 (week), and 0 (none), as shown in Figure 2.2A. Data represents scoring of tissues in pL1 vector for all sequences tested.

Construct	Size (bp)	Sample #	Flower	Leaf	Silique 1	Seed 2	Silique 2	Seed 2
pL1 control: ter35S-hptII-P35S←→Pnapin-GUS-ternos		1	3	4	4	0	3	1
		2	4	3	4	0	4	2
		3	0	1	0	1	0	0
		4	0	1	0	0	0	1
		5	3	4	3	2	4	1
		6	1	3	1	1	1	2
		7	0	0	0	0	0	0
		8	2	4	4	1	3	1
		9	0	1	0	2	0	1
		10	3	3	1	2	1	1
		11	4	4	4	1	4	1
		12	4	4	3	1	3	2
		13	3	4	4	1	3	0
		14	2	2	1	1	1	1
		15	3	4	4	1	4	1
		16	3	4	3	1	3	1
		17	4	4	4	1	2	1
		18	0	1	1	1	0	1
		19	3	3	2	1	2	2
		20	3	3	4	2	3	1
		21	4	4	4	2	3	1
		22	2	4	3	1	2	1
		23	1	2	2	2	2	3
		24	4	4	4	3	3	1
		25	0	1	0	0	0	1
		26	1	1	1	0	0	2
		27	0	0	0	0	0	2
		28	2	3	2	2	1	2
		29	2	3	2	2	4	1
		30	3	3	4	1	3	1
		31	0	3	1	0	0	2
		32	0	3	4	0	0	1
		33	0	0	0	1	0	0
		34	4	4	4	4	4	4
		35	2	4	1	0	1	0
		36	1	2	0	0	0	0
		37	4	4	4	0	4	0
		38	3	3	4	0	3	0
		39	3	3	4	0	2	0
		40	4	4	4	1	2	0
		41	1	2	4	1	1	0
		42	2	4	4	1	3	0
		43	1	2	1	0	1	0
		44	1	3	1	0	2	0

		45	1	3	1	0	1	0
		46	0	0	0	0	0	0
		47	0	1	0	2	0	0
		48	4	4	4	0	2	0
		49	3	3	2	0	1	0
		50	2	3	4	1	4	1
		51	1	1	0	0	0	2
		52	4	4	4	1	4	1
		53	1	4	4	1	3	2
		Total	41	49	42	41		
		Specific	2					
		Nonspecific	49					
		None	2					
UAS_5'end (pL1: pCAMBIA 1391 napin-GUS)	149	1	0	0	0	3	0	2
		2	0	0	0	3	1	2
		3	0	0	0	1	0	1
		4	0	1	1	1	0	2
		5	0	1	0	0	0	0
		6	0	0	0	2	0	0
		7	0	0	0	1	0	0
		8	0	0	0	0	0	0
		9	0	0	0	2	0	1
		10	0	0	0	1	0	1
		11	0	0	0	0	0	1
		12	0	1	0	2	0	2
		13	0	0	0	1	0	1
		14	0	0	0	2	0	1
		15	0	0	0	1	0	2
		16	0	0	0	0	0	0
		17	0	0	0	1	0	1
		18	0	0	0	2	0	1
		19	0	0	0	1	0	2
		20	0	0	0	1	0	2
		21	0	0	0	2	0	1
		22	0	0	0	1	0	1
		23	0	1	0	1	0	1
		24	0	1	0	1	0	1
		25	0	0	1	1	0	1
		26	0	0	0	1	0	1
		27	0	0	0	1	0	2
		28	0	0	0	2	0	1
		29	0	1	1	2	0	1
		30	0	0	1	2	0	1
		31	0	1	4	4	4	1
		32	0	0	0	2	0	2
		33	0	1	0	2	0	1
		34	0	3	0	1	0	2
		35	0	0	1	0	1	1
		36	0	0	0	3	0	2
		37	0	0	0	2	1	1
		Total (37)	0	9	8	34		

