Analysis of Potential DNA Insulators in

Arabidopsis thaliana

by

Lara Rasooli

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirement for the Master of Science degree
Ottawa-Carleton Institute of Biology
University of Ottawa

Thèse soumise à la Faculté des Études Supérieures et Postdoctorales
En vue de l’obtention de la maitrise ès science
L’institut de biologie d’Ottawa-Carleton
Université d’Ottawa

Abstract

The control of transgene expression is the key to producing transgenic organisms with beneficial traits; however, the prevention of transgene misexpression in transformed organisms has been a recurring challenge and may lead to the expression of undesirable traits. Insulators are DNA sequences that regulate gene expression through their interaction with protein factors. They function as cis-acting DNA sequences with the ability to act as enhancer-blocking elements or barrier elements, or both depending on the molecular mechanism. The functions of insulators have been tested using a model system based upon transgenic *Arabidopsis thaliana* plants to determine the expression of transgenes and to decrease non-specific interactions resulting in misexpression. To date, various insulators have been identified in invertebrates and vertebrates but few have been discovered and analyzed in plants.

This study aims to identify insulators that can function to block enhancer-promoter interactions in transgenic plants preventing transgene misexpression. An oligonucleotide library was produced allowing for subsequent screening of putative insulators. Screening of recombinant clones was initiated to identify strong insulator candidates using vectors with different promoter/reporter systems in *Arabidopsis*. To date, three insulators (CLOI-3, CLOII-10, and CLOII-12) have been identified as candidate insulators that block enhancer-promoter interaction. The CLOI-3 insulator has been further analyzed through the introduction of deletions to identify crucial regions. Two replicates for both CLOII-10 and CLOII-12 indicate possible insulator activity. By characterizing new insulator sequences we will be able to further understand the mechanisms involved in gene expression and apply them to transgenic research involving other biological systems, ultimately addressing important issues such as crop development.

Résumé

Le contrôle de l'expression du transgène est la clé qui permet de produire des organismes transgéniques aux caractères bénéfiques; cependant prévenir la mauvaise expression du transgène dans des organismes transformés constitue un défi récurrent et peut mener à l'expression de caractères non désirés. Les isolateurs sont des séquences d'ADN qui régulent l'expression du gène à travers leur interaction avec des facteurs protéiniques. Ils fonctionnent telles des séquences d'ADN agissant en cis, avec la capacité d'agir comme éléments bloquant l'amplification ou des éléments barrière, ou bien les deux, en fonction du mécanisme moléculaire à l'œuvre. Les fonctions des isolateurs ont été testées en utilisant un système modélisé, basé sur des plantes transgéniques *Arabidopsis thaliana*, afin de déterminer l'expression des transgènes et de diminuer les interactions non spécifiques, résultant d'une mauvaise expression du transgène. Jusqu'à aujourd'hui plusieurs isolateurs ont été identifiés chez des invertébrés et des vertébrés mais peu ont été découverts et analysés dans les plantes.

Cette étude a pour but d'identifier des isolateurs fonctionnels qui peuvent bloquer les interactions entre des séquences régulatrices amplificatrices (enhancer) et des promoteurs chez des plantes transgéniques et ainsi prévenir une mauvaise expression du transgène. Une banque d'oligonucléotides a été créée permettant ensuite de conduire une recherche d'isolateurs putatifs. Une recherche de clones recombinants a été réalisée afin de mettre en évidence d'éventuels isolateurs en utilisant des vecteurs avec différents systèmes de promoteurs/reporteurs chez *Arabidopsis*. Pour l'heure, trois isolateurs (CLOI-3, CLOII-10 et CLOII-12) ont été identifiés comme potentiels candidats qui bloquent l'interaction amplificateur-promoteur. L'isolateur CLOI-3 a été analysé en profondeur à travers l'introduction de suppressions pour identifier les régions cruciales. Des résultats pour des répliques des deux éléments CLOII-10 et CLOII-12 indiquent une possible activité isolante. En caractérisant les séquences de nouveaux isolateurs, nous serons capable de mieux

comprendre les mécanismes dont relève l'expression des gènes et les appliquer à la recherche transgénique avec l'implication d'autres systèmes biologiques. Cela permettra éventuellement de se pencher sur des questions importantes telle que le développement de cultures.

Acknowledgments

I would like to thank all my friends and family for their support throughout my education. I would especially like to thank my mom for her patience, advice, and support. Without her constant strength and presence in my life I wouldn’t be who I am today. Thank you to my supervisor Dr. Johnson, who gave me the opportunity to work in his lab. I was very lucky to have learned the many skills in our field first hand from such a great professor and scientist. Thanks for being such a great mentor throughout my Masters at uOttawa. In addition, thank you to my committee members Dr. Ouellet and Dr. Xing for all their help and advice throughout my Masters. Next, I have to thank all the people who have played such an important role in making this project happen. Thank you to Batool, Anh, Richard, and Hassan for all their support and help in finishing this thesis, I couldn’t have done it without you. Thank you to Ram for his help in regards to bioinformatics. I am thankful for the opportunity to complete my Master of Science as I have learned skills that will help me in my future endeavors.

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

Section **Page**

List of Tables and Figures

List of Tables and Figures (continued)

List of Abbreviations

AGIP promoter	*AGAMOUS* second intron-derived promoter
bp	Base pairs
°C	Celsius degree
CaMV	Cauliflower Mosaic Virus
CodA	cytosine deaminase
Col	Columbia
CP190	centrosomal protein 190
CTCF	CCCTC-binding factor
ddH_2O	Distilled de-ionized water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonuleotide triphosphate, any of the four bases
dTopors	topoisomerase I-interacting RS protein
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
g	Gram
GUS	*β*- glucuronidase
hptII	Hygromycinphosphotransferase type II
Kb	Kilobase
KCI	Potassium chloride
$K_4Fe(CN)_6^{-}3(H_2O)$	Potassium ferrocyanidetrihydrate
$K_3Fe(CN)_6$	Potassium ferricyanide
L	Liter
LB	Luria broth
mg	Milligram
µg	Microgram
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulfate
µl	Microliter
mL	Milliliter
µM	Micromolar
mM	Millimolar
NaCI	Sodium chloride
$NaH_2PO_4^{-}H_2O$	Sodium dihydrogen phosphate monohydrate
$Na_2HPO_4^{-}2(H_2O)$	Sodium phosphate dibasic dihydrate
NaOH	Sodium hydroxide
Nos	nopaline synthase
PCR	Polymerase chain reaction
rpm	Revolutions per minute
siRNA	Small interfering ribonucleic acid
Su(Hw)	Suppressor of hairy wing
Taq	*Thermus aquaticus* DNA polymerase
TBE (10X)	0.89 M Tris, 0.89 M Borate, 20 mM EDTA, pH 8.3
Tbf1	Telobox-containing general regulatory factor
T-DNA	Transferred DNA
Tris	Tris (hydroxymethyl) aminomethane
Reb1	RNA polymerase I enhancer binding protein
X-Gluc	5-bromo-4-chloro-3-indolyl-*β*-D-glucuronic acid

35S	Cauliflower mosaic virus 35S promoter/enhancer
35S46	Core cauliflower mosaic virus 35S promoter
5’HS4	5’ constitutive hypersensitive site

1.0 Introduction

Genetic engineering of plant species is an ongoing topic in research, directly impacting agricultural industries, internationally. Biotechnology has greatly influenced findings in this crucial area of research by enhancing efficiency and sustainability of crops and selecting for desirable traits. We are proposing the use of insulators as a method to facilitate the selection of desired traits by blocking enhancer-promoter interaction in transgenic plants.

This thesis describes the cloning of DNA fragments with insulator activity in *Arabidopsis thaliana*. The analysis of these insulators will help us understand the mechanism of insulator activity *in planta* and potentially lead to improving the recovery of transgenic plants.

1.1 Chromatin structure and gene regulation

Gene expression in specific tissues and at specific times of development is determined by regulatory elements such as promoters, repressors, activators, enhancers and silencers. These components are regulatory elements within chromatin, which is a combination of DNA and proteins (e.g. histones) within the nucleus of a cell called nucleosomes. Chromatin can be defined as active (euchromatin) or inactive (heterochromatin) through its functional role in gene expression within a cell. Euchromatin has a loosely packed configuration with a very high gene concentration where transcription occurs. In contrast, heterochromatin is tightly packed and inactive because of its silencing effects on the genes located within it (West, 2002). The distinct positioning of euchromatin and heterochromatin within the nucleus of eukaryotic cells is thought to correlate with particular environments for chromatin activation and repression. This is very important for the various ways chromatin structure is established, as histone modifications are implemented through acetylation and methylation. The acetylation of chromatin increases gene expression producing areas of activity (euchromatin) in contrast to methylation, which causes repression (heterochromatin) by preventing the binding of transcription factors. The way euchromatin and heterochromatin function in these elements through histone-modifying complexes (acetylase/deacetylase; methylase/demethylase) is crucial to the functioning of insulator activity within a given species. These conversions ultimately affect the results of gene expression through silencing or activation (Valenzuela, 2006). The presence of these genomic zones of activity/repression often proves to be a hindrance to plant biotechnology, as chromatin-mediated silencing of the introduced transgene can occur if integration takes place within or near a region of heterochromatin (Singer et al., 2012).

The successful selection of transgenic plants requires strong constitutive promoters to express selectable marker genes to avoid unnecessary and time-consuming screening for the trait of interest. Promoters are components that stimulate gene expression, depending on their specific function. The constitutive cauliflower mosaic virus *(CaMV) 35S* promoter is widely used in plant research to drive expression of selectable marker genes but also has enhancer elements that can influence nearby promoters. Our increasing requirement to genetically modify plants with targeted transgene expression has driven our need to better understand the complexity of enhancer-promoter interactions and to devise means to control or eliminate them. Enhancers and silencers are small *cis*-acting DNA elements that act from a distal location to activate or repress transcription. Enhancers stimulate the activation of promoters to drive gene expression within a given domain. Enhancers in transcriptionally active chromatin (euchromatin) bind the transcription factors that help initiate RNA transcription at the promoter. In contrast, silencers, found in transcriptionally inactive chromatin (heterochromatin), recruit repressor proteins producing inactive chromatin domains. In some cases this heterochromatin stimulates the formation of more heterochromatin in adjacent DNA, a process known as heterochromatinization. Enhancer-bound transcription factors have displayed associated protein attachments at the promoter region that are involved in structural changes such as looping that contribute to producing chromatin remodeling complexes, which control chromatin architecture (Valenzuela, 2006).

1.1.1 Selectable traits and gene expression

Advantageous traits, both genotypic and phenotypic, can be selected by using transformation vectors that harbor multiple transcriptional gene units to enhance several traits in the specified plant at the same time. Unfortunately the expression of several traits in the same vector may cause gene misexpression leading to undesirable interactions. This is caused by the presence of multiple enhancer/promoter elements within a single vector, resulting in undesired interactions between the enhancers and nearby promoters of other genes (Singer, 2011). Furthermore, misexpression may be more frequent in transgenic plants containing vectors in which the enhancer is derived from the *CaMV 35S* promoter and is situated near a plant-derived promoter. The *35S* promoter sequence functions as an enhancer in either orientation (Odell et al., 1988) and is capable of overriding the control of transgene expression when present in the same construct. This result was seen by Hily et al. (2009), where the *35S* enhancer is able to effectively activate the stamen- and carpel-specific *AGIP* promoter in non-targeted leaf tissue in *Arabidopsis*. These interactions could be within the vector or between elements within the vector and within the genomic region surrounding the insertion site. Therefore, we need to find and incorporate mechanisms into vectors to prevent these interactions in such DNA regulatory elements in order for appropriate gene expression to take place. Thus, we are proposing that *cis*-acting DNA regulatory sequences called insulators have the ability to block enhancer-promoter interactions when positioned in between the two elements to better control gene expression in transgenic plants (Valenzuela, 2006).

1.2 Regulatory elements and insulator function

In the past, research in plant biotechnology focused on improvement of a single trait; however since plants cope with numerous conditions such as temperature, physical barriers, and lack of resources such as water, indicates the need for an approach that is able to use multiple traits simultaneously using transformation constructs that have several transcriptional units. This is often achieved using a strong, constitutive promoter/enhancer to direct the expression of a selectable marker gene in combination with tissue-, organ- or developmental stage-specific promoters to drive the expression of transgenes such as the *CaMV 35S* enhancer/promoter that is commonly used in vectors within plants (Odell, 1988). Unfortunately, this approach can be problematic due to the position- and orientation-independent ability of enhancers to trigger enhancer-promoter interference, which can influence both the strength and the specificity of transgene expression (Singer, 2012) as indicated above. This leads us to believe that components such as insulators are a solution to expressing a gene of interest within a transformation vector to produce the transgene of interest. Furthermore, the identification and subsequent analysis of conserved insulator functions in plants, allow us to infer that mechanisms of insulator function seen in other species as explained in **section 1.4**, such as in *Drosophila melanogaster* and *Gallus gallus* to name a few (Chung, 1993; Geyer, 1997), exist in plants.

1.2.1 Insulators and Enhancer-Blocking Activity

Enhancer-blocking elements are considered insulators when they are situated between the enhancer and promoter, blocking communication (**Figure 1)**. This may impede interaction, preventing gene expression. Three major models have been proposed to explain the ability of insulators to act as enhancer-blocking elements: the promoter decoy model, the facilitator model, and the structural (looping) model, all described further in **section 1.3** (Valenzuela, 2006). The onset and extent of enhancer-promoter crosstalk that occurs within a transgene is primarily influenced by the strength of the enhancer, its distance to the target promoter, and the responsive nature of the promoter. The responsiveness can be due to binding proteins that either causes a physical blockage of communication or chromatin loop formation that separates the enhancer from the promoter through the formation of 'insulator bodies' (Singer, 2012). As an example, in *Drosophila* insulator bodies are located on the nuclear matrix and are made up of many suppressor of hairy-wing Su(Hw)-insulator DNA complexes, brought together and held by interactions with Mod(mdg4) and centrosomal protein CP190, producing separate chromatin loop domains which establish the separation of the enhancer from the promoter (Geyer, 1997). In a study of the 5'HS4 insulator in chicken, Bell et al. (1999) described a ubiquitous DNA binding protein, CTCF (CCCTC-binding factor), as a key conserved functional component of boundary domains within various vertebrates as summarized in **Table 1**. This binding protein in addition to other components specific to the insulator, produce similar 'looping domains' as seen in *Drosophila* for enhancer-blocking activity. Numerous insulators with enhancer-blocking activity have been identified in plants and non-plant systems (**Table 2 and 3**) and a well-described enhancer-blocking insulator is the *gypsy* insulator from *Drosophila* (**section 1.4.1**) (Gerasimova, 2000).

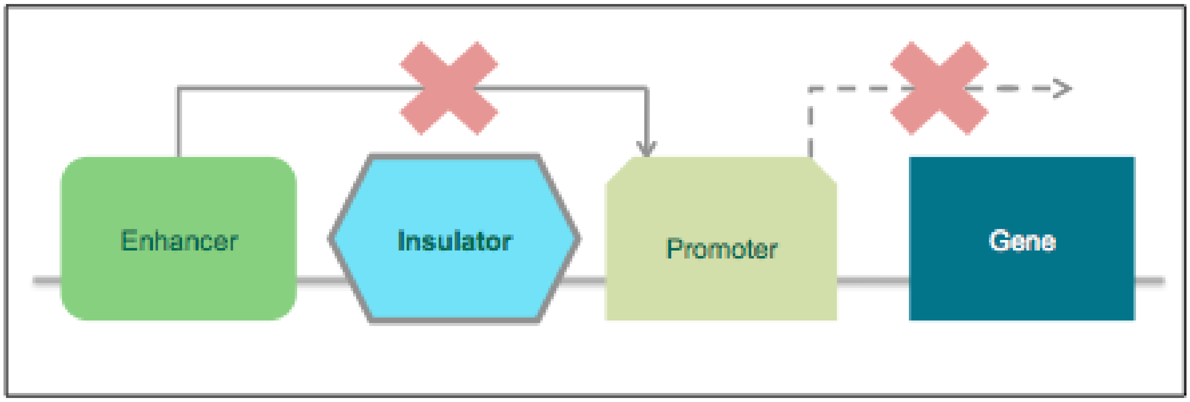


Figure 1. Model representing enhancer-blocking activity. The insulator is positioned between the enhancer and its target promoter preventing their interaction and subsequently blocking gene expression (modified from Valenzuela et al., 2006).

1.2.2 Insulators and Barrier Activity

Barrier elements block the linear spread of heterochromatin to prevent silencing of the flanking active genes. Insulators act as a gatekeeper (barrier) that separates the transcriptionally active gene in the euchromatin domain from the silenced gene in the heterochromatin domain (repressed gene in adjacent inactive domain) (**Figure 2**). Alternatively, specific DNA sequences and associated proteins may have the role of establishing fixed boundaries. Proteins bound to barrier insulators are believed to function by decreasing the silencing affects of heterochromatin in neighboring genomic regions. For example, proteins that bind the 5'HS4 insulator (**section 1.4.4**) are involved in the functioning of 5'HS4-mediated acetylation and methylation of H3K4 in nucleosomes as well as protection against DNA methylation (Singer, 2012; Pikaart et al., 1998). These are strategies through which 5'HS4 may be able to counteract heterochromatic silencing in transgenic plants. Due to the random nature of transgene insertion in the majority of higher eukaryotes, transgenic DNA may integrate into regions of the genome that are heterochromatic leading to transgene silencing. Since a majority of a plant genome can be in the form of heterochromatin, the chance that foreign DNA will integrate in or near these regions, and consequently be silenced, is relatively high. Kim et al. (2007) identified 117 T-DNA insertion sites from *Arabidopsis* suspension cells transformed by *Agrobacterium* and propagated under non-selective conditions to determine the integration frequency into heterochromatic regions. High frequency of T-DNA insertions (x/117) within heterochromatic regions after propagation indicates the increasing possibility of integrating transgenes in or near these regions (Kim et al., 2007). In addition, transgenes may be integrated near transcriptional enhancers or silencers, which could cause mis-expression (Singer et al., 2012). To protect against the encroachment of heterochromatin into a euchromatic region, insulator elements may recruit factors that put 'activating' histone modifications on adjacent nucleosomes, blocking the spread of 'silencing' histone marks. Interference of silencing by heterochromatin may also involve RNA dependent mechanisms, such as small interfering RNAs (siRNAs) implicated in the initiation of transcriptional gene silencing (Sugiyama et al., 2005). By flanking a transgene with elements that block the spread of heterochromatin, a foreign gene may be expressed appropriately regardless of its insertion site within the host genome (Valenzuela, 2006).

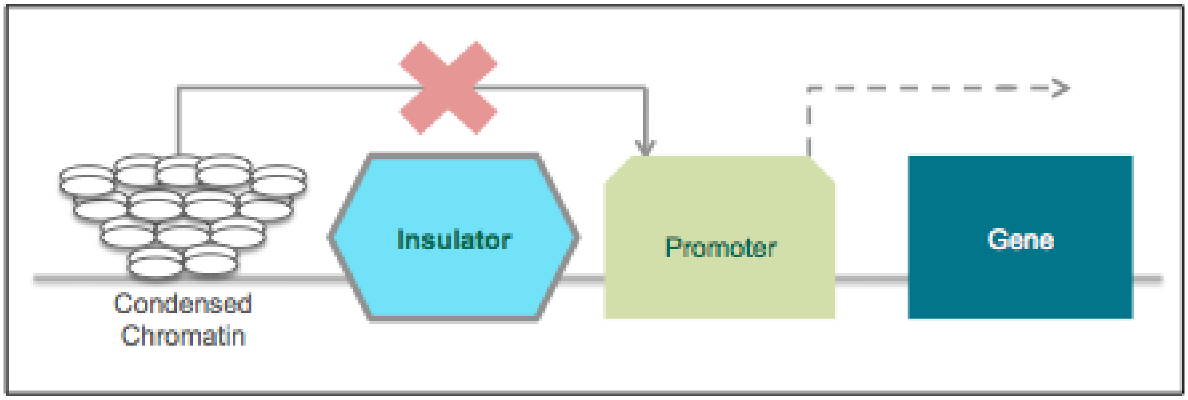


Figure 2. Model representing barrier activity. The insulator is positioned to block the spread of heterochromatin to prevent the silencing of the euchromatin domain (modified from Valenzuela et al., 2006).

1.2.3 Insulators and Position Effects

Position effect refers to the variability of expression levels that is observed when a gene is relocated from its native context to a new chromatin environment. Chromosomal position effects can strongly influence the transcription of foreign genes in transgenic plants and animals. This results in low frequencies and levels of gene expression and, in some cases, in unusual patterns of expression. When insulator elements are integrated into a given domain they flank the transgene, which generates a functionally independent domain preventing the occurrence of position effects in insulators such as the *gypsy* insulator in *Drosophila* and *Arabidopsis* (**section 1.4.1 and 1.5.1.4**) (Clark, 1994; Valenzuela, 2006) (Gerasimova, 2000; She et al., 2010).

1.2.4 Proteins that interact with insulators

Insulator activity has been identified in a number of species, but defining all the components involved in facilitating this activity such as binding proteins and investigating the pathway by which insulation is achieved remains difficult. CCCTC is a common binding site found among vertebrate insulators that are made up of 11 zinc-finger domain proteins, which bind DNA, that aid in domain structure and functionality. For example, the vertebrate 5'HS4 insulator binds the CCCTC site, which in turn recruits other specified proteins, thus facilitating the subsequent attachment of an insulator to the nuclear lamina creating 'insulator bodies'. Insulator bodies function to separate the enhancers and promoters into distinct domains (Gaszner et al., 2006; Walters et al., 1999). Furthermore, 'insulator bodies' are also found in *Drosophila*, the Su(Hw) protein is used to undergo this structural model indicating a common mechanism involved in enhancer-blocking activity. Boundary element-associated factor (BEAF) is a protein factor that interacts within the scs' insulator to block both enhancer-promoter communication and silencing by heterochromatin (Emberly, 2008). As of yet, no functional equivalents of CTCF binding sites have been identified in plants; however, a number of plant zinc-finger gene families share similarity at the amino acid level with the zinc-finger domains of vertebrate CTCF (Singer, 2011). Both enhancer-blocking insulators and barrier insulators appear to require the binding of protein factors. However, while proteins bound to enhancer-blocking insulators (**Table 1**) seem to provide either a physical blockage of communication or the formation of chromatin loops, those bound to barrier insulators are believed to function through heterochromatic silencing from neighboring genomic regions as described by Singer et al., (2012).

1.2.5 Insulator interactions and gene regulation within a domain

How insulators interact within the same domain and the components involved in efficient insulator activity have been hypothesized, but many of the detailed mechanisms are still unknown. As described in **section 1.2.1** and **1.2.2** insulators can be categorized into two functions, enhancer blocking elements, barrier elements, or both. It is important to note that not all insulators have the same sequence, but the proteins that bind to them may have conserved motifs such as zinc-finger proteins as summarized in **Table 1**. Subsequently these functions are described through regulatory models that allow us to understand how insulators are able to block enhancer promoter communication or block the spread of heterochromatin to prevent silencing of an active domain described in **section 1.3**. These models are based on transcriptional modifications through complexes made up of heterochromatin and euchromatin, which is carried out by histone modifications. Histone-modifying complexes catalyze the change in binding affinity between histones and DNA causing the change in structure that can be seen as heterochromatin or euchromatin through acetylation and methylation, ultimately affecting gene expression through activation or silencing. A structural model that prevents enhancer-promoter communication was initially proposed through the experimental data originating from the discovery of the insulator from the *Drosophila* retrotransposon *gypsy*. They defined insulators as *cis*-acting elements, which, via the binding of specific *trans*-acting proteins cause the formation of chromatin loops, which define domains through this model (Mongelard & Corces, 2001). Therefore, an enhancer outside of the ‘loop’ domain is prevented from interacting with a promoter within, thus blocking interaction and misexpression. Binding factors associated with individual insulators such as common proteins directly involved in these models are key components in the structural models involved in creating discrete domains. Proteins such as the Su(Hw) from *Drosophila* in addition to other components specified to the insulator produce the ‘insulator bodies’ that have shown to cause looping in order to separate the enhancer from the promoter producing distinct looping domains. There are other models of insulator function as described in **section 1.3** but it is interesting to point out that different insulators have been proposed to function through common models such as the structural formation of loops to separate the enhancer and promoter. Furthermore, you are able to find various proteins involved in these structures (**Table 1**) such as the CTCF factor, and Su(Hw) protein from multiple systems like *Drosophila* and *Homo sapiens*.

A number of experiments have looked at various characterized insulators and their interactions in tandem. This allows us to determine whether or not insulator activity will be strengthened by one or more insulators in addition to the binding proteins that may be involved in its function and its subsequent role in re-modeling chromatin activity. Cai et al. (2001) determined that when a direct tandem repeat of the *gypsy* insulator was used instead of a single copy, insulator activity was abolished. Muravyova et al. (2001) using the *yellow* and *white* reporter genes and a tandem repeat of *gypsy* insulators obtained the same results.

The results obtained for in tandem *gypsy* elements were not reproduced using combinations of other insulators (Majumder et al., 2003); in fact, increased or similar insulator activity was normally observed. In *Drosophila* six insulators have been identified, *gypsy*, scs, scs', Fab7, SF1 and 1A2 (**Tables 1 and 2**) with DNA sites and binding partners, with different zinc finger proteins only known for *gypsy*, scs, and scs' (**Table 1**). When pairs of different insulators (*gypsy* & scs; *gypsy* & Fab7; *gypsy* & SF1) or the homologous pairing of scs & scs' were arranged in tandem, enhancer blocking activity was significantly stronger than that of either insulator alone. Lastly scs was tested in tandem with Fab7 and SF1, and demonstrated strong enhancer-blocking activity as well. Cai et al. (2001) and Muravyova et al. (2001) obtained similar results using different promoter and enhancer combinations. These results suggest that insulator pairs can exhibit increased activity than either element alone, suggesting that both insulators can function independently in an additive or parallel manner.

Organism	Insulator name	DNA target	Binding Protein	References
Drosophila melanogaster	*gypsy*	TGCTGCATACTT (core)	Su(Hw) (Suppressor of Hairy Wing, 12 zinc fingers)	Geyer et al. (1986)
Drosophila melanogaster	*scs*	TTCGTGCGAA (core)	Zw5 (zest-white-5, 8 zinc fingers)	Kellum et al. (1991) & (1992)
Drosophila melanogaster	*scs'*	dual core CGATA	BEAF (boundary element-associated factor, "1 zinc finger")	Kellum et al. (1991) & (1992)
Drosophila melanogaster	Fab7	GAGA (core)	GAF	Hagstrom et al. (1996)
Ashbya gossypii	UASrpg	TAGCCCATACATC; GCATCCATACATTT	Rap1 (repressor activator protein)	Bi et al. (1999)
Saccharomyces cerevisiae	STARs	Unknown	Tbf1, Reb1	Fourel et al. (1999)
Homo sapiens	BEAD-1	CCGCGNGGNGGCAG	CTCF (CCCTC-binding factor or 11-zinc finger protein)	Zhong et al. (1997)
Homo sapiens	BEAD-1C	CCGCGNGGNGGCAG	CTCF (CCCTC-binding factor or 11-zinc finger protein)	Zhong et al. (1997)
Homo sapiens	5'HS4	TAATTACGTCCCTCCCCCGCTA	CTCF (CCCTC-binding factor or 11-zinc finger protein)	Chung et al. (1993)
1-kb bacteriophage lambda fragment	EXOB	Unknown	Binding partner unknown	Singer et al. (2010)
Arabidopsis thaliana	*Atgypsy*-like element	Unknown	Binding partner unknown	Singer et al. (2013)
Oryza sativa	*HS185*	GGTCCCACA	Binding partner unknown	Zhang et al. (2012)
Petunia hybrida	*TBS*	Unknown	Binding partner unknown	Hily et al. (2009)

Table 1. Characterized selected insulators in a range of organisms with their associated binding proteins.

1.3 Models for Insulator Function

The enhanced insulator activity of the tested paired insulators (with the exception of the *gypsy* insulator), suggests that two insulators could function in parallel or additively. The native targets of these proteins could be chromatin-remodeling complexes such as enhancer-binding proteins (promoter-decoy model) or facilitator proteins (enhancer-facilitator model). Cai et al. (2001) and Muravyova et al. (2001) obtained similar results, using different promoter and enhancer combinations. The implication is that two insulators interact, through the protein complexes bound to them. A possible explanation is that the "looping out" of the sequences separating enhancer and promoter displaces the insulators and, by bringing the enhancer and promoter closer, may even stimulate expression.

The decoy model allows us to predict that two insulators would be better than one since a tandem repeat of insulators would reinforce the trapping of the enhancer, but the results of these two papers are more consistent with the structural model where two insulators physically interact and promote the looping of the intervening sequence. As a result the distance between enhancer and promoter is reduced, and transcription activation then takes place. This demonstrates that the mere interaction between repeated copies of the insulator is not responsible for the inability to interfere with enhancer-promoter communication but rather the special chromatin structure generated by the interaction of neighboring insulators (Mongelard & Corces, 2001).

By understanding all the factors involved in insulator interaction and the similarities that contribute to their function, we can ultimately predict that there may be multiple systems since different insulators work in the same species such as in *Drosophila* (**Table 2**). Furthermore, we can see that different systems can recruit common proteins (**Table 1**) that conclusively produce a common model such as loop formation.

Three main models for insulator function have been postulated and I will describe each in turn. As can be seen from **Tables 2** and **3**, and the discussion of the effects of multiple insulators found in the previous section, a single species may have more than one functional insulator raising the possibility that each sequence functions through binding of specific proteins (**Table 1**). Since the exact pathway by which binding leads to insulator function is unknown, it is possible that all or a subset of models could explain function within a single cell. The simplest assumption would then be that the different pathways feed into a common pathway by binding the same proteins necessary for insulator activity e.g. the formation of insulator bodies. This idea is explored later.

There has been little progress in the literature indicating new models for insulator activity. The review by Valenzuela, (2006) is the most recent summary of the various models. We can only indicate that recently identified insulators have some similarities to known insulators, which indicated conservation of mechanisms involved in enhancer-blocking activity such as looping domains used in structural models (Bell, et al., 1999).

1.3.1 Promoter-decoy model

The promoter-decoy model represents the direct prevention of enhancer-promoter interaction by the insulator (**Figure 3**) by promoting protein-protein interactions between enhancer-bound proteins and insulator proteins instead of the promoter (Valenzuela, 2006; Geyer, 1997). Several modifiers of the *gypsy* insulator are gene specific, suggesting that the mechanism of insulation is related to that of enhancer function (Georgiev & Kozycina, 1996). Also, there is a correlation between enhancer strength and the level of insulation when a scs insulator is used experimentally to block the interaction between the yolk protein-1 enhancer element and an *hsp70* promoter:*lacZ* fusion gene driving β-galactosidase in *Drosophila* transgenics (Vazquez & Schedl, 1994). Therefore, insulators can establish independent functional domains simply by assembling a structure, which interferes with enhancer-promoter communication. Another example of an insulator that is proposed to use this model is the 5’HS4 insulator (Chung et al., 1997). They have suggested that the 5'HS4 insulator might work by preventing 'tracking' of a distant enhancer complex along the chromatin, by isolating the region in the nuclear matrix by acting like a signal to physically isolate the region (Chung et al., 1997).

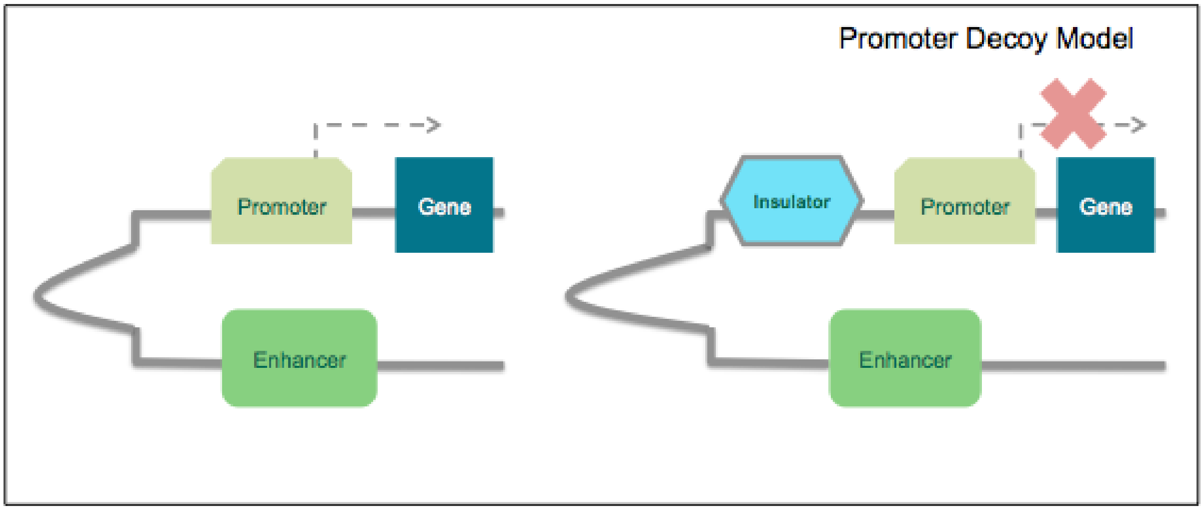


Figure 3. Schematic of the promoter decoy model. The insulator protein complex prevents the enhancer-bound proteins from interacting with the promoter. The insulator acts as a 'decoy' by intercepting the enhancer signal from interacting with the promoter (modified from Valenzuela et al., 2006).

1.3.2 Facilitator model

The facilitator model involves enhancer utilization of facilitator factors to interact with the promoter. The presence of an insulator interferes with the function of facilitator factors. This prevents the enhancer from effectively communicating with the promoter (**Figure 4**). This model displays insulator interference with facilitator proteins preventing the enhancer from communicating with the promoter, in contrast to the promoter-decoy model that focuses on promoting protein-protein interactions with the enhancer and insulator in order to block communication with the promoter (Valenzuela, 2006). Insulators in the facilitator model interfere with enhancer-promoter communication without inactivating enhancers, and without competing with promoters to block communication. An example of the role of facilitator factors has been proposed in the *Drosophila Ultrabithorax regulatory region*, whereby the region between the *iab-7* enhancer and promoter bind unknown proteins that tether the *iab-7* enhancer to the *cis* promoter (Sipos, 1998). When these sites are deleted, the enhancer is released to interact with the *trans* promoter which indicates that the region between the *iab-7* enhancer and its promoter contains multiple interaction sites for facilitators and that deletion of the sites may cause communication with the active promoter (Dorsett, 1999). In addition, this model can help to explain the result that tandem insulator sequences improve insulator activity. Two insulators will be able to better prevent the function of facilitator factors in aiding enhancer-promoter communication.

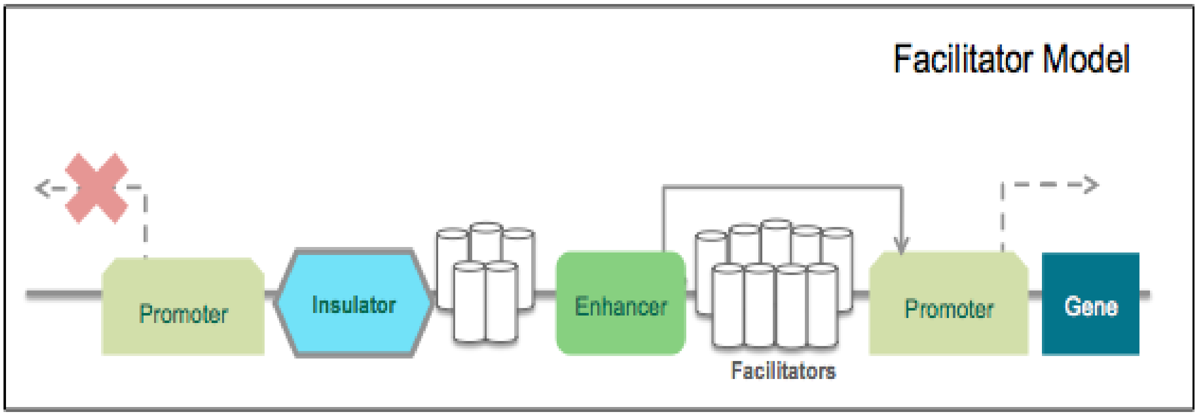


Figure 4. Schematic of the facilitator model. The facilitator factors are used by the enhancer to interact with the promoter. The insulator is positioned between the enhancer and promoter interfering with the function of the facilitator factors (modified from Valenzuela et al., 2006).

1.3.3 Loop model

The loop model relies on the formation of ‘loops’ that represent an independent structural and functional domain. The attachment of the insulator to the nuclear lamina divides the chromatin fiber into separated looped domains, which isolates the signal generated in a domain (**Figure 5**). As a result, the interaction between the enhancer and the promoter only occurs if they are located in the same loop. Loops can be generated by the pairing of insulator elements or by the attachment of an insulator to a fixed structure (Valenzuela, 2006). Furthermore, binding proteins specific to the insulator play an important role in the configuration of the loop model. Data from all characterized insulators thus far seem to support a mechanism known as the looping model, where defined chromatin domains bring some elements into proximity, allowing activation of some genes, or isolating others, causing repression (Bell et al., 2001). For example, the *gypsy* insulator interacts with the suppressor of hairy wing Su(Hw), a protein with twelve zinc-finger DNA binding domains. Su(Hw) further recruits factors such as Mod(mdg4)2.2, topoisomerase I-interacting RS protein (dTopors) and the centrosomal protein CP190 (**Figure 6**) (Pai et al, 2004; Gerasimova, 2000). These factors combine to form ‘insulator bodies’ which have been proposed to organize chromatin into loops to create distinct domains to separate the enhancer from its target promoter (Bushey et al., 2008).

Evidence from *Drosophila* supports the idea that different insulators can function through a common mechanism using similar proteins. This can be seen through the comparison of a number of *Drosophila* insulators such as *gypsy*, scs/scs’, SF1 and 1A2 (**Table 2**) that all obtain enhancer-blocking activity through the proposed structural ‘looping’ model. We should note that there are common interacting proteins that are essential to their function such as CP190 (Geyer, 1986; Kellum, 1992; Majumder, 2003; Chetverina, 2012) but the insulator sequences are not the same and also have proteins that are essential to their function that are not found in the other insulators. It is still interesting to see that there are some similarities within some of the insulators found in the same species even though there is no sequence correlation.

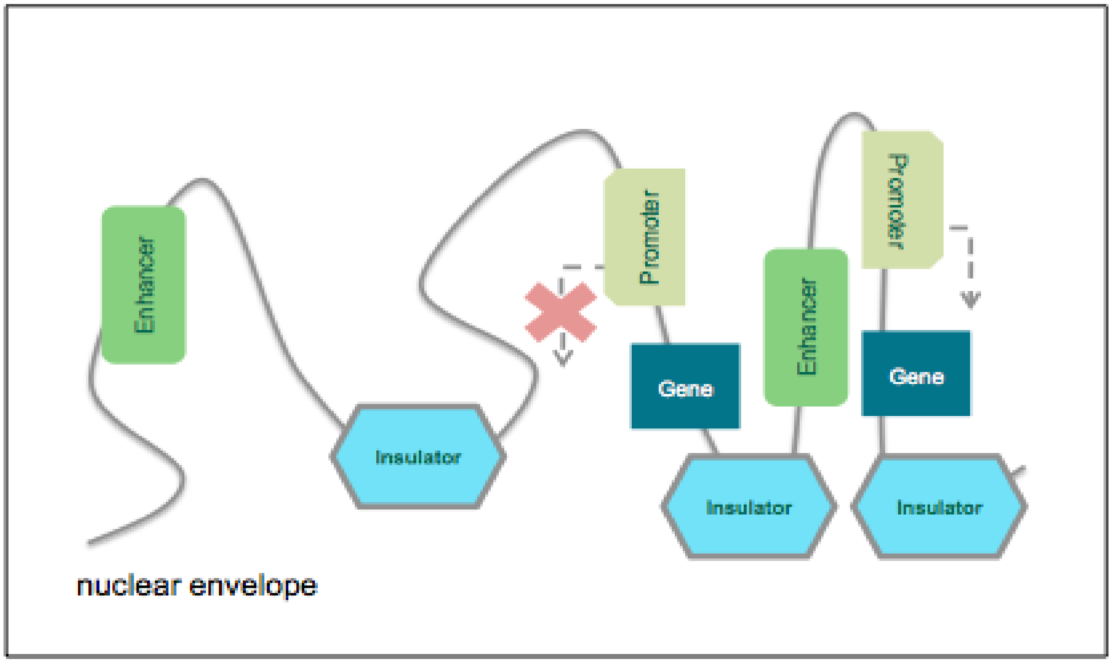


Figure 5. Schematic of the structural (looping) model. The attachment of the insulator causes a loop configuration that produces defined chromatin domains that bring some elements into proximity, allowing activation of some genes, or isolating others, causing repression (modified from Valenzuela et al., 2006).

1.4 Insulators identified in non-plant systems (see Table 2 for summary)

1.4.1 The *Gypsy* insulator from *Drosophila melanogaster*

The *gypsy* insulator is a 350bp DNA sequence derived from the *gypsy* retrotransposon in *Drosophila melanogaster*. Geyer et al. (1986), determined that this element is able to effectively block the action of enhancers distal to the *yellow* locus without affecting any of the enhancers more proximal to the promoter. The essential protein, the suppressor of hairy wing Su(Hw), comprises of twelve direct repeats for the binding of the zinc-finger protein. Su(Hw) further recruits factors such as Mod(mdg4)2.2, topoisomerase I-interacting RS protein (dTopors) and the centrosomal protein CP190 (**Figure 6**) (Pai et al, 2004, Gerasimova, 2000). These factors combine to form ‘insulator bodies’ which have been proposed to organize chromatin into loops to create distinct domains to separate the enhancer from its target promoter (Bushey et al., 2008). The *gypsy* insulator has the ability to impede enhancer-promoter activity, while also protecting the gene from position effects. In *Drosophila*, the Su(Hw) protein is necessary for chromatin loop formation and enhancer/repressor blockage (Gerasimova, 2000).

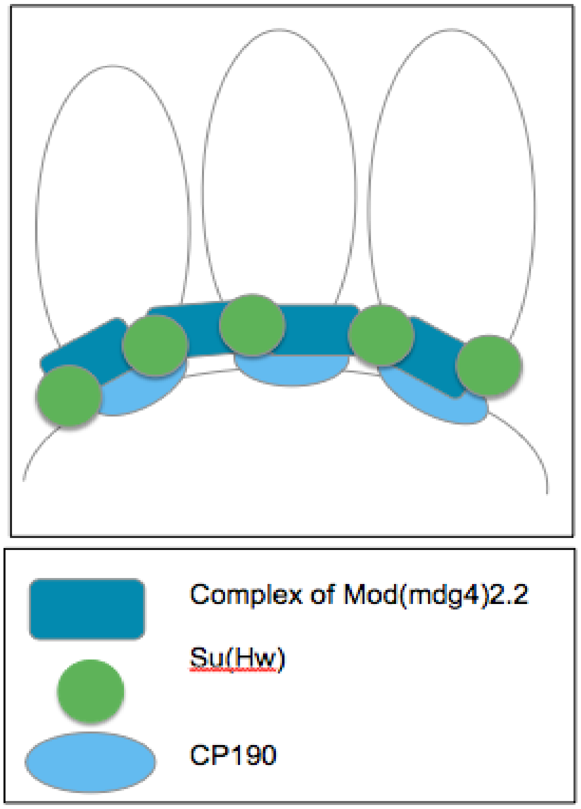


Figure 6. Schematic of insulator body formation of the *gypsy* insulator in *Drosophila*. The formation is due to the interaction of the main components Mod (mdg4)2.2, the Suppressor of Hairy wing Su(Hw) protein, a zinc-finger protein and the centrosomal protein CP190 (modified from Gaszner and Felsenfeld, 2006; Ong and Corces, 2009).

1.4.2 The Scs and Scs' insulators from *Drosophila melanogaster*

Scs and scs' (specialized chromatin structure) elements are 1.6kb DNA sequences that interact with the zinc finger Zeste-white 5 (Zw5) protein and the A and B isoforms of BEAF32 in *Drosophila melanogaster*. Kellum et al. (1992) determined its enhancer-blocking ability by using the yolk protein-1 enhancer element and an *hsp70* promoter:*lacZ* fusion gene driving β-galactosidase expression. They specifically flank the endogenous Hsp70 (heat-shock protein 70) locus at cytological position 87A7. This resulted in the insulator impeding activating and inactivating effects derived from the chromosomal environment that may have affected gene expression, in addition to blocking enhancer-promoter interaction (Kellum et al., 1992).

1.4.3 The 5’HS4 insulator from *Gallus gallus*

The 5’ constitutive hypersensitive site (5’HS4) insulator is predominantly found in vertebrates, located near the 5' end of the active chicken β–globin (Chung, 1993). It has the ability to block enhancer-promoter activity (Chung et al., 1997) as well as protecting the gene from position effects in *Drosophila* and in early–erythroid chicken cell lines (Pikaart et al., 1998). Within a 1.2kb DNA is the binding site for CTCF, a protein that initiates its enhancer-blocking abilities as seen in other species (**Table 1**). Furthermore they tested its enhancer-blocking activity using a γ-globin promoter/γ–neo reporter gene driving expression of a strong β–globin LCR element (Chung, 1993). This 1.2kb DNA has enhancer-blocking activity, in addition it also demonstrates barrier activity, which is independent of the CTCF factor and instead involves proteins that initiate 5’HS4-mediated acetylation and methylation of nucleosomes, to protect against heterochromatinization (Recillas-Targa et al., 2002).

It is generally assumed that an insulator can function as a barrier element and vice versa. This assumption was tested by Chung et al., (1997) who made deletions of the 1.2kb 5’HS4 DNA and tested for both activities. Their analysis concluded that the insulator activity appears to be distributed throughout the fragment, but about half resides in a 250bp core. Two copies of this core were sufficient to give insulator activity equivalent to that of the entire 1.2kb fragment (Chung, 1997). The same two copies of the 250bp core insulator were able to function as a barrier element and the reporter was protected from silencing. Further analysis demonstrated that the two activities could be separated by measuring their function using clones that had deletions of the five different core sites (Recillas-Targa et al., 2002). Their results demonstrated that deletion of all the elements except footprint 2 abolished protection against silencing. Furthermore, footprint 2 was concluded to be the single binding site within the 250bp core insulator element that displayed enhancer-blocking activity, which binds the CTCF protein. We can conclude that both activities appear to function within the larger “insulator sequence” but independently from each other (Recillas-Targa et al., 2002). This result is not generally applicable as detailed sequence/function analyses are rarely done. Clearly they should be.

Organism	Insulator Name	*Drosophila melanogaster*		*Saccharomyces cerevisiae*		*Schizosaccharomyces pombe*		Murine Cells *Mus musculus*		Human Cells *Homo sapiens*	
		EBA	BA	EBA	BA	EBA	BA	EBA	BA	EBA	BA
Drosophila melanogaster	*gypsy* retrotransposon	Geyer, 1986			Donze, 2001						
Drosophila melanogaster	*scs (special chromatin structures)*		Kellum, 1991								
Drosophila melanogaster	*scs' (special chromatin structures)*		Kellum, 1991								
Drosophila melanogaster	Fab-7 (Front abdominal-7)	Hagstrom, 1996									
Drosophila melanogaster	Fab8 (Front abdominal-8)	Barges, 2000									
Drosophila melanogaster	SF1	Majumder, 2003									
Drosophila melanogaster	1A2	Chetverina, 2012									
A. gossypii & S. cerevisiae	UASrpg				Bi, 1999						
Saccharomyces cerevisiae	STARs (subtelomeric anti-silencing regions)				Fourel, 1999						
Saccharomyces cerevisiae	*CHA1*				Donze, 2001						
Saccharomyces cerevisiae	*tDNA*				Donze, 1999						
Saccharomyces cerevisiae	*HMR* tRNA				Donze, 2001						
Mus musculus	SINE B2							Lunyak, 2007			
Mus musculus	SINE B1							(Roman, 2011)			
Mus musculus	*tRNA* genes								Ebersole, 2011		
Homo sapiens	BEAD-1 (blocking element alpha/delta)									Zhong, 1997	
Homo sapiens	tDNAs						Raab, 2012	Raab, 2012		Raab, 2012	
Gallus gallus	5'HS4 (5' constitutive hypersensitive site)		Chung, 1993							Chung, 1993	Ghirlando, 2012

Table 2. Summary of identified insulators in non-plant systems. This table represents various insulators originating from different organisms that have been tested positively for insulator activity. The reference has been filled in for the specific activity the insulator has been proven to function as: EBA – Enhancer Blocking Activity; BA – Barrier Activity.