		Specific	21					
		Nonspecific	14					
		None	2					
UAS_mR1pL1 (pL1: pCAMBIA 1391 napin-GUS)	149	1	1	2	2	1	3	1
		2	0	0	0	1	0	2
		3	0	0	0	0	0	1
		4	0	0	0	0	0	0
		5	2	3	1	2	3	2
		6	0	0	0	1	0	1
		7	1	3	3	3	3	3
		8	0	4	1	3	1	3
		9	0	0	0	1	0	2
		10	0	4	1	1	2	2
		11	1	4	0	2	2	2
		12	0	0	0	2	0	2
		13	0	4	0	0	1	0
		14	3	3	3	0	4	1
		15	0	2	0	0	1	0
		16	0	0	0	0	0	0
		17	0	0	1	0	1	0
		18	0	4	1	1	2	2
		Total	5	10	11	13		
		Specific	5					
		Nonspecific	11					
		None	2					
UAS_mR2pL1 (pL1: pCAMBIA 1391 napin-GUS)	149	1	0	2	0	1	0	1
		2	0	1	0	1	0	1
		3	2	3	1	1	1	1
		4	1	2	1	1	1	0
		5	0	3	0	0	0	2
		6	0	0	0	1	0	1
		7	1	3	0	0	1	1
		8	1	3	1	2	1	0
		9	2	3	1	1	1	0
		10	0	2	3	0	2	0
		11	0	1	0	2	2	2
		12	1	4	1	1	1	0
		13	1	4	0	1	0	2
		14	0	1	0	0	2	1
		15	3	3	1	1	3	2
		16	1	4	3	2	3	0
		17	1	4	1	1	2	0
		18	1	3	0	1	0	2
		19	1	3	1	0	2	2
		20	0	0	0	1	0	1
		21	0	0	1	2	0	3
		Total	12	18	14	20		
		Specific	2					
		Nonspecific	19					

		None	0					
UAS_5' INV (pL1: pCAMBIA 1391 napin-GUS)	149	1	0	0	0	0	0	0
		2	1	1	1	1	3	1
		3	1	1	0	1	1	1
		4	1	0	1	1	3	2
		5	2	2	2	1	2	2
		6	1	1	1	1	2	1
		7	1	1	2	1	2	1
		8	1	1	1	0	0	1
		9	1	1	1	0	2	2
		10	3	2	3	3	4	0
		11	2	2	3	1	2	0
		12	0	1	0	1	0	1
		13	0	0	0	1	0	2
		14	2	1	2	2	2	2
		15	1	1	2	2	3	2
		16	1	1	1	1	0	0
		17	3	3	4	4	4	4
		18	0	1	1	1	1	1
		19	1	1	3	1	1	2
		20	0	3	3	1	3	0
		21	0	1	0	1	1	0
		22	1	2	3	2	3	1
		23	0	0	0	1	0	1
		24	0	0	2	1	0	2
		25	0	1	1	1	2	2
		26	0	0	1	2	2	2
		Total	16	20	22	25		
		Specific	2					
		Nonspecific	23					
		None	1					
UAS_3'endpL1 (pL1: pCAMBIA 1391 napin-GUS)	109	1	0	0	0	1	0	0
		2	0	0	0	0	0	0
		3	0	0	0	0	0	0
		4	0	0	0	0	0	0
		5	0	0	0	1	0	0
		6	0	0	0	0	0	0
		7	0	0	0	0	0	1
		8	0	0	0	0	0	0
		9	0	0	0	0	0	0
		10	0	0	0	1	0	0
		11	0	0	0	1	0	0
		12	0	0	0	1	0	1
		13	0	0	0	0	0	1
		14	0	0	0	1	0	1
		15	0	0	0	0	0	1
		16	0	0	0	1	0	0
		17	0	0	0	0	0	1
		18	0	0	0	1	0	1
		19	0	0	0	0	0	1
		20	0	0	0	1	0	1

		21	0	0	0	0	0	1
		22	0	0	0	0	0	1
		23	0	0	0	1	0	0
		1	0	0	0	3	0	3
		2	0	0	0	2	0	2
		3	0	0	0	0	0	3
		4	0	0	0	0	0	3
		5	0	0	0	0	0	0
		6	0	0	0	0	0	1
		7	0	0	0	1	0	2
		8	0	0	0	2	0	3
		9	0	0	0	2	0	2
		10	0	0	0	1	0	2
		11	0	0	0	2	0	1
		12	0	0	0	1	0	1
		13	0	0	0	0	0	2
		14	0	0	0	0	0	1
		15	1	4	2	2	2	2
		16	0	0	0	0	0	0
		17	0	0	0	3	0	3
		18	0	0	0	2	0	2
		Total (41)	1	1	1	33		
		Specific	32					
		Nonspecific	1					
		None	8					
UAS_3'ΔSu(Hw)pL1 (pL1: pCAMBIA 1391 napin-GUS)	109	1	1	3	3	0	3	0
		2	0	0	0	2	0	1
		3	0	0	0	1	0	1
		4	0	0	0	2	0	1
		5	0	0	1	2	1	2
		6	0	0	0	1	0	2
		7	0	0	0	1	0	1
		8	0	0	0	1	0	1
		9	0	0	1	2	0	1
		10	0	0	1	0	1	1
		11	0	0	1	0	0	1
		12	0	0	0	1	1	1
		13	0	0	0	0	0	0
		14	0	0	0	0	0	1
		15	0	0	1	1	0	1
		16	0	0	1	1	1	1
		17	0	0	0	1	0	1
		18	0	1	1	2	1	2
		19	0	0	0	2	1	2
		20	0	0	1	1	0	1
		TOTAL	1	2	11	18		
		Specific	8					
		Nonspecific	11					
		None	1					
UAS_3'ΔCTC pL1(pL1: pCAMBIA 1391 napin-GUS)	104	1	0	0	0	0	0	0