Organism	Insulator Name	*Arabidopsis thaliana*		*Nicotiana tabacum*		*Oryza sativa*		*Brassica juncea*		*Theobroma cacao*		*Populus tremula*	
		EBA	BA	EBA	BA	EBA	BA	EBA	BA	EBA	BA	EBA	BA
Drosophila melanogaster	*gypsy* retrotransposon		She, 2010										
Drosophila melanogaster	Fab-7 (Front abdominal-7)	Gudynaite-Savitch,1999											
A. gossypii & S. cervisiae	UASrpg	Gudynaite-Savitch,1999											
Saccharomyces cerevisiae	ARS-1 SAR/MAR				Allan, 1993								
Proteus vulgaris	*β-phaseolin* coding and genomic			van der Geest, 1997									
Proteus vulgaris	*β-phaseolin* 3' MAR & 5' MAR			van der Geest, 1997									
Homo sapiens	BEAD-1 (blocking element alpha/delta)	Gudynaite-Savitch,1999											
Homo sapiens	BEAD-1C (blocking element alpha/delta)	Gudynaite-Savitch,1999											
Bacteriophage Lambda	λ (*EXOB*)	Singer, 2010											
Petunia hybrida	*TBS* (transformation booster sequence)	Hily, 2009											
Arabidopsis thaliana	*Atgypsy*-like element (*Atgypsy*-like)	Singer, 2013											
Oryza sativa (Rice)	*HS185*						Zhang, 2012						
Nicotiana tabacum	*Rb7* 3' MAR				Allen, 1996		Vain, 1999				Maximova , 2003		Han, 1997
Nicotiana tabacum	TM2 MAR						Xue, 2005						
Glycine max	P1-SAR/MAR				Breyne , 1992								
Glycine max	*Gmhsp* 17.6L MAR				Schoff, 1993		Oh, 2005						

Table 3. Summary of identified insulators in plant systems. This table represents various insulators originating from different organisms that have been tested positively for insulator activity. The reference has been filled in for the specific activity the insulator has been proven to function as: EBA – Enhancer Blocking Activity; BA – Barrier Activity.

1.5 Functional Plant Insulators (see Table 3 for summary)

1.5.1 Insulators from other species that function in plants

An increased number of insulators originating from a variety of species have been tested in plants, mainly *Arabidopsis*, with several demonstrating insulator activity. This finding suggests that components involved in insulator activity in these organisms are evolutionarily conserved. To date none have been tested for barrier activity. Thus we can propose that heterologous insulators reduce misexpression through conserved processes and that there is conservation of the insulator sequence and binding protein. However we cannot exclude the possibility that a system unique to plants may exist. Elements that are very successful at reducing misexpression may play a role in the recovery of transgenic plants with stable expression of desired genes (Gudynaite-Savitch et al., 2009).

1.5.1.1 The BEAD-1 and BEAD-1C insulators from *Homo sapiens*

BEAD-1 and BEAD-1C (blocking element alpha/delta) is a 1.6kb enhancer-blocking element with CTCF binding sites from the human T cell receptor (TCR) α / δ locus (Zhong, 1997). T lymphocytes express either an α/β or a γ/δ T cell receptor (TCR) heterodimer that is critical for T cell development and function. The genes encoding the four TCR proteins consist of multiple variable (V), diversity (D), and joining (J) gene segments that are assembled by the process of VDJ recombination during T cell development in the thymus. They determined insulator activity by using TCR δ and J_α gene segments, such that it would prevent the TCR δ enhancer (E_δ) from opening J segments for VDJ recombination during the early stage of T cell development. In their experiments BEAD-1 was capable of blocking the activity of the TCR δ enhancer in the human TCR α / δ locus. In addition, this insulator reduced nonspecific enhancer-promoter interactions in *Arabidopsis* blocking interaction between the *35S* enhancer and the napin promoter driving *GUS* gene expression (Gudynaite-Savitch, 2009).

1.5.1.2 The UASrpg insulator from *Ashbya gossypii*

A 149bp fragment, UASrpg (upstream activation site for ribosomal protein genes) from *Ashbya gossypii,* was tested for barrier activity in yeast by monitoring the ability to block the spread of heterochromatin (Bi & Broach, 1999). Insulator activity was defined when UASrpg was inserted between the E silencer and the α-mating genes in HML∆I. They tested the effects of silencing on the α-mating gene fragments of UASrpg, which resulted in the blocking of heterochromatinization, thus confirming barrier activity. This region of yeast DNA contains 3 Rap1 sites, while the homologous sequence from *A. gossypii* contains 2 Rap1 sites. Interestingly UASrpg from *A. gossypii* also significantly reduced nonspecific enhancer-promoter interactions in *Arabidopsis*, blocking the interaction between the *35S* enhancer and the napin promoter driving *GUS* gene expression.

1.5.1.3 The *Gypsy* insulator from *Drosophila melanogaster*

She et. al (2010), introduced the *gypsy* insulator (**section 1.4.1**) and its binding protein Su(Hw) into *Arabidopsis* to look at expression levels and to determine if it is functional in a plant system. The *gypsy* insulator has been extensively studied and is known to regulate gene expression by blocking interaction between enhancer and promoter, causing an insulating affect within the gene. Therefore, research was conducted by inserting the identical *gypsy* insulator sequence from *Drosophila* into *Arabidopsis* using a vector consisting of a fused EGFP-G*US* reporter gene and the *HPT* driven by the *nopaline synthase promoter (NOS)* as its selectable marker. Since *Arabidopsis* does not have a homolog for Su(Hw), a clone expressing Su(Hw) was also introduced. This resulted in improved expression levels of transgenes, the reduction of position effects and improved specificity of promoter activity. Coexpression of Su(Hw) increased specificity of transgene expression. Furthermore, Su(Hw) was detected in the nucleus in *Arabidopsis*, which is identical to that observed in *Drosophila*.

1.5.1.4 The EXOB insulator from bacteriophage lambda

Singer et al., (2010) tested fragments of bacteriophage λ for insulator activity. The insertion of the 1-kb EXOB DNA between the *CaMV 35S* promoter/enhancer and a flower specific *AGAMOUS* second-intron-derived promoter (*AGIP)* fused to *GUS* resulted in diminished *GUS* expression in transgenic leaves (Singer et al., 2012), while neither the 2kb nor 4kb λ fragments could do so. Since there is no evidence for insulators in *E. coli*, the usual host for λ, this suggests that random cloned DNA may contain a sequence similar to a functional insulator.

1.5.2 Insulators isolated from plants

Similarities between animal and plant cells leads us to believe that plants may have DNA sequences that work as barriers or enhancer blocking elements or both. This notion is supported by the functional expression in plants of heterologous insulators. Therefore, studies have been conducted to find DNA sequences that function to block non-specific enhancer-promoter interaction in transgenic plants. To date many insulators from a variety of species has been tested in plants and do function as insulators. This finding suggests that components involved in insulator activity in these organisms are evolutionarily conserved between plants and that finding an insulator capable of working with different enhancer/reporter systems in *Arabidopsis* can lead us to insulators that may work in different plant systems. This will directly impact the way we work with genetic selection of desirable traits in plants.

1.5.2.1 The Transformation booster sequence (*TBS*) insulator from *Petunia hybrida*

The *TBS* MAR sequence from *Petunia hybrida* possesses enhancer-blocking activity in *Arabidopsis thaliana*. Matrix attachment regions (MARs) are characterized by their ability to bind a network of nonhistone proteins in the nuclear matrix. Their proposed function is to organize chromatin into loop domains as has been proposed for insulators, blocking gene expression. Many studies have shown that flanking a transgene with MAR sequences increases the level of transgene expression and/or reduces plant-to-plant variations in transgene expression. For example Singer et al. (2011) using the *AGIP* promoter-*GUS* gene test system described above demonstrated that approximately 70% of transgenic lines in which the *TBS* element was introduced exhibited blocking of *35S* enhancer-mediated activation in *Arabidopsis*. The full length *TBS* fragment has also been found to be effective for impeding enhancer-promoter interference in transgenic tobacco plants and is not promoter specific, indicating that it may function as an insulator in other plant species (Singer, 2011).

One of the most well-studied class of putative insulator elements with potential applications in plant transgenic technology are MARs, which have been suggested to trigger the formation of chromatin loops. MARs are described as non-transcribed, AT-rich sequences that bind a nuclear network of non-histone proteins. MARs have been shown to reduce chromosomal position effects, which result in an increase in the level of transgene expression and or a reduction in plant-to-plant variability (Singer, 2012). An analysis of the enhancer-blocking capabilities of three MAR-containing sequences in transgenic *Arabidopsis* was carried out. The results indicated that the MAR from *Petunia TBS* fragment, but not the maize or tobacco MARs, is able to impede *35S* enhancer activation of the *AGIP* without compromising the performance of either, with elements from which they are not insulated (Hily, 2009). Since MARs have been shown to function as both enhancer blocking and barrier elements, we have yet to know whether the same sequence can act as both. I would like to point out that it seems that MARs act as enhancer-blocking elements within plant systems and barrier elements within non-plant systems as seen in **Table 2** and **3**. The mechanism behind MARs activity is still unknown, but is interesting to see the increasing number of MARs sequences that act as insulators in both plant and non-plant systems. Hence, further analysis needs to be done on these elements to identify their function.

1.5.2.2 The HS185 insulator from *Oryza sativa*

Zhang et al. (2012) performed GUS analysis on the *HS185* sequence that is 1.8kb in length, which originated from deletion analysis of *HSP70* promoter in rice. Its insulator activity was defined through two constructs. They observed significantly low GUS activity when they inserted *HS185* between the *CaMV 35S* enhancer and the *HSP70* minimal promoter. Furthermore, they tested insulator activity through the insertion between the *CaMV 35S* double enhancer and the *CaMV 35S* minimal promoter driving *GUS* gene expression. *HS185* inhibited the enhancer interaction with the *CaMV 35S* minimal promoter, which directly implicated *HS185* as an insulator in rice.

1.5.2.3 The *Atgypsy*-like element insulator from *Arabidopsis thaliana*

As indicated in **section 1.5.1.4** the *gypsy* insulator and its binding protein Su(Hw) decrease transgene variability within *Drosophila* and *Arabidopsis* (She et al., 2002). Similarly Singer et al. (2012), tested a 2.2kb gypsy-like element from *A. thaliana* (*Atgypsy*-like element) to determine if it functions as a true enhancer-blocking element, as gene misexpression in transgenic plants were significantly reduced when *Atgypsy* was inserted between the *CaMV 35S* enhancer and petal- and stamen-specific *PISTILLATA* promoter (*PIp*) driving *GUS* gene expression. Therefore, the *Atgypsy*-like insulator was effective in reducing transgene variations without the Su(Hw) protein and implies that other factors in *Arabidopsis* are compensating for the absence of the Su(Hw) protein as there does not appear to be a Su(Hw) homologue in *Arabidopsis*. However there is no direct evidence that the insulator sequence within the 2.2kb fragment is the one inferred from studies of the *Drosophila gypsy* element. There are multiple repeats in the *Atgypsy*-like sequence that could act as binding sites for other proteins that facilitate enhancer-blocking activity (Singer et al., 2012).

1.6 Hypothesis and Objectives

Many insulators and barrier elements have been described in a wide range of species including some in plants (**Tables 2 and 3**). This study aims to identify functional insulators defined as DNA sequences that prevent non-specific interaction between enhancers and promoters in *Arabidopsis*.

Our hypothesis is that the production and subsequent screening of a random oligonucleotide library will identify enhancer-blocking insulators. To test this hypothesis we will:

i. Analyse 3 putative sequences (CLOI-3, CLOII-12, and CLOII-12) cloned into pB31 from a random oligonucleotide library produced in our lab using *Arabidopsis*.
ii. Determine enhancer-blocking activity of the candidate sequences in **i.** through GUS staining, PCR and sequence analysis.
iii. From the candidate sequences, reproducibility identified by screening will use bioinformatics tools to verify the presence and functionality of the plasmid vector components.

Our long-term goal is to determine whether these insulators can function in organisms such as yeast and tobacco so that we can identify universal insulators. Thus, this study aims to identify, isolate and characterize DNA insulators possessing enhancer-blocking activity in the model plant system *Arabidopsis thaliana.*

2.0 Materials and Methods

2.1 Vector formation: pC1, pB31, and pL1

All vectors were constructed using the pCAMBIA series of plasmids described at http://www.cambia.org/daisy/cambia/materials/vectors.html. The vectors used in the DNA insulator project as a whole are: pC1, pB31 and pL1, constructed by Dr. L. Gudynaite-Savitch. pC1 was used first to select for potential insulator sequences which were then sublconed into pB31 for further screening (**Figure 7**).

2.2 Generation of the random oligonucleotide library

This project was initiated through the production of a random oligonucleotide library. Fragments consisting of an 18 base PCR primer binding sequence (including restriction sites) linked to 124 random bases linked to an 18 base PCR primer binding sequence (including different restriction sites) were synthesized at McMaster University (5'-AGTGGATCCGAGACAAGC (N124) CCTCCTCCTGAATTCTGC 3') and cloned into pC1 for selection as described in Results.

2.3 Molecular techniques for selection and screening of the oligonucleotide library

2.3.1 Polymerase chain reaction (PCR) utilized for cloning

All PCR reactions were done in 20μL reaction mixtures containing 1x PCR buffer (final concentrations of 20 mM Tris-HCl pH 8.75, 10 mM KCl, 10mM $(NH_4)_2SO_4$), 2 mM $MgSO_4$, 200μM of each deoxynucleotide triphosphate and 0.2μM of each specific primer. With the addition of 1μL of DNA and 1 unit of Taq polymerase (UBI Calgary), the PCR product was amplified using the following cycle parameters: (1) 95^o C denaturation for 5 minutes; (2) 94^o C for 30 seconds; (3) 55^oC annealing for 30 seconds; (4) 72^o extension for 30 seconds to 1 minute. Steps 2 – 4 were repeated 31 times. A final extension step at 72°C for 5 minutes was performed to ensure the extension of the PCR product. PCR products were then run on 1.25% agarose gel.

2.3.2 Bacterial transformation utilized for cloning

Plasmid constructs with a DNA fragment of interest were transformed using *E. coli* DH5α (Life Technologies Inc. Burlington, Canada) according to the manufacturer's protocol. *E. coli* transformants were selected on 100μg/mL ampicillin or 50μg/mL kanamycin depending on the plasmid construct being used.

Plasmid constructs with a DNA fragment of interest were also transformed into *Agrobacterium tumefaciens* strain GV3101 (pMP90) (Koncz and Schell 1986) by electroporation (Wang *et al*. 2006). Transformants were selected on 50μg/mL kanamycin and 100μg/mL rifampicin.

2.3.3 Plasmid preparation utilized for the sequencing of clones

The GeneElute Plasmid Mini-Prep Kit (Sigma-Aldrich) was used following the manufacturers instructions to isolate DNA for cloning and sequencing.

2.3.4 General cloning techniques

Restriction and ligation for fragment cloning, and cloning into pGEM-t (Promega Corporation) followed the manufacturers' recommendations.

2.4 Plant molecular techniques for cloning

2.4.1 *Arabidopsis* genomic DNA mini-preparation

DNA extracted from transgenic plants for PCR and/or cloning was isolated using a small-scale procedure (McKinney et al., 1995, with minor modifications). Leaf material was stored at -20^{o}C prior to extraction. A cauline leaf (approximately 0.5 cm^2) of the genotype to be examined was placed in a sterile microfuge tube. The plant tissue was ground with a sterile plastic pestle. An aliquot of 0.5mL of DNA extraction buffer (100mM Tris-HCl pH 8.0, 250mM NaCl, 10mM EDTA, 0.5% SDS) was added to the tube and the tissue sample was ground a second time with the pestle. The tube was placed on ice to homogenize for 2 minutes. The samples were microfuged for 2 minutes at 13000 rpm. To recover the DNA, 0.3mL of supernatant was transferred to tubes containing 0.3mL 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 0.75mL 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 0.1mL of TE mix (10mM Tris, 1mM EDTA, pH8.0) and were mixed by vortexing. The mix was spun using the microfuge for 2 minutes and 0.04mL of the supernatant was transferred into a sterile, labeled microfuge tube. These DNA preparations were stored in a 4^{o}C unit for storage and use.

2.4.2 PCR screening of transgenic plants utilized for DNA and insert verification

The extracted DNA from each transgenic *A. thaliana* sample was tested with two rounds of PCR. The first round 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 762bp; the presence of this sequence confirmed successful DNA extraction. The second round used 1300LaczF/GUS-Rev primers to amplify the vector including the insert to confirm successful cloning of the sequence of interest, with an expected size of 600bp or 900bp (**Figure 15**). PCR conditions as indicated above were used for all reactions. All PCR products were analyzed by gel electrophoresis, using 1.25% agarose gels. All primers used in this thesis are summarized in **Table 4.**

2.4.3 Plant transformation utilized for transgenic seed production

Arabidopsis thaliana (Columbia ecotype) 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 30µg/ml *hygromycin-B* and 250µg/ml *timentin*; (h*ygromycin-B* for selection of plants carrying the plasmid vectors, and *timentin* to inhibit bacterial growth). Seeds were grown on selective media for approximately 2 weeks in a growth cabinet until visible roots and leaves formed. Transgenic plants were selected and then transferred to soil for further growth.

2.4.4 GUS staining of transgenic plants

Transgenic plants that were transferred from selective media to soil are grown in a growth cabinet for an additional 10 days. Tissue is then collected from each transgenic plant into a 1.5ml microtube when there are visible siliques, flowers and leaves. Each sample contained a minimum of three tissues from a single plant, i.e. 3 siliques (2 samples, one week apart), flowers, or leaves. In addition two cauline leaves were collected for molecular analysis of insert and plasmid presence were collected and stored at -20^{o}C.

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) (see Supplementary Figure S1 of Gudynaite-Savitch et al., 2007).

2.5 Plant material and growth conditions

2.5.1 Plant growth

Arabidopsis wild-type seeds (Columbia) were initially grown on soil (Premier, Pro-Mix PGX) that was rehydrated with water and a small amount (5g/L) of MiracleGro. The soil was compacted into individual pots and covered with a plastic casing for 2 days. They were then grown under 100 µmol/m^2/s photosynthetically active radiation (PAR) day/8h at 22^{o}C. The plants were watered consistently every 2-3 days using Hoaglands nutrient solution (Gudynaite-Savitch *et al*. 2007), mixture of fertilizer and tap water and, regular tap water once a week.

2.5.2 Seed sterilization

Seed sterilization was performed on all seeds prior to planting. The desired number of seeds was transferred into a 15mL conical tube. Approximately 10mL of tap water was added to the tube and 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 10mL of 0.1% agarose is added to the tube and mixed, followed by the application of seeds using a Pasteur pipette onto premixed soil.

2.6 Bioinformatics and sequence analysis

All DNA sequences were analysed using DNAMAN software (Lynnon Corporation). The software was used to determine and match sequences, and analyze restriction sites.

Searches for potential insulator or regulatory sequences were made using programs available on the WEB at sites indicated in the text.

Motif sequences were searched on the JASPER CORE (http://jaspar.genereg.net/cgi-bin/jaspar_db.pl?rm=browse&db=core&tax_group=plants) plants database online.

ActivePerl software (ActiveState Software Inc.) was used to determine the presence and location of motifs and consensus sequences within the potential insulator sequences that were being analyzed.

Repressor sites were located within sequences using a specialized database of plant cis-acting regulatory DNA elements called, PLACE (http://www.dna.affrc.go.jp/PLACE/) and PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

Primer	Target of Amplification	Sequence
SALK_049131_RP2	Genomic AtDNA, positive control	5’-GTCTCTACCGTACGCGCTTC-3’ (20mer)
SALK_049131_LP2	Genomic AtDNA, positive control	5’-GGTTTGCATTTGACCTTTCG-3’ (20mer)
1300LacZ-F	CLO DNA transgene	5' CACTCATTAGGCACCCCAGG3’ (20mer)
GUS-Rev	GUS gene in transgene	5'-GTGGCTAGCTTGTTTGCCTC-3' (20mer)
M13For	CLO DNA transgene	5’-CGCCAGGGTTTTCCCAGTCACGAC-3’ (24mer)
M13Rev	CLO DNA transgene	5'-AGCGGATAACAATTTCACACAGGA-3’ (24mer)
CLO-F-BPX	CLO DNA transgene insert	5'-CTTAAGCTTCGCAAGACCCTTCCTCTATATAAG-3' (31mer)
CLO-R-ESX	CLO DNA transgene insert	5'-AGTTCTAGAGTCGACGAATTCAGGAGGAGG-3' (30mer)

Table 4. List of all primers used in this thesis with their target of amplification and sequence.

3.0 Results

3.1 Summary of transformation vectors described in this study

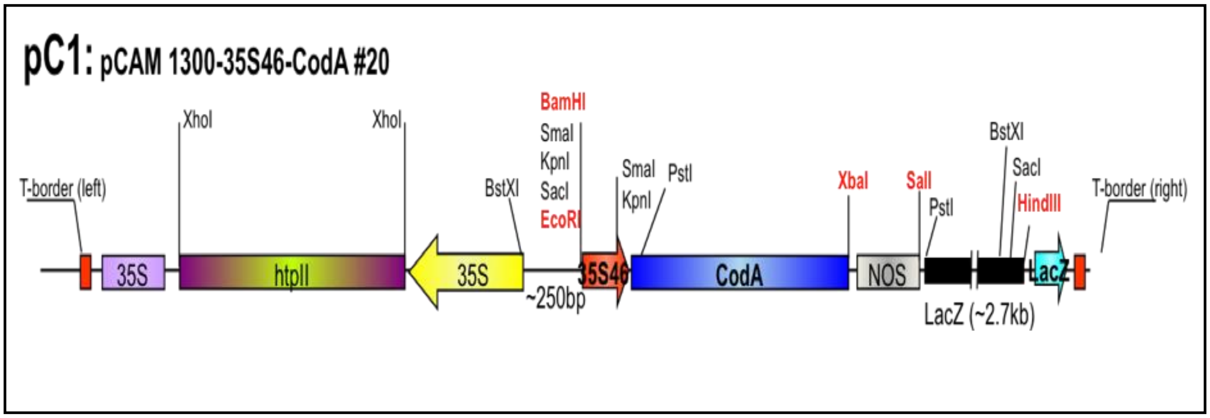
The transformation vectors used for the selection and screening of sequences from the random oligonucleotide library are pC1, pB31 and pL1 (**Figure 7**), all derived from common pCAMBIA vectors (http://www.cambia.org/daisy/cambia/585.html). These vectors were developed by Dr. Gudynaite-Savitch, a former post-doctoral fellow in our lab. The use of these vectors will allow us to isolate and partially characterize novel plant insulators. This project was initiated through the synthesis of DNA fragments consisting of random sequences and flanking regions to provide restriction sites and sites for PCR amplification.

The DNA fragments were initially ligated into pC1 (pCAM1300-35S46-*CodA*) (**Figure 7A**); to generate a random oligonucleotide library in which putative insulator sequences are cloned between the *CaMV 35S* enhancer/promoter and *CaMV 35S* core promoter (Ott et al., 1990), which drives the expression of *codA*. pC1 (**Figure 7A**) is a negative selection vector, with T-DNA borders flanking a gene encoding antibiotic resistance (*hygromycin*), driven by a full *CaMV 35S* enhancer/promoter system for selection, and a conditional negative selective marker gene, *codA* (cytosine deaminase), driven by the core *CaMV 35S* promoter (Ott et al., 1990). *Agrobacterium tumefaciens GV3101* was transformed with pC1 vectors carrying putative insulator sequences. A functional *codA* gene converts 5-FC (5-Fluorocytosine) on selective plates into 5-FU (5-Fluorouracil), which is toxic leading to plant death. In plants and mammals, enzymatic conversion of cytosine to uracil is absent; therefore, 5-FC does not affect the growth of wild-type plants at the levels used, while transgenic plants expressing cytosine deaminase are negatively selected on 5-FC plates (see **Appendix Figure 1**). Growth on selective plates may be due to an insulator but there are other possibilities as well including mutagenesis of the *codA* gene, integration into heterochromatin and the cloning of a DNA element with strong repressor binding activity that could overcome activation by 35S enhancers.

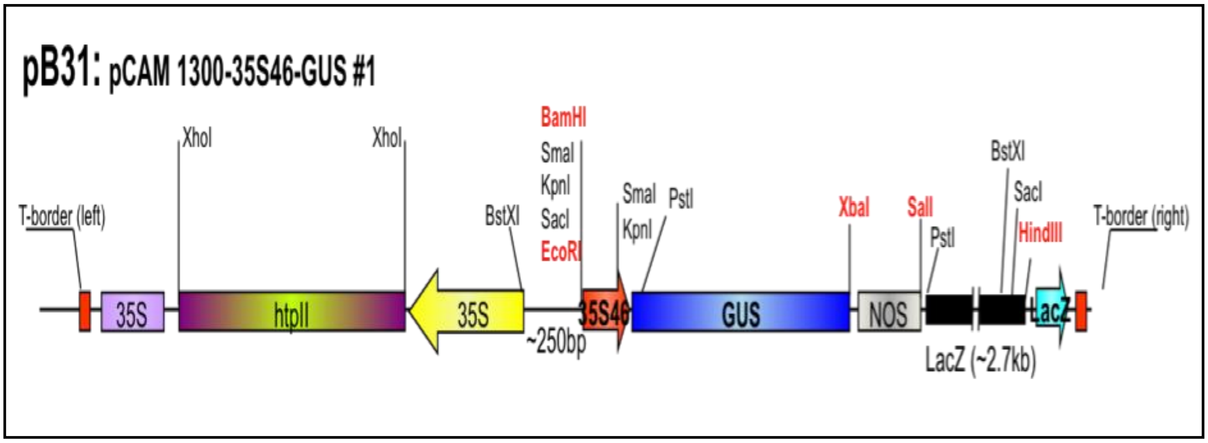
pB31 (pCAM1300-35S46-*GUS*) is a similar vector used for GUS screening. Putative insulator sequences are cloned between the *CaMV 35S* enhancer/promoter and *CaMV 35S* core promoter (Ott et al., 1990), which drives the expression of the *GUS* gene as shown in **Figure 7B.** Expression levels were measured by screening for GUS enzyme activity in different plant tissues (flowers, leaves, siliques) of transgenic *Arabidopsis* plants. All sequences analyzed in this thesis were cloned into pB31 using the *Eco*RI and *Bam*HI restriction enzymes.

Lastly, pL1 (pCAM1391-Napin) is a screening plasmid vector, which has a *CaMV 35S* enhancer/promoter and napin (seed-specific) promoter that drives expression of the *GUS* gene as shown in **Figure 7C**. The pL1 vector was not used for screening of putative insulators in this thesis, but should be used for further analysis of putative sequences as described in the Discussion.

A


pC1: pCAM 1300-35S46-CodA #20
XhoI
XhoI
BamHI
SmaI
KpnI
SacI
EcoRI
BstXI
SmaI
KpnI
PstI
XbaI
SalI
PstI
BstXI
SacI
HindIII
T-border (left)
T-border (right)
35S
htpII
35S
~250bp
35S46
CodA
NOS
LacZ
LacZ (~2.7kb)

B


pB31: pCAM 1300-35S46-GUS #1
XhoI
XhoI
BamHI
SmaI
KpnI
SacI
EcoRI
BstXI
SmaI
KpnI
PstI
XbaI
SalI
PstI
BstXI
SacI
HindIII
T-border (left)
T-border (right)
35S
htpII
35S
~250bp
35S46
GUS
NOS
LacZ
LacZ (~2.7kb)

C

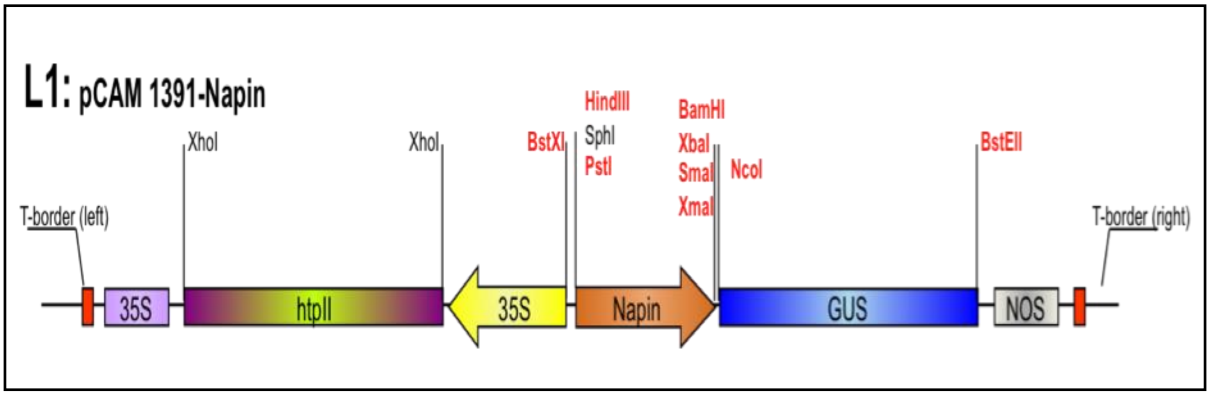

L1: pCAM 1391-Napin
XhoI
XhoI
BstXI
HindIII
SphI
PstI
BamHI
XbaI
SmaI
XmaI
NcoI
BstEII
T-border (left)
T-border (right)
35S
htpII
35S
Napin
GUS
NOS

Figure 7. Summary of the plasmid vectors used for selection and screening of the sequences in the random oligonucleotide library for the determination of insulator activity. (**A**) Plasmid vector pC1 (pCAM1300-35S46-*CodA*); used for the isolation of potential insulator sequences by negative selection using the *CodA* gene (blue box). Made by Loreta Gudynaite-Savitch. (**B**) Plasmid vector pB31 (pCAM1300-35S46-*GUS*); used to confirm the presence of a potential insulator sequence using the *GUS* gene for screening. Made by Loreta Gudynaite-Savitch. (**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. Made by Loreta 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 box) or napin seed-specific promoter (orange box).

3.1.1 Statement of contributions

All pCAMBIA plasmid vectors used in this study were previously constructed by Dr. L. Gudynaite-Savitch, and others. The random DNA oligonucleotide library for cloning into pC1 was synthesized at McMaster University. Fragments of 154bp in size consist of 124bp of random sequence flanked by sites for PCR primer that are 30bp long and contain *Bam*HI and *Eco*RI sites respectively for cloning. Thousands of seeds 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 ~450bp and the original 154bp sequences. Mapping all of the *Eco*RI and *Bam*HI restriction sites indicated that in most cases three different ~154bp sequences ligated to each other at their *Eco*RI and *Bam*HI flanking sites generating a long sequence consisting of ~450bp. See the MSc thesis of Batool Gandorah, 2012, **section 3.4 Figures 13 and 14** for an in depth explanation. These sequences were identified, sequenced, and classified into four groups (I, II, III, and IV; represented by Roman numerals, based on the selection trial completed). Twenty out of the sixty were chosen based on plant growth/health on selective media to be cloned into pB31 by Dr. Gudynaite-Savitch. Subsequently seventeen out of twenty of these potential insulator sequences were cloned into pCAMBIA pB31 (pCAM 1300-35S 46-*GUS*) by Dr. Gudynaite-Savitch. The three remaining insulator sequences were cloned by Batool Gandorah. The candidate insulator CLOI-3 was tested by Batool Gandorah through the completion of two transformation events that showed minimal GUS staining (**Figure 8**). Subsequently, PCR and sequence analysis for one transformation indicated CLOI-3 as a strong candidate insulator.

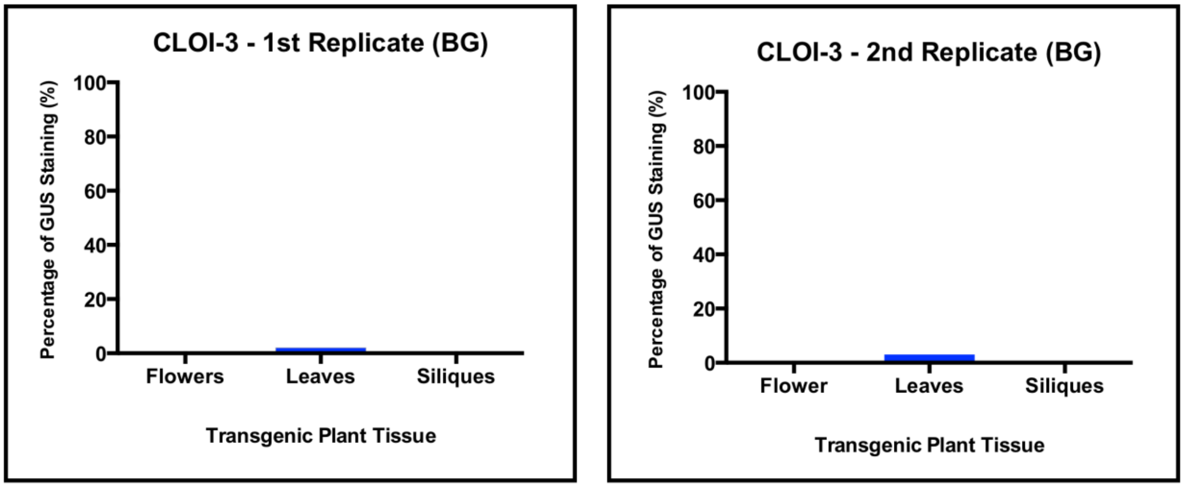


Figure 8. CLOI-3 (438bp) GUS staining analysis. CLOI-3 is a candidate sequence in which two individual screening events indicated no GUS staining in flower and silique tissue, and minimal GUS staining in the leaves. This sequence was cloned into vector pB31 using the *GUS* gene for screening. 1^{st} replicate N=24 transformed plants; 2^{nd} replicate N=25 transformed plants. Reproduced from the thesis of B. Gandorah with permission.

3.1.2 Defining insulator activity

In the experiments described in this thesis we initially defined a putative insulator as a sequence isolated from transgenic *Arabidopsis* that survived selection in pC1. Following successful cloning into pB31 and transformation, GUS staining of several plant tissues as described in Material and Methods **section 2.4.4** was used to confirm potential insulator activity (the lack of staining in any of the tissues). Scoring was stringent and if only a small region of any of the tissues sampled had GUS staining, it was scored as positive. The stained samples were scored independently by at least 2 individuals to reduce bias. At least ten *Arabidopsis* transformants were tested for each construct. In addition to this criterion, PCR analysis was performed to test for the presence of the insert and at least two PCR products were sequenced. Therefore lack of GUS staining, a positive PCR result for insert presence, and correct sequence analysis of a small number of clones allows us to confirm insulator activity and define putative insulators in our work. Finally two or more replicates, starting from *Arabidopsis* transformation need to be performed.

3.2 Identification of potential insulator sequences

From the twenty chosen sequences cloned into pB31, histochemical GUS staining produced three promising DNA sequences (CLO I-3, CLO II-10, and CLOII-12; see **Figures 9-14**) respectively. These sequences were identified as potential sequences that may possess insulator activity, as they appear to block the influence of the *35S* enhancer on the core *35S* promoter in pB31 as indicated by minimal or no GUS staining. DNA analysis was done on at least 10 samples per transformation using SALK primers which amplifies the Atchrom-1 region 11,078,981-11,078,219 with expected size of 762bp in the *A. thaliana* genome, to test the quality of DNA, and 1300LacF/GUS5'R primers which anneal to sites flanking the CLO insertion site which amplify inserts along with the flanking regions in pB31 for confirmation of insert (900bp) (**Figure 15**). In addition preliminary data obtained by Batool Gandorah (BG) indicated that these sequences may be putative insulators that require further analysis to verify function (Batool Gandorah, 2012). Replicates are labeled with initials to indicate the individual who completed a given transformation. 'LR' represents Lara Rasooli's replicates and 'BG' represents replicates done by the former Masters student Batool Gandorah. The replicates that are labeled LR are defined as a successful transformation that has minimal GUS staining, PCR confirmation, and sequence confirmation.

3.2.1 Analysis of CLO transformed *Arabidopsis thaliana*

Figure 9 displays the results from the transformation of the pB31 vector as a control to indicate the extent of GUS staining in the absence of an insert. Transformation with the pB31 vector lacking an insert indicates 19% GUS staining in flower samples, 67% GUS staining in leaf samples, and 64% in silique samples. The first replicate of CLOI-3 (438bp) displayed no GUS staining in flowers, leaves and siliques as seen in **Figure 10**. Previous data from two separate transformation events had shown that few tissues were GUS positive, but PCR and sequence data was only collected from the second replicate that was completed by BG (**Figure 8**) (Gandorah, 2012). The PCR analyses (**Figure 16)** and sequence confirmation (4/4 clones) indicated the presence of DNA of expected size (**Appendix Figure 3**) and confirmed that the sequence was CLOI-3. Lack of GUS staining for CLOI-3 and the sequence data confirmed the possible insulator activity of this sequence. The first replicate of CLOII-10, which is 154bp in length, displayed no GUS staining in flowers and leaves, shown in **Figure 11**. No data from siliques was collected for this experiment due to staining error, but PCR (**Appendix Figure 3**) and sequence confirmation is ongoing. Previous data from one transformation completed by BG also demonstrated minimal GUS staining but no PCR or sequence data was collected. Furthermore a second replicate of CLOII-10 showed no GUS staining in any of the three tissues stained (**Figure 12**). PCR data from this second replicate confirmed the presence of the DNA of expected size (**Appendix Figure 3**). Lack of staining for CLOII-10 and the PCR data for both replicates confirms the possible insulator activity of this sequence. **Figure 13** and **Figure 14** displays the staining patterns for CLOII-12 (427bp), which also displayed no GUS staining in flowers, leaves and siliques. CLOII-12 also had previous data from one transformation completed by BG that showed minimal staining but did not have PCR or sequence data. The two replicates with complete data include PCR results (**Appendix Figure 3**) and sequence confirmation (3/3 clones) for each replicate. Lack of staining for CLOII-12 and the sequence data confirms the possible insulator activity of this sequence.

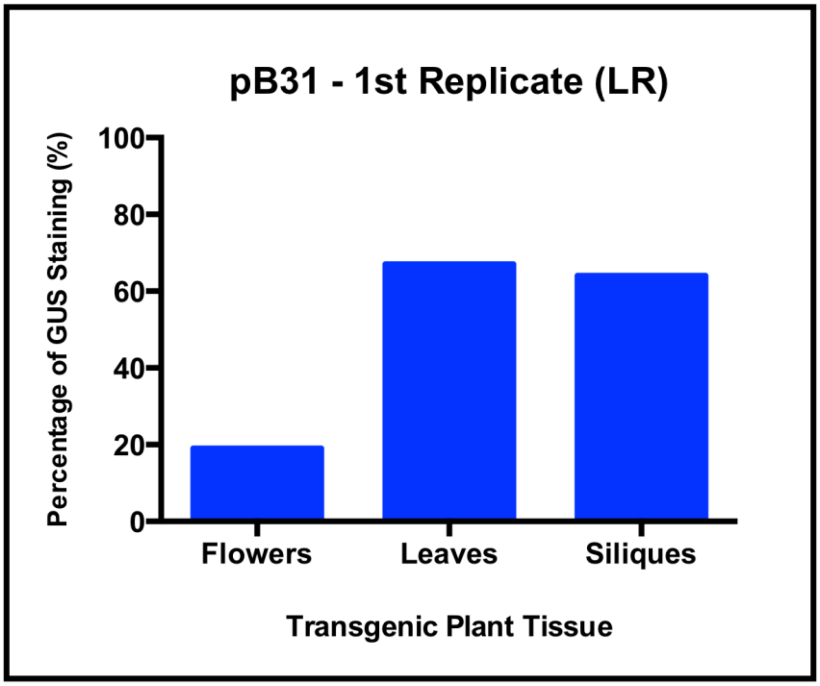


Figure 9. GUS staining of pB31 transformants. The vector pB31 was used as a control for the absence of an insert. pB31 displayed GUS staining in the flowers (19%), leaves (67%) and siliques (64%) in the first replicate of transformants. N=42 transformed plants.

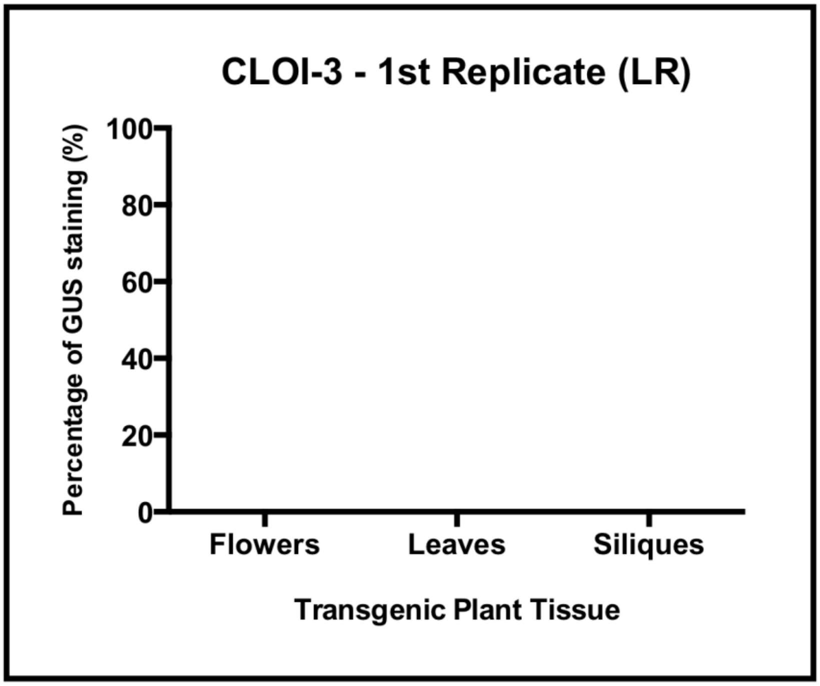


Figure 10. GUS staining of CLOI-3 (438bp) transformants in pB31. CLOI-3 is a sequence that was cloned into pB31 that displayed no GUS staining in the flowers, leaves and siliques in the first replicate of transformants. N=51 transformed plants.

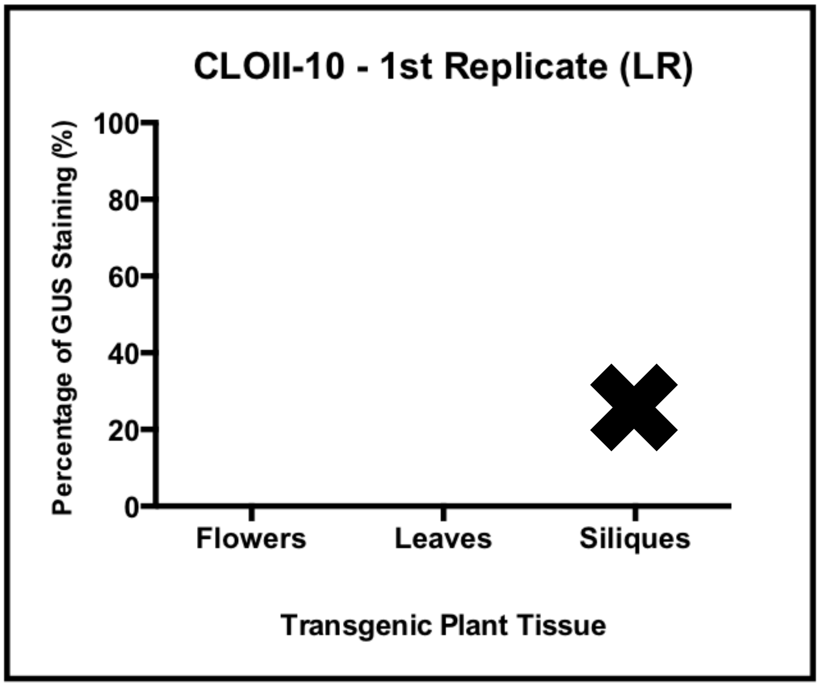


Figure 11. GUS staining of CLOII-10 (154bp) transformants in pB31. CLOII-10 is a sequence that was cloned into pB31 that displayed no GUS staining in the flowers and leaves in the first replicate of transformants. Silique staining was not done for this replicate (indicated by the **X**). N=14 transformed plants.

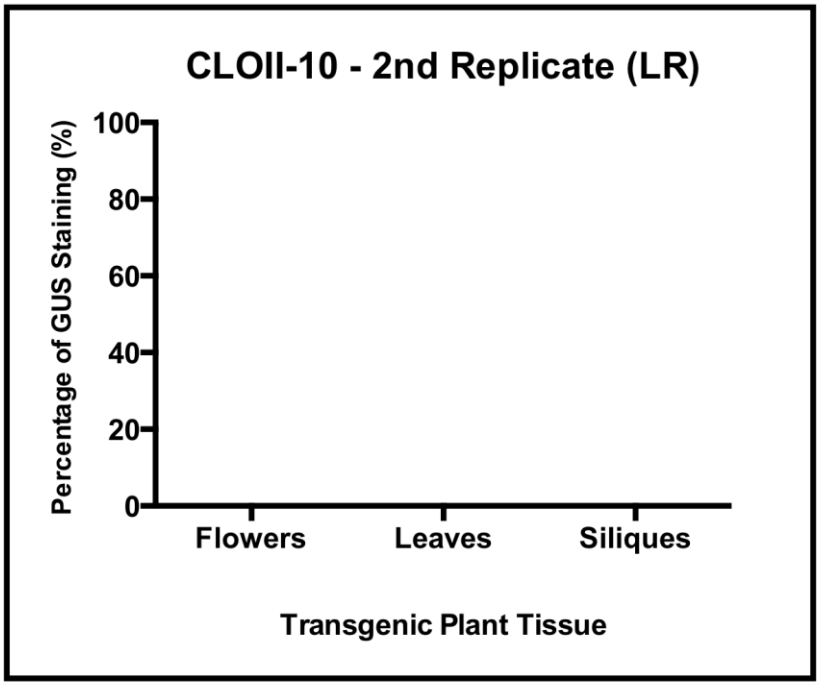


Figure 12. GUS staining of CLOII-10 (154bp) transformants in pB31. CLOII-10 is a sequence that was cloned into pB31 that displayed no GUS staining in the flowers, leaves and siliques in the second replicate of transformants. N=20 transformed plants.

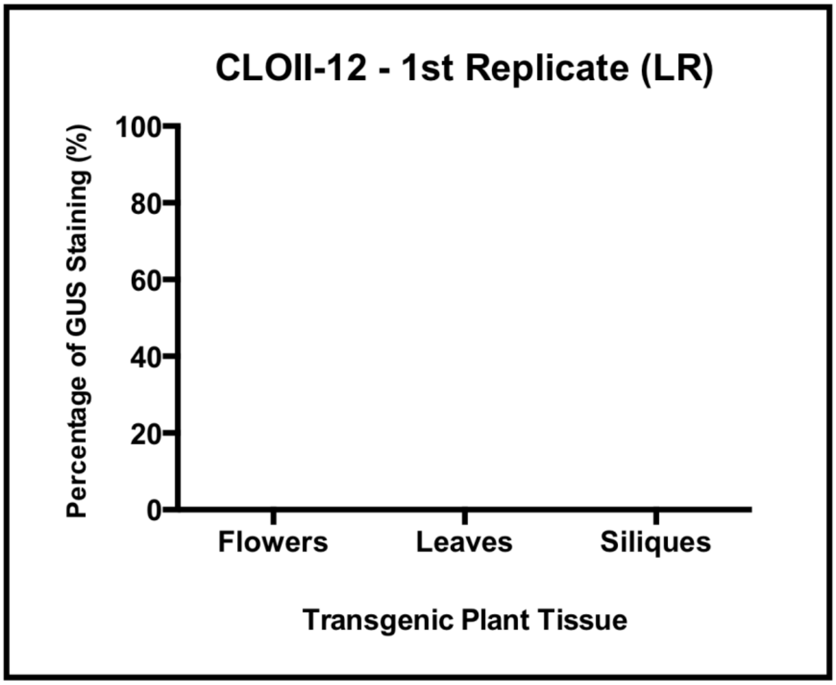


Figure 13. GUS staining of CLOII-12 (427bp) transformants in pB31. CLOII-12 is a sequence that was cloned into pB31 that displayed no GUS staining in the flowers, leaves and siliques in the first replicate of transformants. N=17 transformed plants.

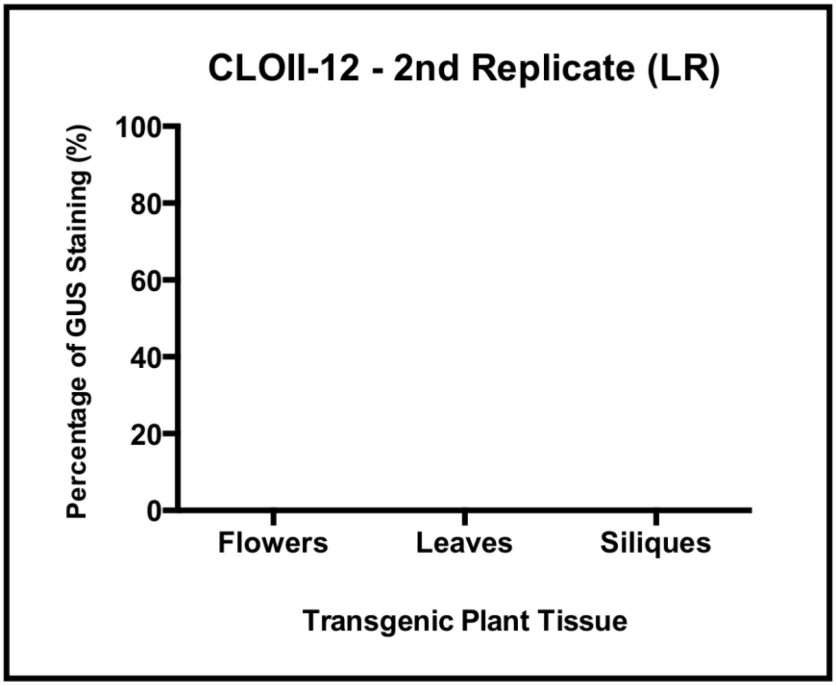


Figure 14. GUS staining of CLOII-12 (427bp) transformants in pB31. CLOII-12 is a sequence that was cloned into pB31 that displayed no GUS staining in the flowers, leaves and siliques in the second replicate of transformants. N=36 transformed plants.