		2	0	0	0	0	0	3
		3	0	0	0	1	0	3
		4	0	0	0	2	0	2
		5	0	0	0	2	0	3
		6	0	0	0	1	0	1
		7	0	0	0	2	0	2
		8	0	0	2	1	0	2
		9	0	0	0	3	0	3
		10	0	0	1	2	0	2
		11	0	0	0	0	0	0
		12	0	0	0	1	0	1
		13	0	0	1	1	1	1
		14	0	0	0	0	0	1
		15	0	0	0	0	0	0
		16	0	0	0	1	0	1
		17	0	3	0	0	0	0
		18	0	0	2	2	2	2
		19	0	0	0	2	0	3
		20	0	0	0	1	0	1
		Total	0	1	4	16		
		Specific	14					
		Nonspecific	3					
		None	3					
UAS_3'INVpL1 (pL1: pCAMBIA 1391 napin-GUS)	109	1	0	0	0	2	0	3
		2	0	0	0	0	0	2
		3	0	0	0	0	0	1
		4	0	0	0	0	0	1
		5	0	0	0	0	0	1
		6	0	0	3	0	3	0
		7	1	3	3	0	3	0
		8	0	0	0	2	0	1
		9	0	1	0	1	0	1
		10	0	1	0	0	0	0
		11	0	0	0	0	0	1
		12	0	0	0	0	0	0
		13	0	0	0	0	0	1
		14	0	0	0	0	0	1
		15	0	0	0	1	0	2
		16	0	0	0	1	0	2
		17	0	0	0	1	0	2
		18	0	0	0	1	0	1
		19	0	0	0	0	0	1
		20	0	1	0	0	0	2
		21	0	2	0	0	0	2
		Total	1	5	2	17		
		Specific	14					
		Nonspecific	6					
		None	1					
UASrpg_INVpL1 (pL1: pCAMBIA 1391 napin-GUS)	246	1	0	0	0	2	0	2
		2	1	3	3	0	3	0

		3	0	0	0	0	0	1
		4	0	0	0	0	0	1
		5	0	0	2	2	2	2
		6	0	0	1	2	1	2
		7	0	0	1	2	1	1
		8	0	0	0	1	0	2
		9	0	0	0	2	0	2
		10	0	0	0	3	0	3
		11	0	0	0	2	0	3
		12	0	0	0	0	0	0
		13	0	0	1	1	1	2
		14	0	0	2	1	3	0
		15	0	0	2	2	1	2
		16	0	0	0	2	0	2
		17	0	0	0	2	0	2
		18	0	1	2	2	2	2
		19	0	0	0	0	0	0
		20	0	0	0	0	1	1
		Total	1	2	10	17		
		Specific	9					
		Nonspecific	9					
		None	2					
BEAD1c_5'end (pL1: pCAMBIA 1391 napin-GUS)	265	1	1	0	0	2	1	2
		2	1	1	1	1	1	2
		3	2	2	1	2	1	2
		4	1	1	0	1	0	1
		5	0	0	0	1	0	1
		6	1	1	1	2	1	2
		7	0	0	1	1	0	1
		8	0	1	0	1	0	1
		9	1	1	1	0	1	1
		10	0	0	0	0	0	1
		11	1	1	1	1	1	1
		12	1	1	1	1	0	1
		13	0	0	0	1	0	3
		14	0	0	0	0	0	2
		15	0	0	0	1	0	1
		16	0	0	0	2	0	2
		17	1	2	1	1	0	1
		18	2	3	2	2	3	2
		19	0	0	0	1	1	1
		20	0	0	0	0	0	1
		21	0	0	0	1	0	1
		22	1	1	0	1	1	3
		23	0	1	0	1	1	3
		24	1	2	0	1	0	1
		25	2	4	1	1	1	1
		26	0	0	0	0	0	1
		27	1	1	0	1	1	1
		28	2	1	3	2	0	3
		29	0	0	0	1	1	1
		30	0	0	1	1	0	1