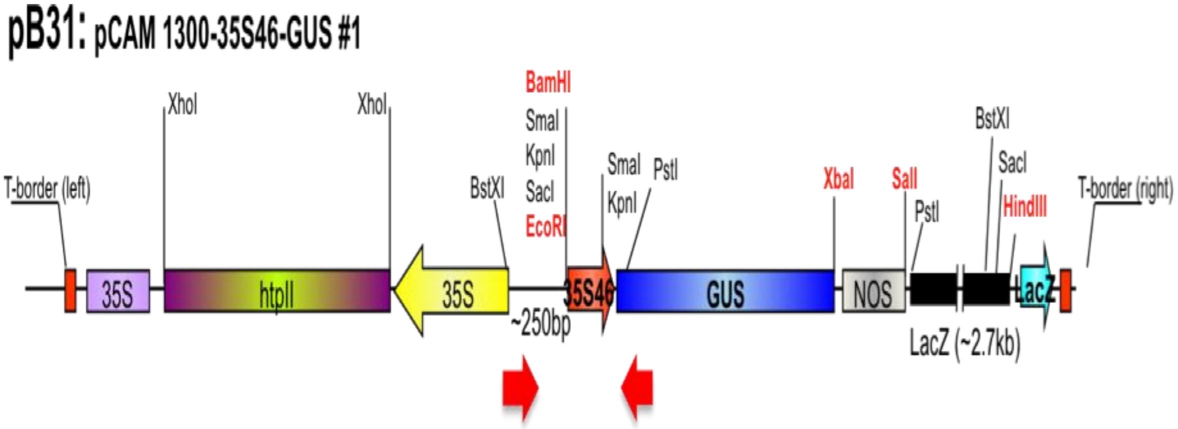


Figure 15. Plasmid vector pCAMBIA pB31 (pCAM 1300-35S 46-GUS); a screening plasmid vector, between *CaMV 35S* enhancer/promoter and *CaMV 35S* core promoter, which drives the expression of the *GUS* gene (Done by L. Gudynaite-Savitch). Red arrows indicate PCR amplification of the insulator to verify their presence using 1300LacZF/GUS5'R primers. SALK primers amplify the genomic DNA to test the quality of the DNA extracted.

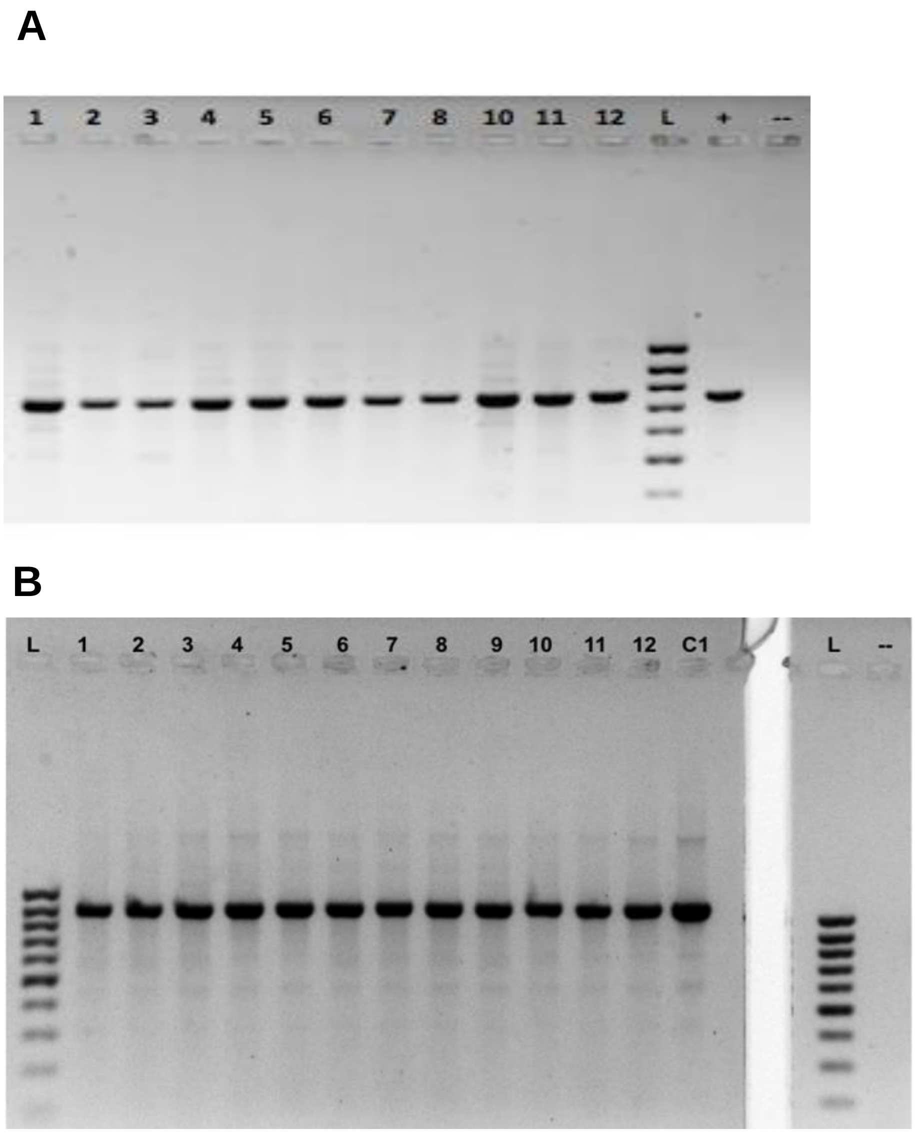


Figure 16. Example of agarose gel electrophoresis showing PCR amplification of CLOI-3. Transgenic *A. thaliana* DNA samples #1-12 containing the respective candidate insulator are shown. The expected size of ~750bp using SALK primers (**A**) and ~900bp using 1300LacZF/GUS5Rev primers (**B**) is shown. All gels include the Fermentas MassRuler Low Range DNA Ladder labeled as '**L**'.

A: + = positive control (wild-type *A. thaliana* ecotype Columbia)
-- = negative control (No DNA)

B: + = positive control (CLOI-3 DNA in pB31 (~900bp) labeled '**C1**'

-- = negative control (No DNA)

3.3 Analysis of subclones of CLOI-3 to define areas of insulator activity

Three separate experimental replicates had indicated that CLOI-3 was a strong insulator candidate (**Figure 8 and 10**). To localize the region that may have insulator activity, three subclones of CLOI-3 of ~150bp each (**Figure 17**) were generated as described in the **Appendix page 92**, cloned into pB31 and used to transform *A. thaliana*. These are preliminary results due to the lack of replicates and no sequencing has been completed to date. Only one transformation was completed for each subclone except for CLOI-3Sub3, which has two replicates. However, PCR was completed to confirm the presence of the insert of the expected size in all replicates except for the second replicate of CLOI-3Sub3 (**Appendix Figure 3**).

The results obtained from subclones CLOI-3Sub1, CLOI-3Sub2 and CLOI-3Sub3 display GUS staining in almost all tissues. For CLOI-3Sub1 (**Figure 18**) GUS staining comparable to pB31 was observed in all tissues sampled (**Figure 9**) with GUS staining in 16% of flower samples, 60% of leaf samples, and 55% of silique samples. CLOI-3Sub2 (**Figure 19**) had no GUS staining of flower samples, 45% GUS staining of leaf samples, and 44% of silique samples. Two transformation events were completed for CLOI-3Sub3, (**Figure 20** and **21)** with both results displaying GUS staining in the leaf and silique tissues. The first replicate of CLOI-3Sub3 (**Figure 20**) showed 10% staining of the flower tissue, and 40% staining of the leaf and silique tissue. In addition, the second replicate of CLOI-3Sub3 (**Figure 21**) displayed no GUS staining of the flower tissue, 47% GUS staining of the leaves, and 13% GUS staining of the siliques. Furthermore, all subclones notably demonstrated a low percentage of GUS staining in the flower tissue. These results suggest that these individual regions do not have the components necessary to achieve insulator activity in contrast to almost no staining in any tissue as seen for CLOI-3, CLOII-10 and CLOII-12 that demonstrate possible insulator activity. Possible limitations and interpretations of these results will be included in the Discussion section.

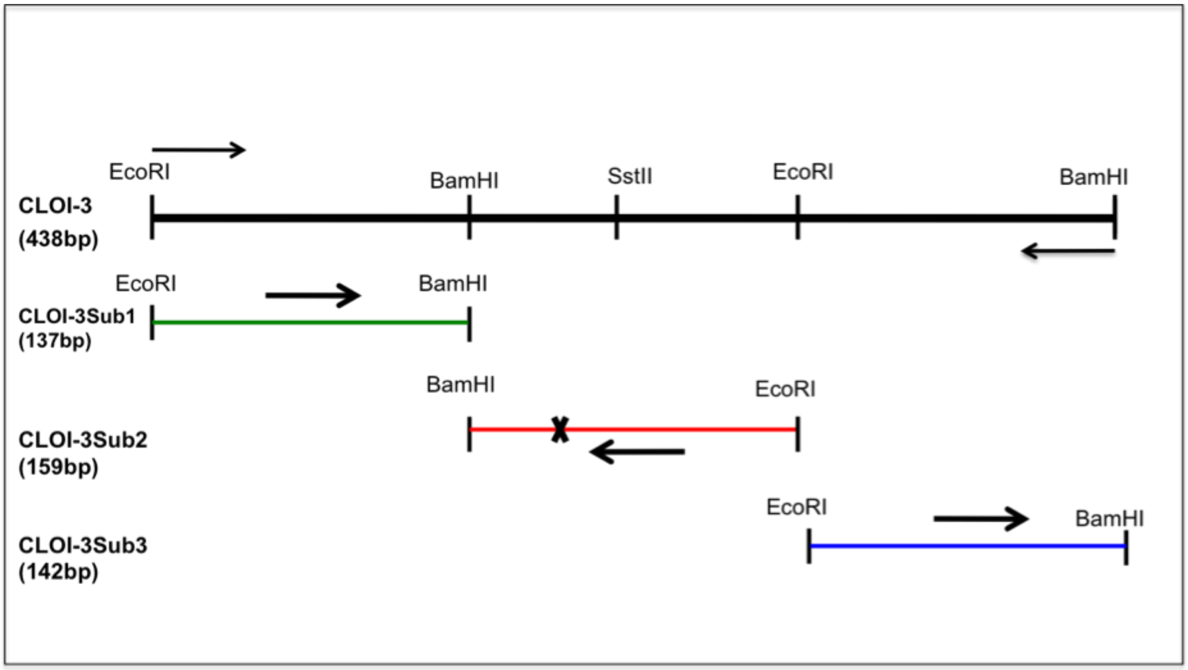


Figure 17. CLOI-3 Subclone Map: Deletions in the candidate CLOI-3 were introduced to determine functional sequences. Each subclone was cloned into pB31 for screening of GUS staining. CLOI-3Sub2 is the only subclone whose sequence after cloning was altered by cloning at the site indicated by an X, where an A was changed to a G. This clone is also the only one cloned in the reverse orientation.

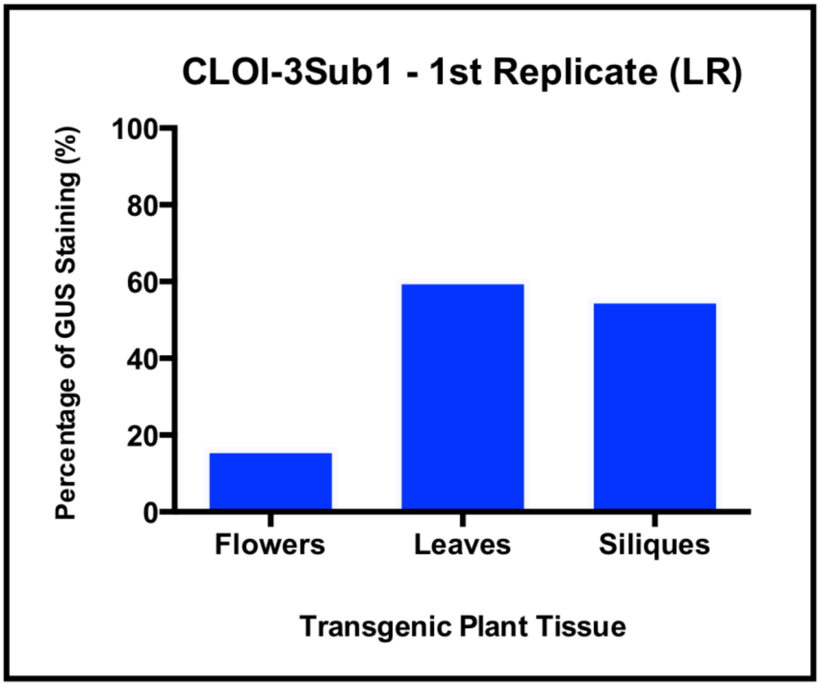


Figure 18. GUS staining of CLOI-3Sub1 (137bp) transformants in pB31. CLOI-3Sub1 (137bp) is a sequence that was cloned into pB31 that displayed GUS staining in the flowers (16%), leaves (60%) and siliques (55%) in the first replicate of transformants. N=82 transformed plants.

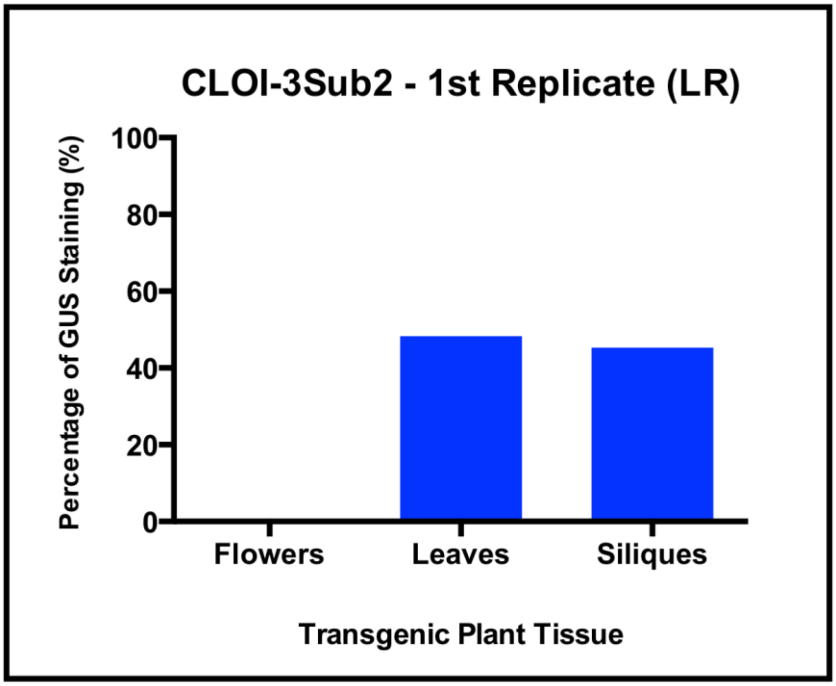


Figure 19. GUS staining of CLOI-3Sub2 (159bp) transformants in pB31. CLOI-3Sub2 is a sequence that was cloned into pB31 that displayed no GUS staining in the flower tissue, GUS staining was seen in leaves (45%) and siliques (44%) in the first replicate of transformants. N=39 transformed plants.

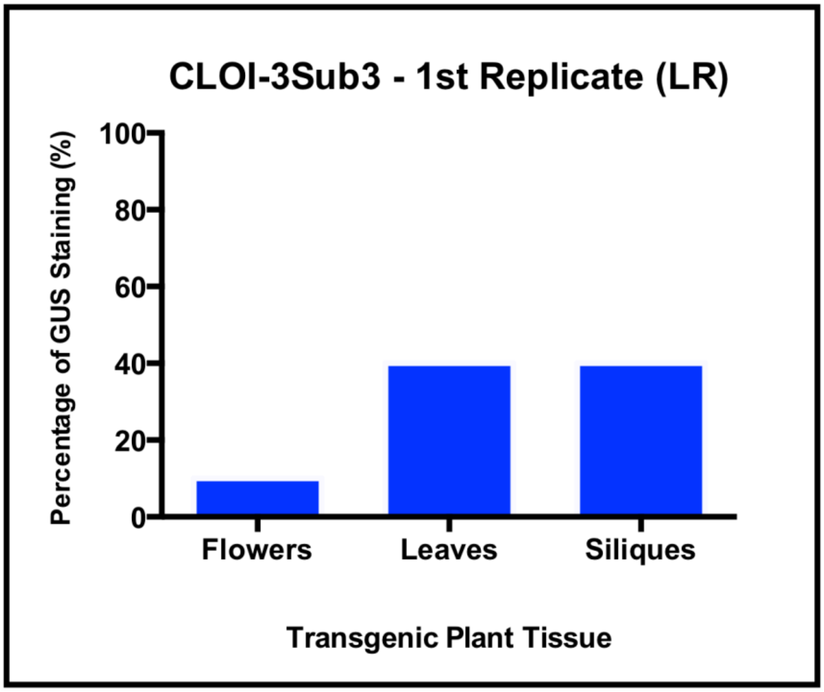


Figure 20. GUS staining of CLOI-3Sub3 (142bp) transformants in pB31. CLOI-3Sub3 (142bp) is a sequence that was cloned into pB31 that displayed GUS staining in the flowers (10%), leaves (40%) and siliques (40%) in the first replicate of transformants. N=10 transformed plants.

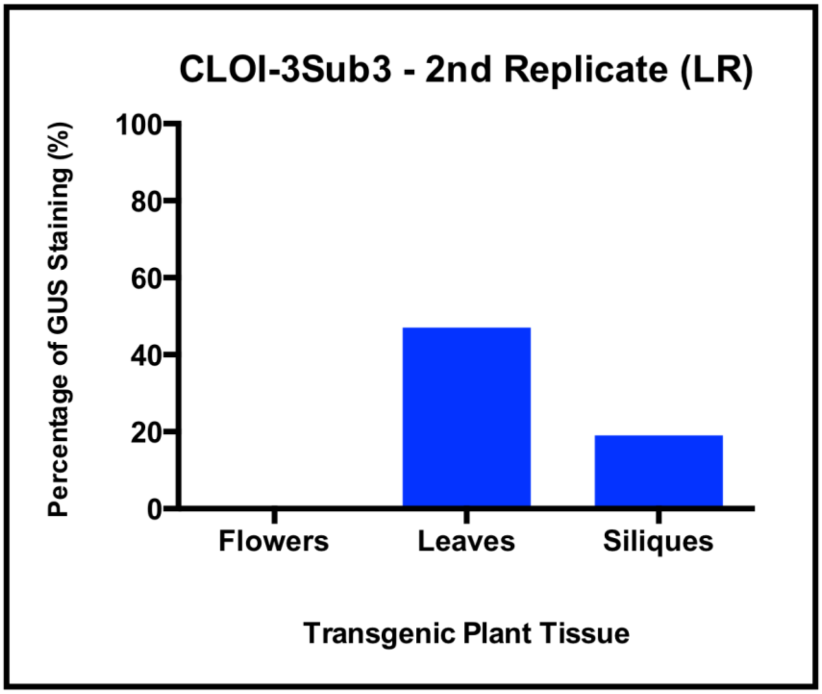


Figure 21. GUS staining of CLOI-3Sub3 (142bp) transformants (2nd replicate) in pB31. CLOI-3Sub3 (142bp) is a sequence that was cloned into pB31 that has displayed no GUS staining in the flower tissue, GUS staining was seen in the leaves (47%) and siliques (13%) in the second replicate of transformants. N=75 transformed plants.

3.4 Bioinformatics and sequence analysis of candidate insulator sequences

3.4.1 Identification of consensus sequences

It is important to determine whether any long putative insulator sequences have the same sequence or whether they contain the same short sequence in either orientation. Long sequences may be derived from multiple ligation events resulting in a long sequence containing three short sequences (B. Gandorah thesis, 2012, **section 3.4 Figures 13 and 14**); therefore it is necessary to search for consensus sequences in both orientations. Dot matrix comparisons of each candidate long sequence (CLOI-3 and CLOII-12) that consists of three short (~150bp) sequence types and one short sequence (CLOII-10) was completed to determine matches. The matrices display matches when sequences are compared to themselves, but none of our candidate insulators display sequence similarity in either the forward orientation or its inverse compliment when compared to each other, except for smaller sequences that may be conserved and the primer sites that are in every sequence. The dot matrices were produced to compare the sequences of interest for major similarities, not to look at small conserved sequences, since it is common to see them due to random matches and primer sites. The dot matrices are available in the **Appendix Figure 2.** Lastly, CLOIII-78 is a sequence that displays putative insulator activity, which needs further research. This sequence was also analyzed for GUS staining in the past by BG and showed minimal staining in the tissues. Further dot matrices analysis (**Appendix Figure 2**) showed no consensus sequences with other putative insulators indicating itself as a sequence that requires further analysis. These results taken together indicate that the sequences we are analyzing are distinct putative insulator sequences that may act as promising enhancer-blocking elements.

3.4.2 Identification of motifs similar to insulator sequences in other species

An in depth analysis of identified sequence motifs and common sequences were searched using the JASPER online database of plant motifs (http://jaspar.genereg.net/cgi-bin/jaspar_db.pl?rm=browse&db=core&tax_group=plants). The database was used to identify plant motifs that correspond to insulator activity. Extensive review of motifs was completed by inputting each sequence and searching the function of each motif given in the output. Motifs were finally chosen based on research that indicates a role in insulator activity. This search identified two motifs, the CTCF binding site (CCCTC) and the Sp1 site (GGGAGG) that have also been reviewed by Bell et al., 1999 (**Table 1**). Furthermore, motifs that have been listed in **Table 1** that have been found in identified insulators were all searched in the CLOI-3, CLOII-10 and CLOII-12 sequences. The results of this search indicated none of the other sites except the CTCF binding site as a common motif that is found in our putative insulators. This method of searching looks for exact matches, but we understand that there may be sequence variation in the motifs due to the presence of alleles in the population, in addition to the divergence of plants from animals and insects. The CTCF binding site has been identified as an important component in insulators found in vertebrates such as *Homo sapiens* and the chicken β–globin locus as discussed in section **1.4.3** (Walters et al., 1999; Chung et al, 1993). Furthermore, Sp1 is a binding site that can act as an inhibitor of enhancer blocking activity in *Homo sapiens* (Bell et al., 1999). The integral function of these binding sites have important affects on insulator activity, therefore we are using these motifs to identify common sequences within our insulators in order to confirm insulator activity.

ActivePerl bioinformatic software was used to locate the exact position of the motifs identified within our sequences. This software allowed us to determine the presence and location of the various motifs.

The localization of CTCF sites and Sp1 sites that are known to be present in our putative insulator sequences was performed first. **Table 5** summarizes the number of sites found within each putative insulator sequence in addition to their positions. These motifs were analyzed in forward and reverse orientation. All the putative sequences (CLOI-3, CLOII-10, CLOII-12, CLOI-3Sub1, CLOI-3Sub2 and CLOI-3Sub3) analyzed in **Table 5** demonstrated at least one CTCF site in either forward or reverse orientation. **Figure 22** displays a schematic of the CTCF and Sp1 sites located in the putative sequences and summarize the results listed in **Table 5**. Sp1 sites were only found in CLOI-3 and its subclones (Sub1 and Sub3), as summarized in **Table 5**. It is interesting to see the Sp1 site at the start (Sub1) and end (Sub3) of the CLOI-3 sequence shown in **Figure 23**. Furthermore, these motifs were searched in other sequences that had been cloned into pB31 by Batool Gandorah and had conclusively shown no insulator activity. These sequences, as summarized in **Table 6**, show varying results. The CTCF site was present in all sequences in either forward or reverse orientation except for CLOII-3, CLOIII-43, CLOIII-53, CLOIII-57 and CLOIII-63. In addition Sp1 sites in either forward or reverse orientation were only found in five sequences (CLOII-3, CLOIII-22, CLOIII-43, CLOIII-53, CLOIII-55). Examples such as unsuccessful candidates, CLOIII-4, which has three CTCF binding sites, and CLOIII-53, CLOIII-57 & CLOIII-63 which has no CTCF binding sites indicates that the CTCF binding site is not a determinant of insulator activity. This simple analysis demonstrates that since the CTCF binding site is present in almost all sequences analyzed, both successful and unsuccessful candidates, it alone cannot define insulator activity. Furthermore, the presence of Sp1 was found in CLOI-3, which is a possible insulator and few unsuccessful sequences (CLOII-3, CLOIII-22, CLOIII-43, CLOIII-53, CLOIII-55), which cannot indicate its role has a repressor of enhancer-blocking activity using this analysis alone.