		31	0	0	0	1	1	1
		32	1	0	2	1	0	1
		Total (32)	16	16	19	32		
		Specific	9	28				
		Nonspecific	23	66				
		None	0					
BEAD1c_3'end (pL1: pCAMBIA 1391 napin-GUS)	301	1	0	0	0	0	0	0
		2	0	0	0	2	0	3
		3	0	0	0	2	0	2
		4	0	0	0	0	0	2
		5	0	0	0	0	1	3
		6	0	0	0	1	0	2
		7	0	0	0	0	0	2
		8	0	0	0	1	1	2
		9	0	0	0	0	0	0
		10	0	0	0	1	0	1
		11	0	0	1	0	0	0
		12	0	0	0	2	0	2
		13	0	0	1	1	1	1
		14	0	0	0	3	0	3
		15	0	0	0	0	0	2
		16	0	0	0	0	0	1
		17	0	0	0	0	0	1
		18	0	0	0	1	0	3
		19	0	1	0	2	0	2
		20	0	0	0	2	0	2
		21	0	0	0	2	0	2
		22	0	0	0	2	0	2
		23	0	0	0	0	0	1
		24	0	0	0	1	0	2
		25	0	0	1	2	0	1
		26	0	0	2	1	0	2
		27	0	0	0	1	0	2
		28	0	0	0	1	0	1
		29	0	0	0	1	3	2
		30	0	0	0	1	0	2
		31	0	0	1	1	0	2
		Total	0	1	8	29		
		Specific	20	69				
		Nonspecific	9	31				
		None	2					
ΔBEADA (pL1: pCAMBIA 1391 napin-GUS)	509	1	0	0	1	1	0	1
		2	0	0	0	1	0	1
		3	0	0	0	1	0	1
		4	0	0	0	1	0	1
		5	0	0	1	1	1	1

		6	0	0	1	1	2	1
		7	0	0	0	2	0	0
		8	0	0	0	2	0	2
		9	0	0	1	1	0	1
		10	0	0	0	1	0	1
		11	0	0	0	1	0	1
		12	0	0	0	2	0	2
		13	0	0	0	1	1	1
		14	2	3	2	2	2	2
		15	1	1	2	0	2	1
		16	0	0	0	1	0	1
		17	0	0	0	0	0	0
		18	0	0	0	0	0	1
		19	0	0	0	1	0	0
		20	0	0	0	1	0	1
		Total	2	2	7	19		
		Specific	12					
		Non-Specific	7					
		None	1					
BEAD1c_INV (pL1: pCAMBIA 1391 napin-GUS)	548	1	0	0	0	0	0	1
		2	0	0	0	0	0	1
		3	0	0	0	0	0	1
		4	0	0	0	0	0	2
		5	0	0	0	0	0	2
		6	0	0	0	0	0	1
		7	0	0	0	0	0	1
		8	0	0	0	0	0	1
		9	0	0	0	0	0	1
		10	0	0	0	1	0	2
		11	0	3	3	0	3	0
		12	0	0	0	0	0	1
		13	0	0	0	0	0	1
		14	0	0	0	0	0	1
		15	0	0	0	1	0	1
		16	0	0	0	0	0	2
		17	0	0	0	0	0	0
		18	0	0	1	2	2	2
		19	0	0	0	2	0	2
		20	0	0	0	2	1	2
		Total	0	1	3	18		
		Specific	16					
		Nonspecific	3					
		None	1					
Gypsy (pL1: pCAMBIA 1391 napin-GUS)	340	1	0	2	1	1	1	1
		2	1	3	2	1	2	0
		3	0	0	0	1	1	1
		4	1	3	3	2	3	1
		5	1	3	3	0	2	0
		6	0	3	3	0	2	1
		7	0	1	1	0	1	0
		8	1	3	3	0	3	1