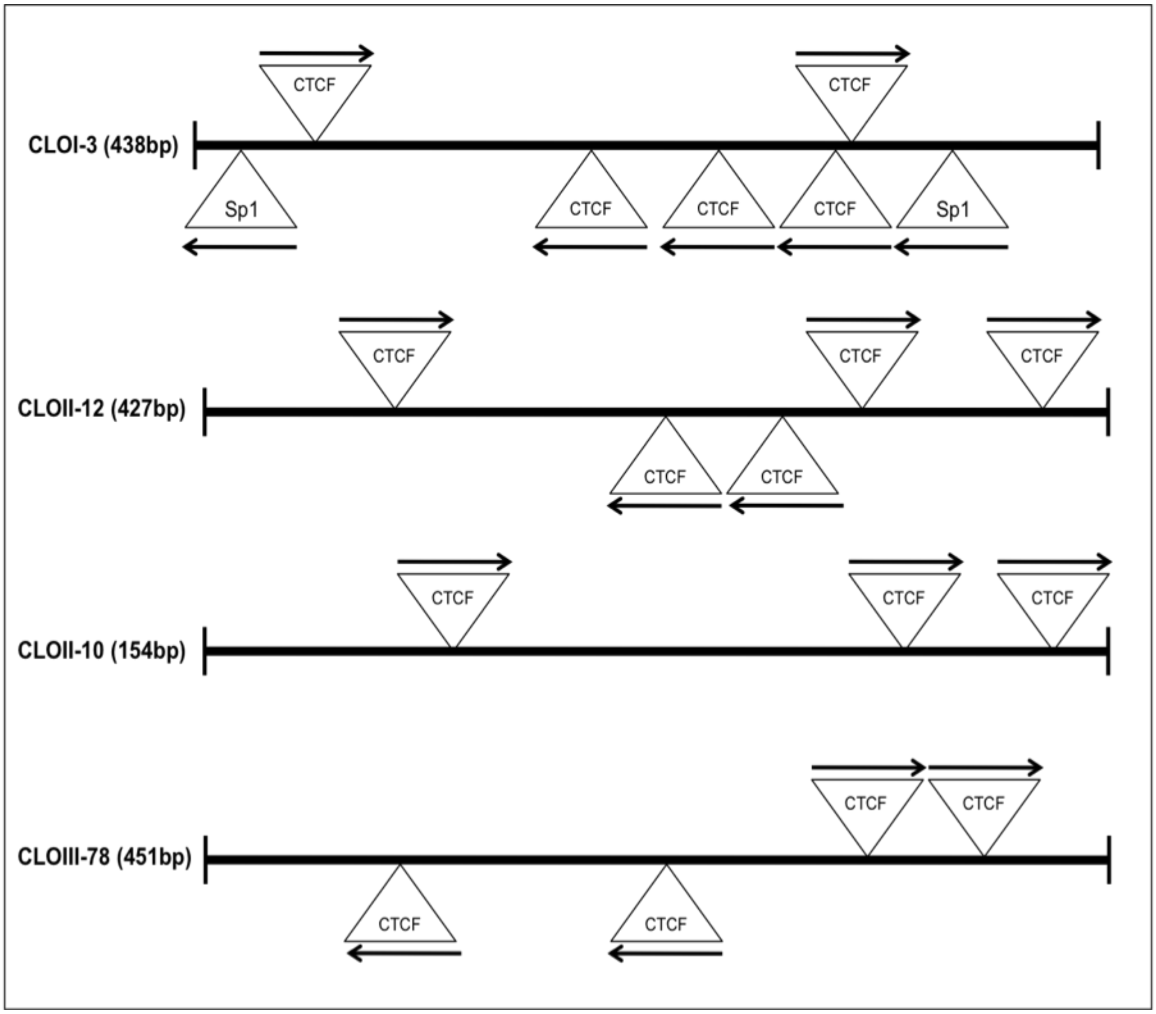


Figure 22. Schematic of CTCF binding site (CCCTC) and Sp1 site (GGGCGG) within putative insulator sequences.

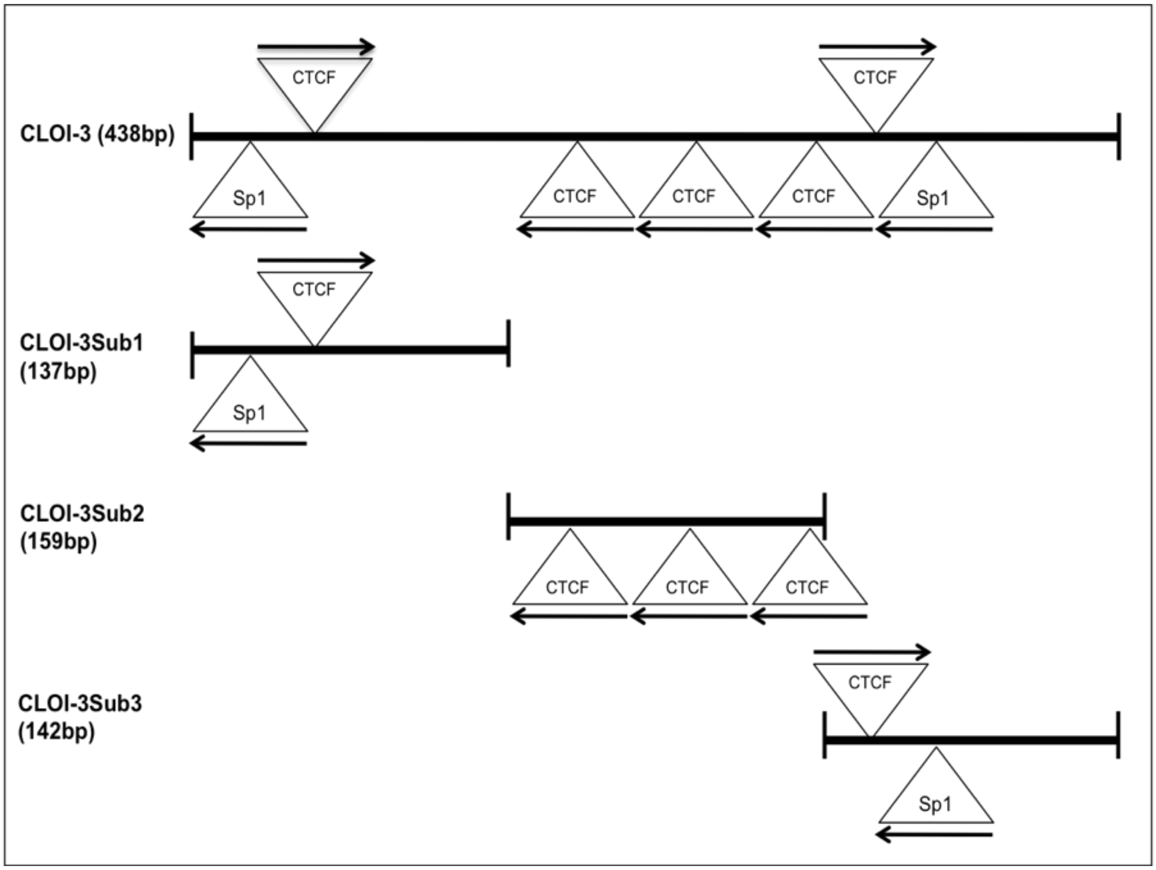


Figure 23. Schematic of CTCF binding site (CCCTC) and Sp1 site (GGGCGG) within CLOI-3 and its subclones.

Inserts in pB31 vector	Insert size (bp)	# of CTCF sites (Fwd)	# of CTCF sites (Rev)	Position of CTCF sites (Fwd)	Position of CTCF sites (Rev)	# of Sp1 sites (Fwd)	# of Sp1 sites (Rev)	Position of Sp1 sites (Fwd)	Position of Sp1 sites (Rev)
CLOII-12	427	3	2	86, 321, 377	235, 254	None	None	None	None
CLOII-10	154	3	0	41, 118, 143	None	None	None	None	None
CLOI-3	438	2	3	57, 335	219, 223, 248	None	2	None	22, 353
CLOI-3Sub1	137	1	0	57	None	None	1	None	22
CLOI-3Sub2	159	0	3	None	53, 72, 101	None	None	None	None
CLOI-3Sub3	142	1	0	335	None	None	1	None	353

Table 5. Summary of motifs listing the number of sites and their exact location within each putative insulator sequence. These sequences are successful insulator candidates as determined by GUS staining using pB31.

Insert in pB31 vector	Insert size (bp)	# of CTCF sites (Fwd)	# of CTCF sites (Rev)	Position of CTCF sites (Fwd)	Position of CTCF sites (Rev)	# of Sp1 sites (Fwd)	# of Sp1 sites (Rev)	Position of Sp1 sites (Fwd)	Position of Sp1 sites (Rev)
CLOII-3	151	0	0	None	None	0	1	None	135
CLOIII-4	440	2	1	92, 348	225	0	0	None	None
CLOIII-17	443	1	0	281	None	0	0	None	None
CLOIII-22	425	2	0	53, 351	None	0	2	None	48, 332
CLOIII-27	128	1	0	73	None	0	0	None	None
CLOIII-43	151	0	0	None	None	0	1	None	128
CLOIII-53	152	0	0	None	None	0	2	None	49, 120
CLOIII-55	432	1	2	44	219, 257	1	1	172	337
CLOIII-57	154	0	0	None	None	0	0	None	None
CLOIII-63	154	0	0	None	None	0	0	None	None
CLOIII-74	153	1	0	61	None	0	0	None	None

Table 6. Summary of motifs listing the number of sites and their exact location within sequences obtained from the random oligonucleotide library. These sequences are unsuccessful insulator candidates as determined by GUS staining using pB31.

3.4.3 Identification of potential repressor sites

Repressors are DNA binding proteins that inhibit expression (transcription) of a gene. They differ from insulators as they inhibit expression by blocking the attachment of RNA polymerase to the promoter. Online searches using the PLACE database (http://www.dna.affrc.go.jp/PLACE/) and a second database called PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), were used to identify known repressors that could potentially block CaMV promoters. These are databases for cis-regulatory elements in plants. This was completed to determine whether lack of expression is due to strong repressor activity instead of insulator activity. The sequence of interest was inputted, and the output list of regulatory elements allows us to search for ones that function as repressors; detailed criteria involving the search can be found in the **Appendix.** This search produced three potential repressor binding sites present within some of our insulator sequences. Two out of the three repressor sites found bind to a W-box (TGAC(C/T)). The first W-box site is WBOXNTERF3, a binding site for the tobacco WRKY proteins involved in wound-activated transcription of the ERF3 gene (Nishiuchi, 2004). It was found in all the putative insulator sequences (CLOI-3, CLOII-10, and CLOII-12) except CLOIII-78 in addition to most of the unsuccessful sequences from the random oligonulceotide library that are listed (**Table 6**). It is important to note that three WBOXNTERF3 sites are present in CLOI-3 indicating the abundance of this site. Furthermore all three of these identical sites are found in CLOI-3Sub3, indicating that it is located in the last ~150bp of the CLOI-3 putative insulator. The implications of this will be explored in further detail in the Discussion section. The second W-box, WRKY71OS, a binding site of rice OsWRKY71 proteins, acts as a transcriptional repressor of the gibberellin signaling pathway in rice and is found in all putative insulator sequences and all unsuccessful insulator candidates. It is interesting to see the presence of the W-box in all the sequences tested, which may indicate a regulatory role. The third repressor site identified is called, CACGTGMOTIF, which is a G-box (CACGTG) that is a binding site of *Arabidopsis* GBF4 (Terzaghi, 1997). This site was found in CLOII-12, CLOII-3 and CLOIII52. CLOII-12 is one of the possible insulator sequences and CLOII-3 & CLOIII-52 are unsuccessful insulator candidates. Furthermore CLOII-12 displays two CACGTGMOTIF sites indicating its frequent presence in this sequence. This analysis demonstrates that the three repressors identified exhibit differing results as the W-box repressor sites are found in all insulator sequences of interest in contrast to the three sequences that have the CACGTG site. The abundant presence of the W-box repressor sites indicates that it cannot act as a strong repressor sequence since expression is seen in all the unsuccessful insulator candidates that also have this repressor site. Furthermore, the CACGTMOTIF site is also found in two unsuccessful candidates that have shown expression, which contradicts its function as a repressor. Detailed description of the database output and corresponding references of each repressor can be found in the **Appendix**.

4.0 Discussion

4.1 Confirmation of insulator activity of putative sequences

To date very few plant insulators have been identified. As a result little is known about insulators and their mechanism of action in plants. We have explored a random oligonucleotide library that may contain sequences with insulator activity in a model *Arabidopsis* system and have identified several candidate insulators.

The pB31 vector used for this study (**Figure 7B**) provided a control for the level of GUS staining when there is no cloned insert (**Figure 9).** While only one replicate was done, we observed GUS staining in 19% of the flower samples while both the leaves and siliques had over 50% GUS staining. These values are comparable but not identical to those found by Singer et al., where their p1 vector without an insert also displayed high GUS staining results (Singer, 2011). Direct comparisons between the two systems are not easy since Singer used a different vector (p1) and different constructs (*AGIP* promoter-*GUS; PIp* promoter-*GUS; SUS1* promoter-*GUS*). In both cases the CaMV35S promoter/enhancer provides the enhancer elements but the influenced promoter is different. We also noticed lower percentages of GUS staining in the flower tissue for samples CLOI-3Sub1, CLOI-3Sub2 and CLOI-3Sub3 in addition to pB31 suggesting that this may be a property of the system we are using. Since we have only one result for pB31, we looked again at staining patterns for insulator candidates previously categorized as not functional (**Appendix Table 2**). While some show the same low flower staining as pB31 and the three sublcones (e.g. CLOIII-57), in general the staining is too variable to see any trends. Further analysis is required to test whether there is tissue specificity.

No GUS staining of plant tissue was observed for CLOI-3, CLOII-10, and CLOII-12 (**Figures 10-14**). CLOI-3 is an insert in pB31 that had two replicates completed by BG (**Figure 8**) of GUS staining analysis, one of which had PCR data and sequence confirmation. The completion of a third replicate (**Figure 10**), which also had PCR (**Figure 16**) and sequence confirmation, demonstrates that CLOI-3 shows promise as an insulator. CLOII-12 also showed no staining of tissues with two replicates completed (**Figure 13 and 14**) for each alongside PCR data (**Appendix Figure 3**) and sequence confirmation. For the third possible insulator, CLOII-10, two replicates were completed with no GUS staining (**Figure 11 and 12**) in addition to PCR data (**Appendix Figure 3**). Sequence data must still be collected.

Typically in transgenic experiments it is difficult to obtain 100% insulator activity. Singer et al. demonstrated that when using the *AGIP* promoter-*GUS* gene test system, approximately 70% of transgenic lines in which the *TBS* element was introduced exhibited blocking of *35S* enhancer-mediated activation in *Arabidopsis* (Singer et al., 2011). Values of less than 100% were also observed in several *Drosophila melanogaster* test systems (Cai et al., 2001; Muravyova et al., 2001). In our system with a large numbers of plants tested we observe very little or no staining. It remains to be investigated whether this is due to the strength of the insulator or is specific to our test system.

It is interesting to see that CLOI-3 and CLOII-12 are sequences that are ~450bp, which is far longer than the original ~150bp pieces that were made during the synthesis of the random oligonucleotide library. These sequences are proving to be promising insulators but further analysis as to where the activity is localized is important as described in **section 4.2**. The insulator CLOII-10 is 154bp long, which makes it an interesting candidate to further evaluate as its short sequence makes it more amenable to mutational analysis e.g. deletion/mutation to localize activity. CLOII-10 is a prime candidate for cloning, as it will be easier to put into various vectors without issues regarding internal restriction sites. A next step would be to clone all candidates into pL1 vector (**Figure 7C**) in order to further confirm insulator function with a different promoter (napin seed-specific promoter). An insulator capable of protecting a range of promoters would be more useful in future cloning experiments.

4.2 CLOI-3 subclone analysis

Deletion mutation analysis is prominently used to identify or localize regions of activity within sequences of interest (Hagstrom et al., 1998; Singer et al., 2011). Thus, we initially chose to produce deletions of CLOI-3 that our results identified as a promising candidate. Therefore three subclones of ~150bp were generated by deletion or sublconing as described in the **Appendix**. While CLOI-3 is about ~450bp, ~150bp is the size of the fragments generated by DNA synthesis. However, GUS staining was still prominent in CLOI-3Sub1, CLOI-3Sub2 and CLOI-3Sub3 (**Figure 18-21**). We conclude that these individual regions alone do not have the DNA components necessary to achieve insulator activity. When deletions of the *TBS* element in *Arabidopsis* (Singer et al., 2011) were analysed, they concluded neither fragment achieved the effectiveness of the full-length insulator, which suggests that there may be an additive effect of several elements within the full-length *TBS*. As indicated above the vectors in our tests are not the same as those described by Singer et al., (2011). To date, no one else has used a vector that is constructed like pC1, pB31 and pL1. In addition we measured staining in three tissue types while Singer et al., only stained the leaf tissue for their deletion analysis with their p1 vector.

The exact components necessary for insulators to function in *Arabidopsis* are not known, as plant insulators are relatively new. Zinc-finger domain proteins have been found to have an essential role in insulator function in species such as *Drosophila*, and *Homo sapiens* as they form key parts of insulators such as Su(Hw) and CTCF (Chung et al, 1993; Geyer, 1986; **Table 1** this thesis). For example, it is known that 12 zinc-finger proteins are required for the *gypsy* insulator and 11 zinger-finger proteins for the 5'HS4 insulator to be functional (Chung et al, 1993; Geyer, 1986). To date homologs of Su(Hw) and CTCF have not been found in *Arabidopsis* but a large number of zinc-finger gene families that share similarity to the zinc-finger domains of vertebrate CTCF proteins have been found (Singer, 2011), Furthermore, She et al. (2010) inserted the identical *gypsy* insulator sequence from *Drosophila* into *Arabidopsis.* Since *Arabidopsis* does not have a homolog for Su(Hw), a clone expressing Su(Hw) was also introduced. This resulted in improved expression levels of transgenes, the reduction of position effects and improved specificity of promoter activity. Emberly (2008), discusses the BEAF binding site in *Drosophila* and how it functions as a dual-core binding site. Each subunit of the BEAF complex targets one CGATA motif. Point mutations within this consensus abolish both its binding and insulating activities. As indicated in **Figure 17**, CLOI-3Sub2 was transformed in the reverse direction in addition to the one-nucleotide point mutation it developed, which may have an implication to the GUS staining that was observed in all the tissues for that subclone. We can conclude that multiple elements are required in combination to exert insulator activity within a domain.

4.3 Bioinformatic analysis and its implications on the putative insulator sequences

Due to the identification of longer sequences than originally generated, it is important to determine whether any long putative insulator sequences contain the same short sequence in either orientation. Long sequences can derive from multiple ligation events; therefore it is necessary to search for matches in both orientations. Dot matrix comparisons of each candidate long sequence (CLOI-3, CLOII-12 and CLOIII-78) that consists of three short (~150bp) sequence types and one short sequence (CLOII-10) was completed to determine matches (**Appendix Figure 2).** The dot matrices established matches when a sequence was compared to itself but displayed no similarity when compared to each other in either the forward orientation or its inverse compliment as seen in these direct comparisons.

A thorough search using the JASPER database and a list of known insulators (**Table 1**) was used to try to identify common motifs found in insulators, and to determine the presence of these motifs in our sequences. The results of this search indicated no sites except the CTCF and Sp1 binding sites were found in our putative insulators. This method of searching looks for exact matches, but we understand that there may be sequence variation in the motifs due to the presence of alleles in the population or the divergence of plants from animals and insects. Further work is required to produce a more accurate search perhaps by using an algorithm that allows queries with degenerate sequences.

We first looked at the distribution of the CCCTC motif and the Sp1 binding site sequences within all of our candidate insulators. CCCTC is a motif that binds the CTCF protein that has been found in many insulators and has been characterized as a factor that is important for insulator activity. As noted above, homologs of CTCF have not been found in *A. thaliana* but the site may have different Zn-finger binding partners in plants. A complete search for this motif displayed that it was found in all candidate sequences that were analyzed in this thesis (**Table 5 and 6**). This may point to a role in insulator activity as found in other systems (**Table 1**); however the presence of a single CCCTC may be insufficient as subclones of CLOI-3, even with multiple copies of this sequence (**Table 5**) did not have activity. In addition some DNAs with no apparent insulator activity e.g. CLOIII-55 (**Table 6**) have multiple copies of this sequence. Thus there appears to be no simple relationship between the number of copies, the presence of the sequence and insulator activity. In *Drosophila* there is evidence for different insulator sequences and binding proteins that are functional in the same species (**Table 1**). Perhaps our inability to find a single insulator sequence unique to all clones reflects diversity of systems in plants too.

Sp1 is a protein that is known to function as an inhibitor of insulator activity in *Homo sapiens* (Bell et al, 1999). We searched for the Sp1 binding motif within all the putative insulators. It was only found in CLOI-3 and its subclones (Sub1 and Sub3) summarized in **Table 5**. It is interesting to see the Sp1 site at the start (Sub1) and end (Sub3) of the CLOI-3 sequence shown in **Figure 23**. We can predict that this motif may be a reason why the subclone sequences may not be able to block enhancer-promoter interaction, as this motif could inhibit that function.

As indicated in the Results, it is possible that the insulator activity we observe is actually due to the cloning of a site for a strong repressor. If this is true then all the tissues we tested for GUS staining must contain the repressor protein. An online search using the PLACE database identified a repressor site, the W-box (TGAC(C/T), which is a binding site of WRKY proteins which have multiple functions in different plants such as wound-activated transcription of the ERF3 gene in tobacco (WBOXNTERF) (Nishiuchi, 2004) or transcriptional repression of the gibberellin signaling pathway in rice (OsWRKY71OS = At1g29860). A search of the BAR database (http://bar.utoronto.ca/welcome.htm) showed that At1g29860 is expressed in all tissues in *Arabidopsis.* However, the W-box is found in all of the unsuccessful sequences from the random oligonucleotide library that are listed in **Table 6** in addition to the putative insulator sequences (CLOI-3, CLOII-10, CLOII-12, and CLOIII-78). Thus we can conclude that the repressor binding sites identified by database searches are not specific to the sequences that function as insulators in our assay system. If we have detected a site for a strong repressor, then it has not been observed before.

4.4 Future work

It is important to look at some of the next steps we can take after obtaining the results described in this thesis. As we move forward with these results, we understand that repetition of our transformation results for each sequence is necessary to validate insulator activity. Furthermore, the subclone analysis that was completed for CLOI-3 indicates that further analysis is needed to localize the area of insulator activity. Combinations such as CLOI-3Sub1 and Sub2 or CLOI-3Sub2 and Sub3 as one sequence would be interesting to analyze. This would allow us to see whether or not we missed a key area when cutting the sequence in the initial deletion/mutation or if the size of the subclones didn't allow the functional components to be present within the sequence (**Figure 17**), which may identify the region where insulator activity is actually localized. In addition, testing these sequences by cloning them into the pL1 screening vector will allow us to further confirm insulator function of our putative insulators. Since pL1 has a napin seed-specific promoter, it will allow us to also confirm if the sequences are acting as repressors/silencers, since staining will only be seen in the seeds if there is a functional insulator, which will further confirm insulator activity. This difference can be seen in the work of Gudynaite-Savitch et al., (2009), Table 1. Insulator activity is detected when either BEAD1C or UASrpg were cloned into pL1 as indicated by GUS staining found only in seeds while repressor activity is indicated by the loss of GUS staining in all tissues as when Fab7PRE was cloned into pL1. Furthermore, the repressor analysis of our candidate insulators touched lightly on an area that requires a much more in depth analysis. Further steps such as creating a system to find and identify repressor sites would be advantageous in the efforts to characterize insulators in plants. Lastly, it would be interesting to clone our sequences into other species such as tobacco and fungi. For example, the full length *TBS* fragment has also been found to be effective for impeding enhancer-promoter interference in transgenic tobacco plants and is not promoter specific, indicating that it may function as an insulator in other plant species (Singer, 2011). Since a random oligonucleotide library produced our sequences, it is possible for these sequences to be functional in a variety of species, which would also confirm the conserved function of insulator activity in a range of species.

An important objective would be the identification of the protein(s) able to bind the putative insulator. One approach to achieving this goal would be to use EMSA (electrophoretic mobility shift assay) to first detect the protein in nuclear extracts and then monitor its purification (Bell et al, 1999). Candidates can then be sequenced and then studied for their binding properties *in vitro*.

4.4.1 Enhancer-blocking activity, barrier activity or both?

Enhancer-blocking and barrier activities within an insulator function separately by instilling functionality through different mechanisms. Both activities ultimately play a role in isolating domains for selecting traits in transgenic plants. The inserts we have identified in this thesis have been defined as enhancer-blocking insulators as we have only assayed for this function. The question is whether or not an insulator can act as both an enhancer-blocking and barrier element. Plant insulators have not been identified as barrier elements as of yet; published plant insulators have only been identified as enhancer-blocking elements. Moreover, the chicken 5'HS4 is the characterized insulator that is able to act as both elements (**Section 1.4.3**) although the activities maps to difference sequences. It would be interesting to develop an assay that is capable of determining barrier activity in our sequences. To be able to have an insulator in plants that is able to block enhancer-promoter interaction and block the spread of heterochromatin would be quite advantageous towards the development of transgenic plants.

4.5 Conclusions

Research in plant biotechnology is becoming increasingly important as the demand for crops increases for energy and food supply as the world population is increasing exponentially. A method to select for desirable traits in transgenic plants is vital, as chromatin position effects and transgene misexpression are problems. This thesis demonstrates that we can isolate candidate sequences with insulator activity. Sequence analysis has not identified a sequence unique to candidate insulators.

REFERENCES

Allen, G.C., Hall, G.E., Jr., Childs, L.C., Weissinger, A.K., Spiker, S., and Thompson, W.F. (1993). Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. *Plant Cell, 5,* 603–613.

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

Bell, A.C., West, A.G., Felsenfeld, G. (1999). The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. *Cell, 98*, 387-396.

Bell, A.C., *et al*. (2001). Insulators and Boundaries: Versatile Regulatory Elements in the Eukaryotic Genome*. Science, 291*, 447.

Bi, X., Broach, J.R. (1999). UASrpg can function as a heterochromatin boundary element in yeast. *Genes & Development, 13,* 1089-1101.

Breyne, P., Vanmontagu, M., Depicker, A., Gheysen, G. (1992). Characterization of a plant scaffold attachment region in a DNA fragment that normalizes transgene expression in tobacco. *Plant Cell*, *4*, 463–471.

Brouwer, C., Bruce, W., Maddock, S., Avramova, Z., Bowen, B. (2002). Suppression of transgene silencing by matrix attachment regions in maize: A dual role for the maize 5′ ADH1 matrix attachment region. *Plant Cell*, *14*, 2251–2264.

Bushey, A.M., Dorman, E.R., Corces, V.G. (2008). Chromatin Insulators: Regulatory Mechanisms and Epigenetic Inheritance. *Molecular Cell, 32*, 1-9.

Cai, H.N., Shen, P. (2001). Effects of cis Arrangement of Chromatin Insulators on Enhancer-Blocking Activity. *Science, 291*, 493-495.

Chetverina, D.A., Elizaryev, P.V., Georgiev, P.G., Erokhin, M.M. (2012). 1A2 Insulator can Interact with Promoter of hsp70 Gene in D. melanogaster. *Molecular Genetics, 49*, 371-379.

Chung, J.H., Whiteley, M., Felsenfeld, G. (1993). A 5’ Element of the Chicken B-Globin Domain Serves as an Insulator in Human Erythroid Cells and Protects against Position Effect in Drosophila. *Cell, 74*, 505-514.

Chung, J.H., Bell, A.C., Felsenfeld, G. (1997). Characterization of the chicken B-globin insulator. *Proceedings of the National Academy of Science, 94*, 575-580.

Clark, A.J., Bissinger, P., Bullock, Damak, S., Wallace, R., Whitelaw, C.B.A., Yull, F. (1994). Chromosomal Position Effects and the Modulation of Transgene Expression. *Reprod. Fertil. Dev., 6,* 589-598.

Clough, S.J., Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant Journal, 6*, 735-743.

Donze, D., Adams, C.R., Rine, J., and Kamakaka, R.T. (1999). The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. *Genes Dev*., *13*, 698–708.

Donze, D., Kamakaka, R.T. (2001). RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in *Saccharomyces cerevisiae*. *The EMBO Journal, 20*, 520–531.

Dorsett, D. (1999). Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Current Opinion in Genetics & Development, 9*, 505-514.

Ebersole, T., Kim, J., Samoshkin, A., Kouprina, N., Pavlicek, A., White, R.J., Larionov, V. (2011). tRNA genes protect a reporter gene from epigenetic silencing in mouse cells. *Cell Cycle, 10,* 2779-2791.

Emberly, E., Blattes, R., Schuettengruber, B., Hennion, M., Jiang, N., Hart, C.M., Kas, E., Cuvier, O. (2008). *Plos Biology, 6*, 2896-2910.

Fourel, G., Revardel, E., Koering, C.E., Gilson, E. (1999). Cohabitation of insulators and silencing elements in yeast subtelomeric regions. *The EMBO Journal, 18,* 2522-2537.

Gandorah, B. (2012) MSc Thesis: Identifying Insulators in *Arabidopsis thaliana*. http://hdl.handle.net/10393/23226

Gatz, C., Kaiser, A., Wendenburg, R. (1991). Regulation of a odified CaMV 35S promoter by the Tn*10*-encoded Tet repressor in transgenic tobacco. *Molecular Genetics and Genomics 227,* 229-237.

Gaszner, M., Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Reviews, 7*, 703-713.

Georgiev, P., Kozycina, M. (1996). Interaction Between Mutations in the *suppressor of Hairy wing* and *modifier of mdg4* Genes of *Drosophila melanogaser* Affecting the Phenotype of *gypsy*-Induced Mutations. *Genetics, 142*, 425-436.

Gerasimova, T.L., Byrd, K., Corces, V.G. (2000). A Chromatin Insulator Determines the Nucelar Localization of DNA. *Molecular Cell, 6*, 1025-1035.

Geyer, P.K., Spana, C., Corces, V.G. (1986). On the molecular mechanism of gypsy-induced mutations at the *yellow* locus of *Drosophila melanogaster*. *The EMBO Journal, 5,* 2657-2662.

Geyer, P.K. (1997). The role of insulator elements in defining domains of gene expression. *Current Opinions in Genetics & Developments, 7,* 242-248.

Ghirlando, R., Giles, K., Gowher, H., Xiao, T., Xu, Z., Yao, H. Felsenfeld, G. (2012). Chromatin domains, insulators, and the regulation of gene expression. *Biochimica et Biophysica Acta*, *1819*, 644-651.

Gudynaite-Savitch, L., Johnson D.A., Miki B.L.A. (2009). Strategies to mitigate transgene-promoter interaction. *Plant Biotechnology Journal, 7,* 472-485.

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

Han, K.H., Ma, C.P., Strauss, S.H. (1997). Matrix attachment regions (MARs) enhance transformation frequency and transgene expression in poplar. *Transgenic Research*, *6*, 415–420.

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

Jagannath, A., Bandyopadhyay, P., Arumugam, N., Gupta, V., Burma, P.K., Pental, D. (2001). The use of a Spacer DNA fragment insulates the tissue-specific expression of a cytotoxic gene (*barnase*) and allows high-frequency generation of transgenic male sterile lines in *Brassica juncea* L. *Molecular Breeding, 8*, 11-23.

Kellum, R., Schedl, P. (1991). A Position-Effect Assay for Boundaries of Higher Order Chromosomal Domains. *Cell, 64,* 941-950.

Kellum, R., Schedl, P. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Molecular and Cellular Biology, 12,* 2424-2431.

Kim, S., Veena, Gelvin, S. (2007). Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. *The Plant Journal, 51*, 779-791.

Kirkland, J.G., Raab, J.R., Kamakaka, R.T. (2013). TFIIIC bound DNA elements in nuclear organization and insulation. *Biochimica et Biophysica Acta, 1829*, 418-424.

Koncz, C. and Schell, J. (1986) The promoter of the TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Mol. Gen. Genet. 204*, 383–396.

Li X., Zhu, Z., Junwang, X., Qian, W., Honglin, X. (2001). Isolation of pea matrix attachment region and study on its function in transgenic tobaccos. *Science China C*, *44*, 400-408.

Lunyak V.V., *et al.* (2007). Developmentally Regulated Activation of a SINE B2 Repeat as a Domain Boundary in Organogenesis. *Science, 317*, 248.

Malik, K., Wu, K., Li, X.-Q et al. (2002). Constitutive gene expression system derived from the tCUP cryptic promoter elements. *Theor. Appl. Genet., 105,* 505-514.

Majumder, P., Cai, H.N. (2003). The functional analysis of insulator interactions in the *Drosophila* embryo. *PNAS, 100,* 5223-5228.

Majumder, P., Roy, S., Belozerov,V.E., Bosu, D., Puppali, M., Cai., H.N. (2009). Diverse transcription influences can be insulated by the *Drosophila* SF1 chromatin boundary. *Nucleic Acids Research, 37*, 4227-4233.

Maximova S, Miller C, Antúnez de Mayolo G, Pishak S, Young A, Guiltinan MJ. (2003). Stable transformation of *Theobroma cacao* L. and influence of matrix attachment regions on GFP expression. *Plant Cell Rep.*, *21*, 872-83.

McKinney, E.C., Ali, N., Traut, A., Feldmann, K.A., Belostotsky, D.A., McDowell, J.M., Meagher, R.B. (1995). Sequence-based identification of T-DNA insertion mutations in *Arabidopsis*: actin mutants *act2-1* and *act4-1*. *Plant J*., *8*, 613-622.

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

Mongelard, F., Corces, V.G. (2001). Two insulators are not better than one. *Nature, 8*, 192-194.

Murashige, T., Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum, 15,* 473-497.

Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V., Georgiev, P. (2001). Loss of Insulator Activity by Paired Su(Hw) Chromatin Insulators. *Science, 291*, 495-498.

Nagaya, S., Yoshida, K., Kato, K., Akasaka, K., Shinmyo, A. (2001). An insulator element from the sea urchin *Hemicentrotus pulcherrimus* suppresses variation in transgene expression in cultured tobacco cells, *Mol Genet Genomics*, *265*, 405-413.

Nisiuchi, T., Shinshi, H., Suzuki, K. (2004). Rapid and Transient Activation of Transcription of the ERF3 Gene by Wounding in Tobacco Leaves. *The Journal of Biological Chemistry, 279*, 55355-55361.

Odell, JT., Knowlton, S., Lin, W., Mauvais, J. (1988). Properties of an isolated transcription stimulating sequence derived from the cauliflower mosaic virus *35S* promoter. *Plant Molecular Biology, 10*, 263-272.

Oh, S-J., Jeong, J.S., Kim E.H., Yi, N.R., Yi, S-I., Jang, I-C., Kim, Y.S., Suh, S-C., Nahm, B.H., Kim, J-K. (2005). Matrix attachment region from the chicken lysozyme locus reduces variability in transgene expression and confers copy number-dependence in transgenic rice plants. *Plant Cell Rep*, *24,* 145-154.

Ott, R.W., Ren, L., Chua, N. (1990). A bidirectional enhancer cloning vehicle for higher plants. *Molecular Genetics and Genomics, 221,* 121-124.

Pai, C., Lei, E.P., Ghosh, D., Corces, V.G (2004). The Centrosomal Protein CP190 Is a Component of the gypsy Chromatin Insulator. *Molecular Cell, 16*, 737-748.

Petersen, K., Leah, R., Knudsen, S., Cameron-Mills, V. (2002). Matrix attachment regions (MARs) enhance transformation frequencies and reduce variance of transgene expression in barley. *Plant Molecular Biology*, *49*, 45-58.

Pikaart, M.J., Recillas-Targa, F., Felsenfeld, G. (2009). Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes & Development, 12*, 2852-2862.

Raab, J.R., Chiu, J., Zhu, J., Katzmann, S., Kurukuti, S., Wade, P.A., Haussler, D., Kamakaka, R.T. (2012). Human tRNA genes function as chromatin insulators. *The EMBO Journal, 31*, 330-350.

Recillas-Targa, F., Pikaart, M.J. Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., Felsenfeld, G. (2002). Position-Effect Protection and Enhancer Blocking by the Chicken Β-Globin Insulator are Separable Activities. *Proceedings of the National Academy, 10*, 6883-6888.

Roman, A.C., Gonzalez-Rico, F.J., Molto, E., *et al.* (2011). Dioxin receptor and SLUG transcription factors regulate the insulator activity of B1 SINE retrotransposons via an RNA polymerase switch. *Genome Research, 21*, 422-432.

Schöffl, F., Schröder, G., Kliem, M., Rieping, M. (1993). A SAR sequence containing 395 bp fragment mediates enhanced, gene-dosage-correlated expression of a chimaeric heat shock gene in transgenic tobacco plants. *Transgenic Res*, *2*, 93–100.

She, W., *et al.* (2010). 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, 185*, 1141-1150.

Singer, S.D., Cox, K.D. (2013). A *gypsy*-like sequence from *Arabidopsis thaliana* exhibits enhancer-blocking activity in transgenic plants. *Journal of Plant Biochemical and Biotechnology*, 22, 35-42.

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

Singer, S.D., Hily, J., Zongrang, L. (2010). A 1-kb Bacteriophage Lambda Fragment Functions as an Insulator to Effectively Block Enhancer-Promoter Interactions in *Arabidopsis thaliana*. *Plant Molecular Biology Reports, 28*, 69-76.

Singer, S.D., Liu, Z., Cox, K.D. (2012). Minimizing the unpredictability of transgene expression in plants: the role of genetic insulators. *Plant Cell Reports, 31,* 13-25.

Sipos, L., Mihaly, J., Karch, F., Schedl, P., Gausz, J., Gyurkovics, H. (1998). Transvection in the Drosophila Abd-B Domain: Extensive Upstream Sequences Are Involved in Anchoring Distant cis-Regulatory Regions to the Promoter. *Genetics, 149*, 1031-1050.

Sugiyama, T., Cam, H., Grewal, S.I.S. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *PNAS, 102*, 152-157.

Terzaghi, W.B., Bertekap, R.L., Cashmore, A.R. (1997). Instracellular localization of GBF proteins and blue light-induced import of GBF2 fusion proteins into the nucleus of cultured Arabidopsis and soybean cells. *The Plant Journal, 11*, 967-982.

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

Valenzuela, L., Kamakaka, R.T. (2006). Chromatin Insulators. *Annual Review of Genetics, 40*, 107-138.

van der Geest, A.H., Hall, T.C. (1997). The beta-phaseolin 5′ matrix attachment region acts as an enhancer facilitator. *Plant Molecular Biology*, *33*, 553–557.

Vazquez, J., Schedl, P. (1994). Sequences required for enhancer blocking activity of scs are located within two nuclease-hypersensitive regions. *The EMBO Journal, 13*, 5984-5993.

Walters, M.C., et al. (1999). The Chicken B-Globin 5'HS4 Boundary Element Blocks Enhancer-Mediated Suppression of Silencing. *Molecular and Cellular Biology, 19,* 3714-3726.

Wang, W., et al. (2006). Identification of Important Regions for Ethylene Binding and Signaling in the Transmembrane Domain of the ETR1 Ethylene Receptor of Arabidopsis. *Plant Cell, 18*, 3429-3442.

West A.G., Gaszner, M., Felsenfeld, G. (2002). Insulators: many fucntions, many mechanisms. *Genes & Development, 16*, 271-288.

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

Yang, Y. Singer, S., D. Liu, Z. (2010). Evaluation and comparison of the insulation efficiency of three enhancer-blocking insulators in plant*. Plant Cell, Tissue and Organ Culture, 105,* 405-414.

Zhang, Y., Zheng, Y., Xiao, N., Wang, L., Zhang, Z., Fang, R., Chen, X. (2012). Functional analysis of the *HS185* regulatory element in the rice *HSP70* promotor. *Molecular Biology Reports*, *39,* 1649-1657.

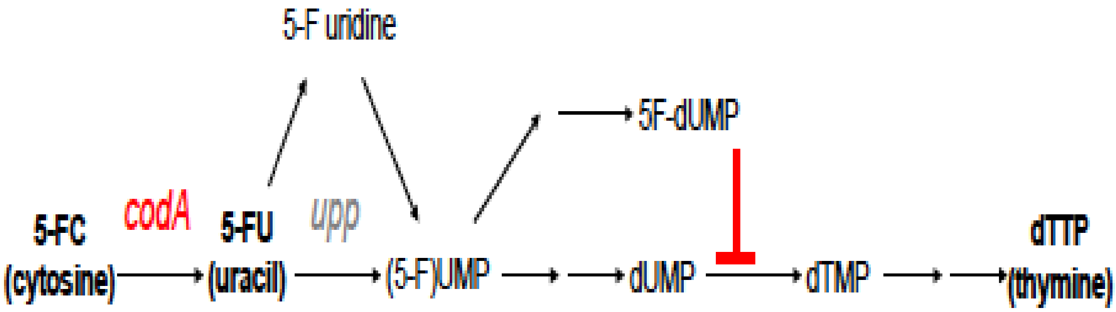
Zhong, X-P., Krangel, M. (1997). An enhancer-blocking element between α and δ gene segments within the human T cell receptor α / δ locus. *PNAS, 94,* 5219-5224.

Appendix

The Pathway of 5-FC Conversion

Figure 1. The toxic metabolic pathway of 5-FC catalyzed by the product of *codA* expression. Normally cytosine is metabolized to uracil, and eventually to dTTP (deoxythymine triphosphate) that is required for all DNA synthesis. Cytosine deaminase enzyme encoded by *codA* converts5-FC to 5-FU, which is then processed to 5F-dUMP (5-fluoro-deoxyuridine monophosphate) indirectly, via the intermediate 5F-uridine or directly, by uracil phosphoribosyltransferase (*upp*). 5F-dUMP irreversibly inhibits thymidylate synthase activity, and as a result the cells are deprived of deoxythymidine triphosphate (dTTP) necessary for DNA synthesis.

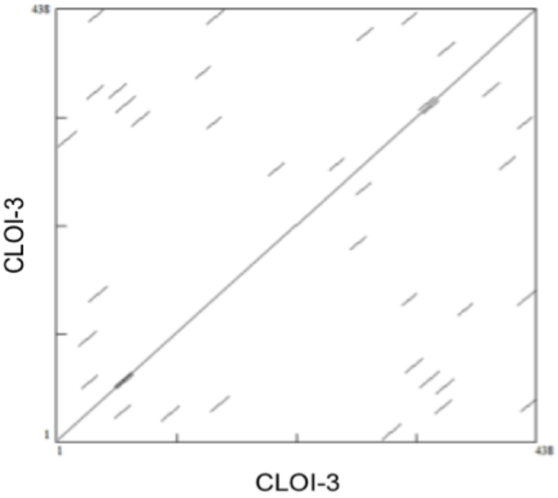
Modified from (http://parts.igem.org/Part:BBa_K404125).


5-F uridine
5F-dUMP
5-FC
(cytosine)
codA
5-FU
(uracil)
upp
(5-F)UMP
dUMP
dTMP
dTTP
(thymine)

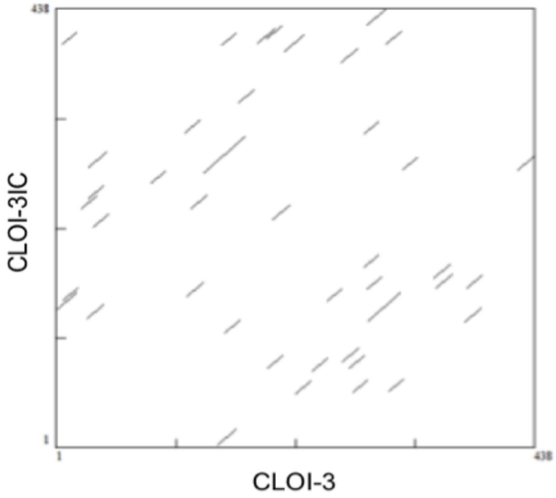
Dot Matrix Comparisons between Selected Potential Insulators

Figure 2. Dot matrix comparisons among each of the four candidate Insulator sequences CLOI-3, CLOII-10, CLOII-12 and CLOIII-78. Samples were compared pairwise, first with the sequence (**A, C, E, G, I, K, M, O, Q, S**) and then with the inverse complement (**B, D, F, H, J, L, N, P, R, T**). Used with permission of Dr. Douglas A. Johnson.