		9	2	3	3	0	3	0
		10	0	0	2	0	1	0
		11	0	2	0	2	0	2
		12	1	2	2	0	2	0
		13	0	2	3	0	3	0
		14	0	0	0	1	1	0
		15	2	3	3	0	2	0
		16	0	3	2	0	1	0
		17	0	3	3	0	3	0
		18	0	3	3	0	3	0
		19	3	3	3	0	3	0
		20	0	1	2	0	2	0
		Total	8	17	19	8		
		Specific	0					
		Nonspecific	20					
		None	0					
gypsy_ΔSu(Hw) (pL1: pCAMBIA 1391 napin-GUS)	386	1	0	0	1	0	1	0
		2	1	3	3	0	2	2
		3	1	3	3	0	3	0
		4	0	0	0	0	1	1
		5	0	0	0	1	0	1
		6	0	0	0	0	0	0
		7	1	3	2	0	2	2
		8	2	3	3	1	2	2
		9	1	2	3	1	3	1
		Total	5	5	7	6		
		Specific	1					
		Nonspecific	7					
		None	1					
BEAD_INVgen2#14 (pL1: pCAMBIA 1391 napin-GUS)		1	0	0	0	2	0	2
		2	0	0	0	1	0	2
		3	0	0	0	1	0	2
		4	0	0	0	2	1	1
		5	0	0	0	1	0	1
		6	0	0	0	2	1	1
		7	0	0	0	1	0	1
		8	0	0	0	1	0	1
		9	0	0	0	0	0	1
		10	0	0	0	1	0	1
BEAD_INVgen2#15 (pL1: pCAMBIA 1391 napin-GUS)		11	0	0	0	0	0	1
		12	0	0	0	1	0	2
		13	0	0	0	1	0	2
		14	0	0	0	1	0	2
		15	0	0	1	0	0	2
		16	0	0	0	1	0	2
		17	0	0	0	1	0	2
		18	0	0	0	0	0	1
		19	0	0	0	1	0	3
		20	0	0	0	0	0	1

BEAD_INVgen2#16 (pL1: pCAMBIA 1391 napin-GUS)	21	0	0	0	0	0	2
	22	0	0	0	0	0	2
	23	0	0	0	0	0	2
	24	0	0	0	0	0	2
	25	0	0	0	0	0	2
	26	0	0	0	2	0	2
	27	0	0	0	0	0	2
	28	0	0	0	2	0	2
	29	0	0	0	1	0	2
	30	0	0	0	3	0	2
BEAD_INVgen2#19 (pL1: pCAMBIA 1391 napin-GUS)	31	0	0	0	3	0	2
	32	0	0	0	2	0	2
	33	0	0	0	2	0	2
	34	0	0	0	2	0	2
	35	0	0	0	2	0	2
	36	0	0	0	3	1	2
	37	0	0	1	2	0	2
	38	0	0	0	2	0	1
	39	0	0	0	1	0	2
	40	0	0	0	2	0	2
	Total	0	0	4	40		
	Specific	36					
	Nonspecific	4					
	None	0					

Analysis of insulator stability in second generation plants

Materials and Methods

In studying the F2 generation of insulators, 4 transgenic plants were selected from the transformation of BEAD1c_INV. The plants were left to dry for approximately 4 weeks, and seeds were collected. These seeds were sterilized, plated on hygromycin-kanamycin plates and selected to grow on soil as described in Chapter 2, Section 2.3. Tissue samples were collected and scored according to previous method as well.

Results:

In addition to characterizing non-plant insulators, our goal was to confirm the stability of these insulator sequences in subsequent generations within our enhancer-blocking assay. This was done by growing seeds of the F1 generation transgenic plants already containing a functional insulator sequence, and testing the transgenic tissue from the F2 generation. A total of 4 transgenic plants from the F1 generation were kept and dried for seed collection. The seeds were collected from transformation of BEAD1cINV which produced GUS expression level of 84% (16/20). Seeds were grown on bacterial plates for selection and these F2 generation plants were tested for GUS expression. Results are shown in Table 3.2.

Table A3.2: Enhancer-promoter interactions in transgenic *Arabidopsis thaliana.* Summary of the results from the testing of F2 generation of BEAD1cINV constructs for β-glucoronidase (GUS) staining. The columns indicate the construct name and size, the

number of individual transformants tested for leaf, flower, silique and seed staining, and a calculation of the specific and nonspecific expression of the GUS transgenes.

			GUS staining				GUS expression		
Construct Name	Size (bp)	No. plants tested	Flowers	Leaves	Siliques	Seeds	Specific	Non specific	None
Gen2_14		10	0	0	0	10	8/10	2/10	0
Gen2_15		10	0	0	1	10	9/10	1/10	0
Gen2_16		10	0	0	0	10	10/10	0/10	0
Gen2_19		10	0	0	2	10	9/10	1/10	0