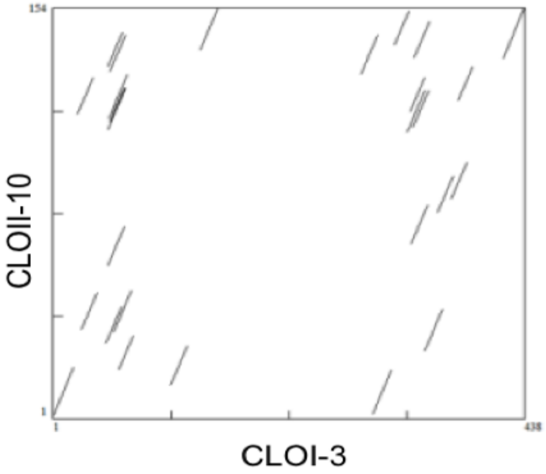
A


438
1
CLOI-3
1
438
CLOI-3

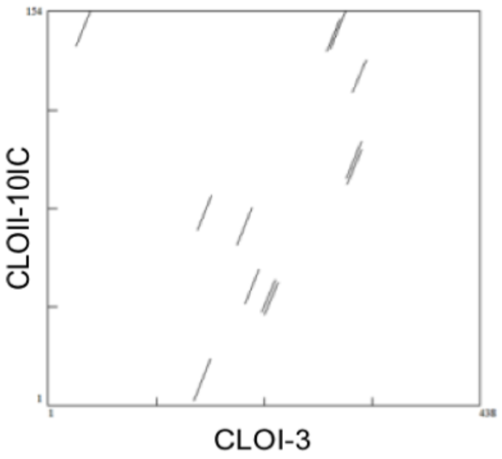
B


438
1
CLOI-3IC
1
438
CLOI-3

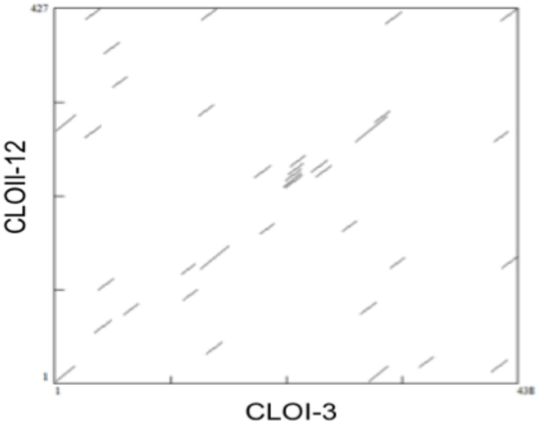
C


154
1
CLOII-10
1
438
CLOI-3

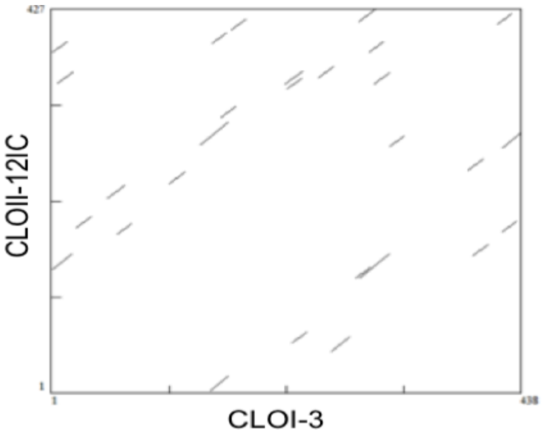
D

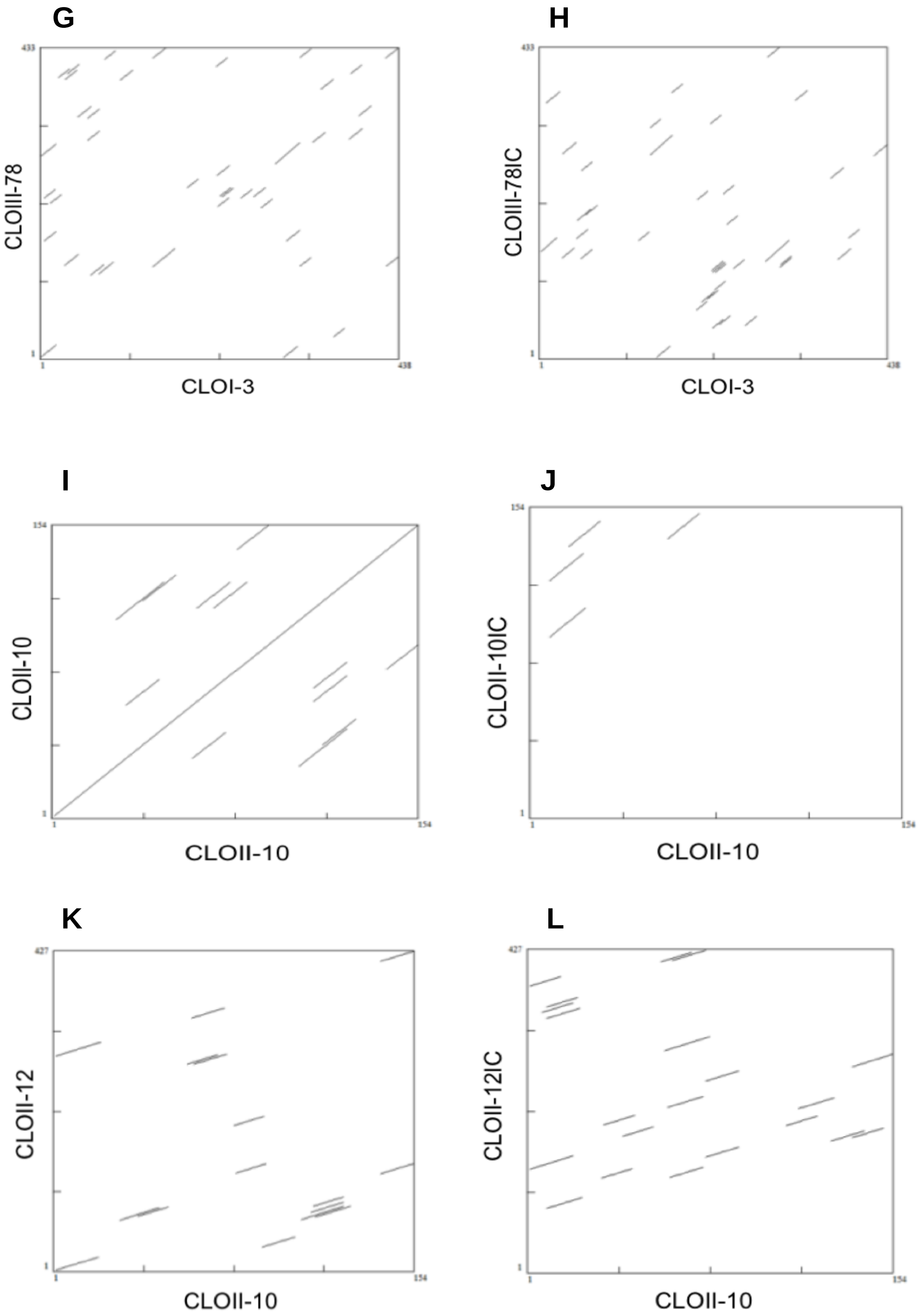

154
1
CLOII-10IC
1
438
CLOI-3

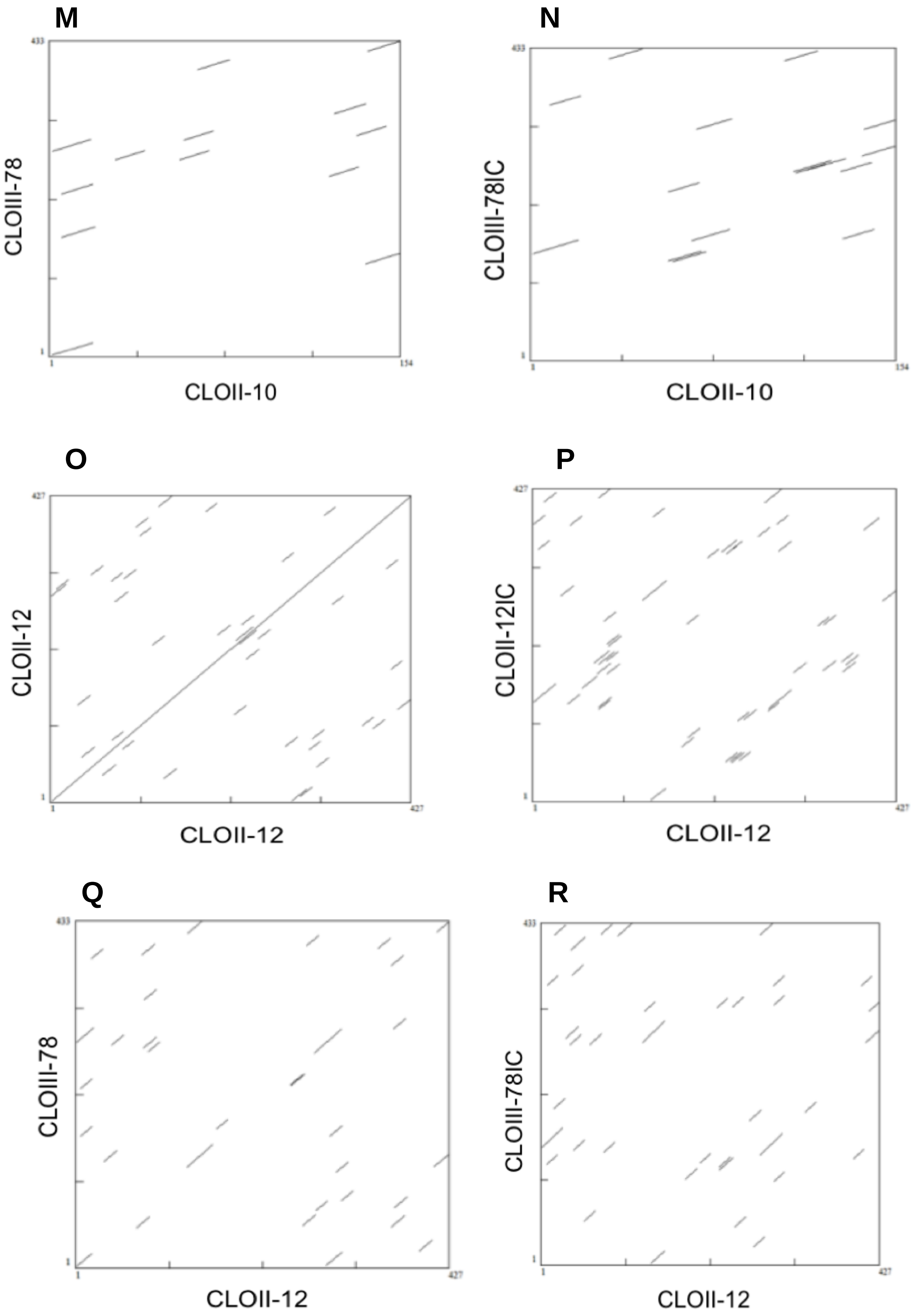
E

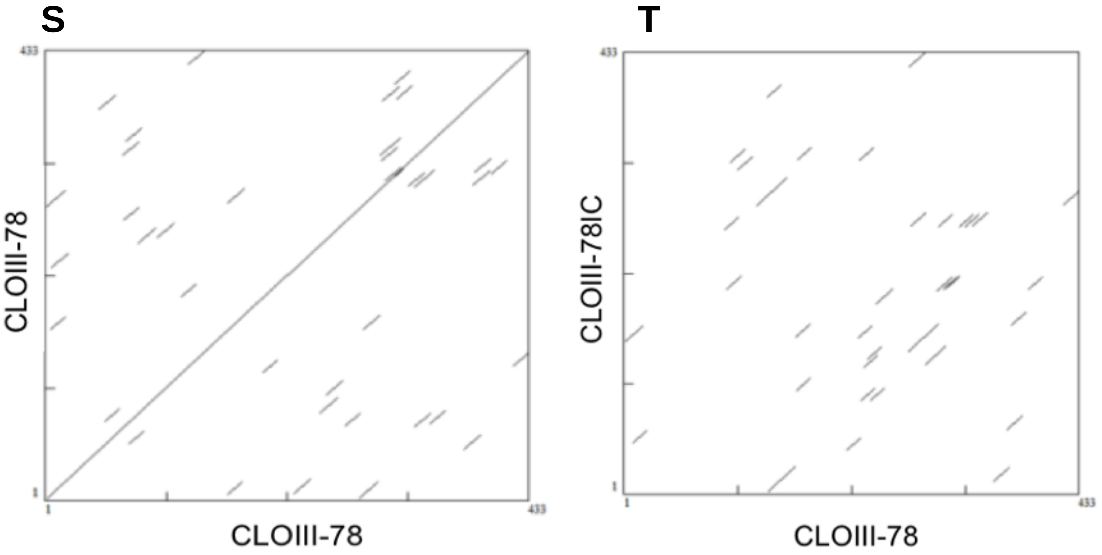

427
1
CLOII-12
1
438
CLOI-3

F


427
1
CLOII-12IC
1
438
CLOI-3


G
433
1
1
438
CLOIII-78
CLOI-3
H
433
1
1
438
CLOIII-78IC
CLOI-3
I
154
1
1
154
CLOII-10
CLOII-10
J
154
1
1
154
CLOII-10IC
CLOII-10
K
427
1
1
154
CLOII-12
CLOII-10
L
427
1
1
154
CLOII-12IC
CLOII-10


M
N
433
1
154
CLOIII-78
CLOII-10
CLOIII-78IC
CLOII-10
O
P
427
1
427
CLOII-12
CLOII-12
CLOII-12IC
CLOII-12
Q
R
433
1
427
CLOIII-78
CLOII-12
CLOIII-78IC
CLOII-12


S
433
1
CLOIII-78
1
433
CLOIII-78
T
433
1
CLOIII-78IC
1
433
CLOIII-78

Examples of PCR Gels of Putative Insulator Sequences

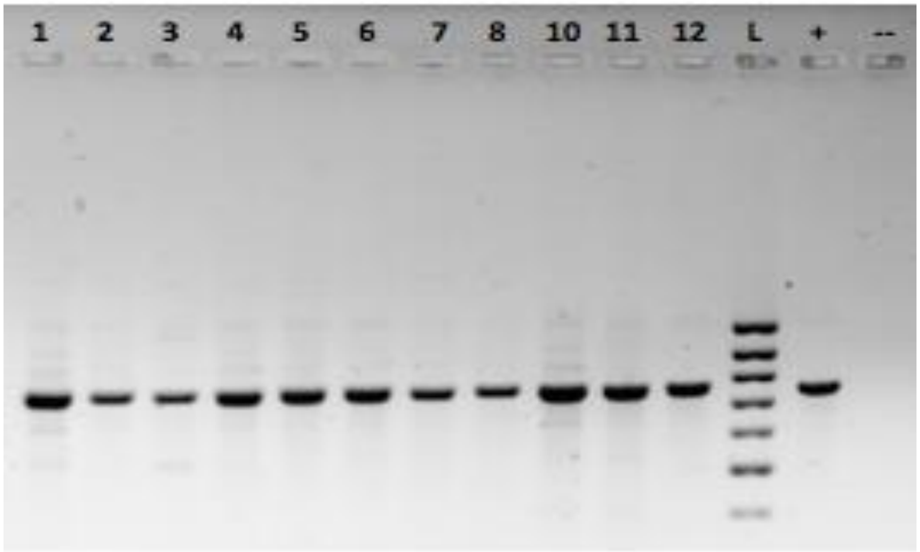
Figure 3. Example of agarose gel electrophoresis showing PCR amplification of CLOI-3, CLOII-10, CLOII-12, CLOI-3Sub1, CLOI-3Sub2, and CLOI-3Sub3. Transgenic *A. thaliana* DNA samples #1-12 containing the respective candidate insulator are shown. The expected size of ~750bp using SALK primers (**A**) and ~900bp using 1300LacZF/GUS5Rev primers (**B**) is shown. All gels include the Fermentas MassRuler Low Range DNA Ladder labeled as '**L**'.

A: + = positive control (wild-type *A. thaliana* ecotype Columbia)
-- = negative control (No DNA)

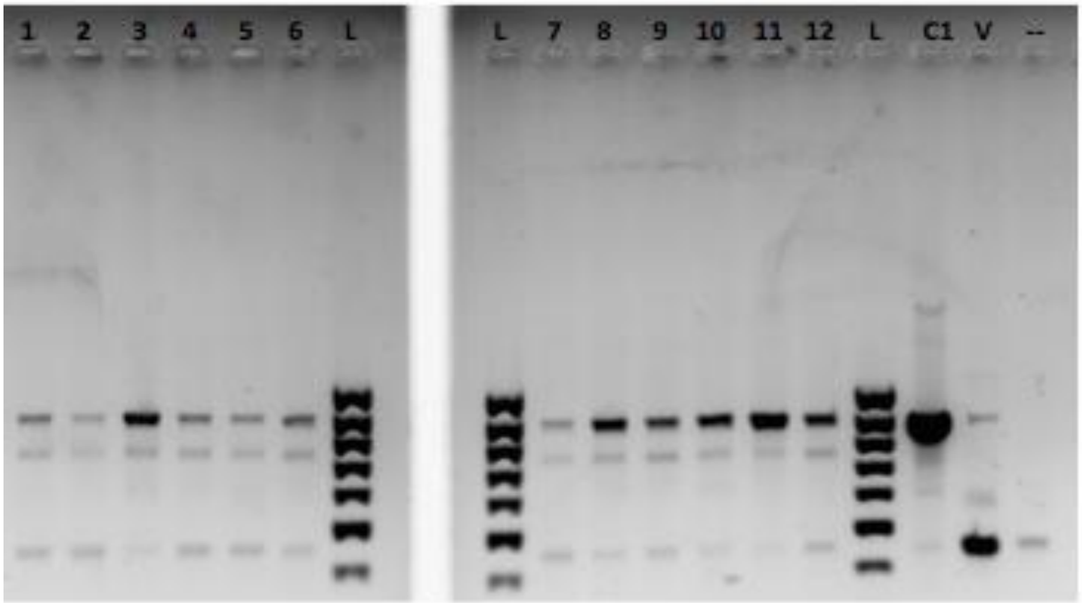
B: + = positive control (CLOI-3 DNA in pB31 (~900bp) labeled '**C1**', pB31 (only vector ~450bp) labeled '**V**', and CLOI-3Δ8 (insert only ~150bp) labeled '**C2**'. Not all positive controls listed are seen in every gel. At least one was used for each gel below.
-- = negative control (No DNA)

CLOI-3 Replicate 1LR:

A

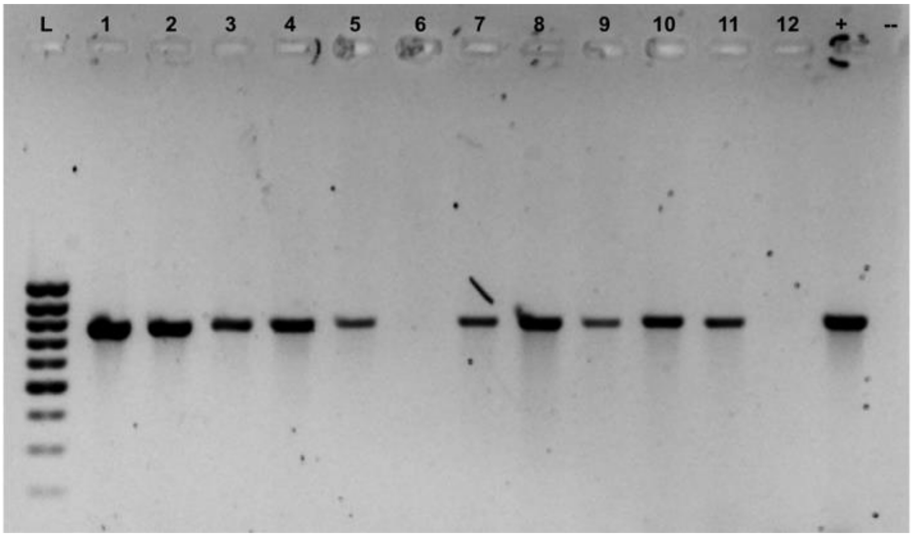


B

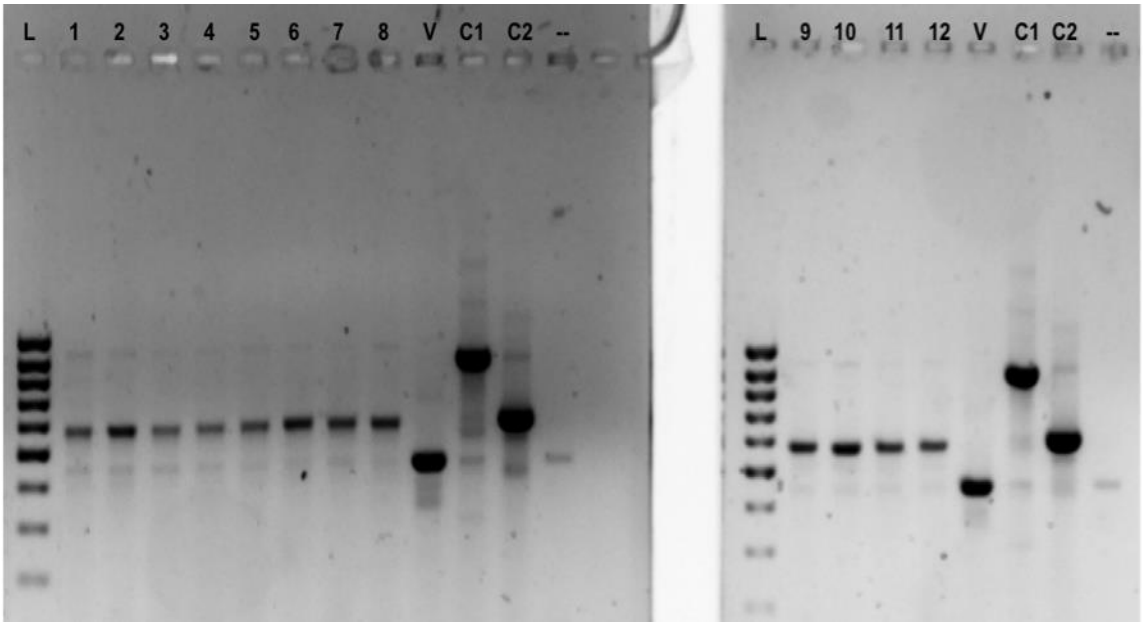


CLO-10 Replicate 1LR:

A



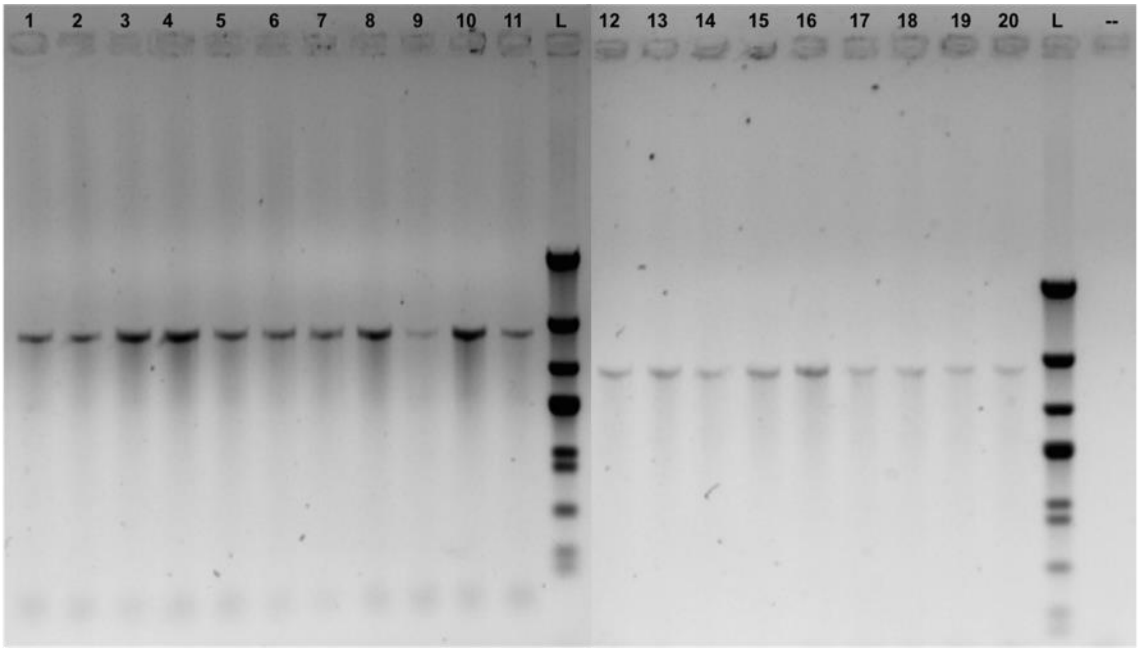
B



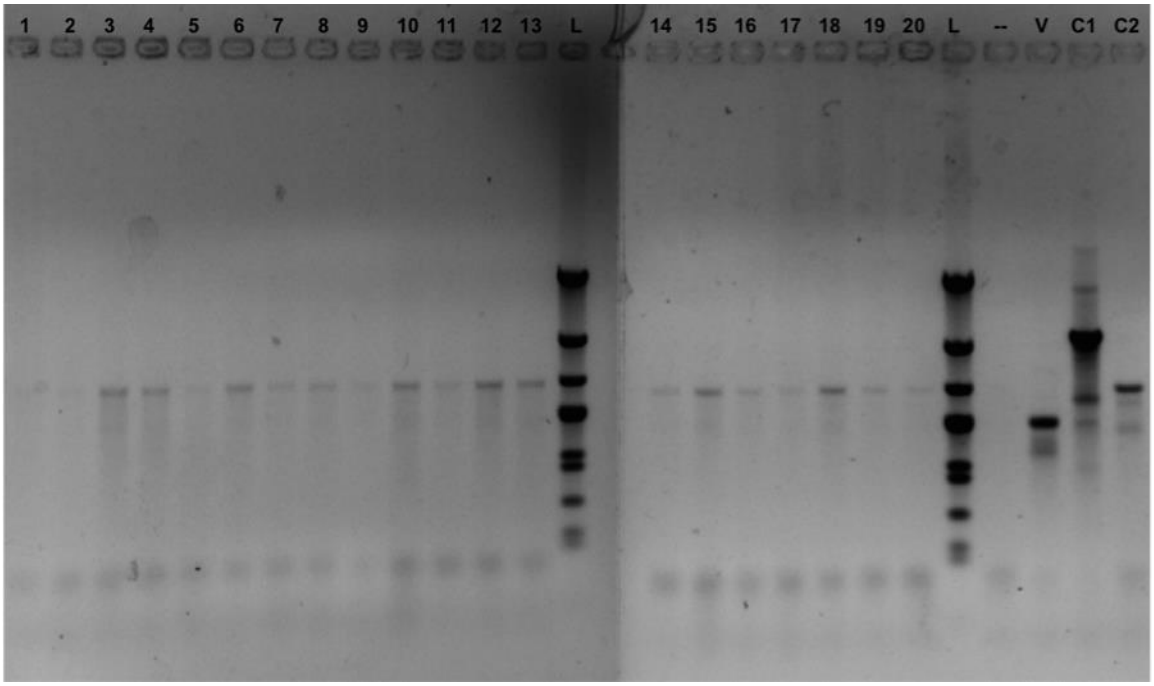
CLO-10 Replicate 2LR:

L = pPhiXHaeIII

A

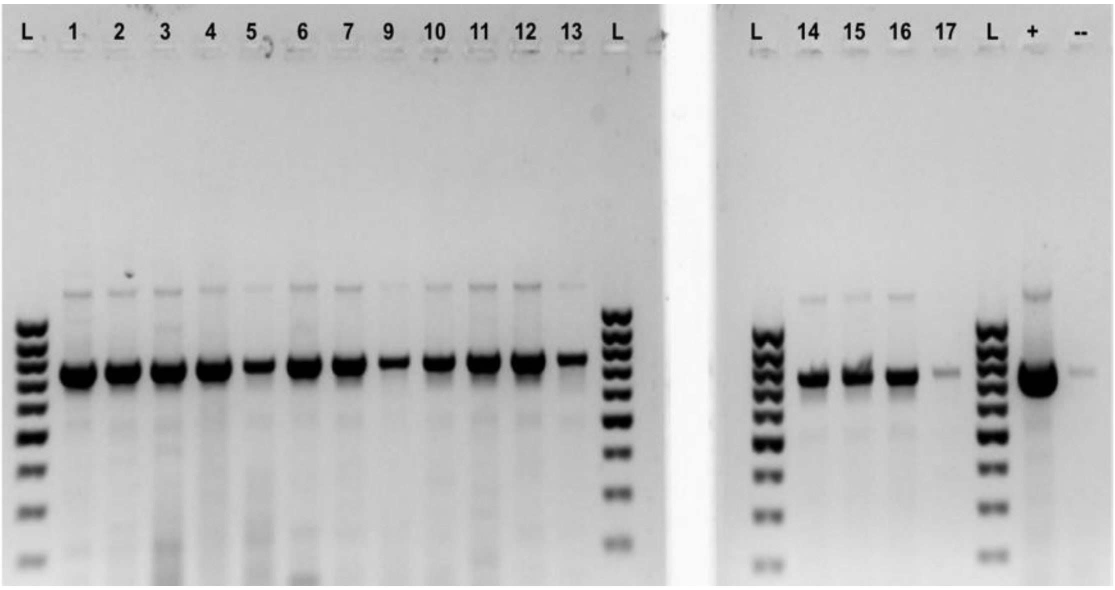


B

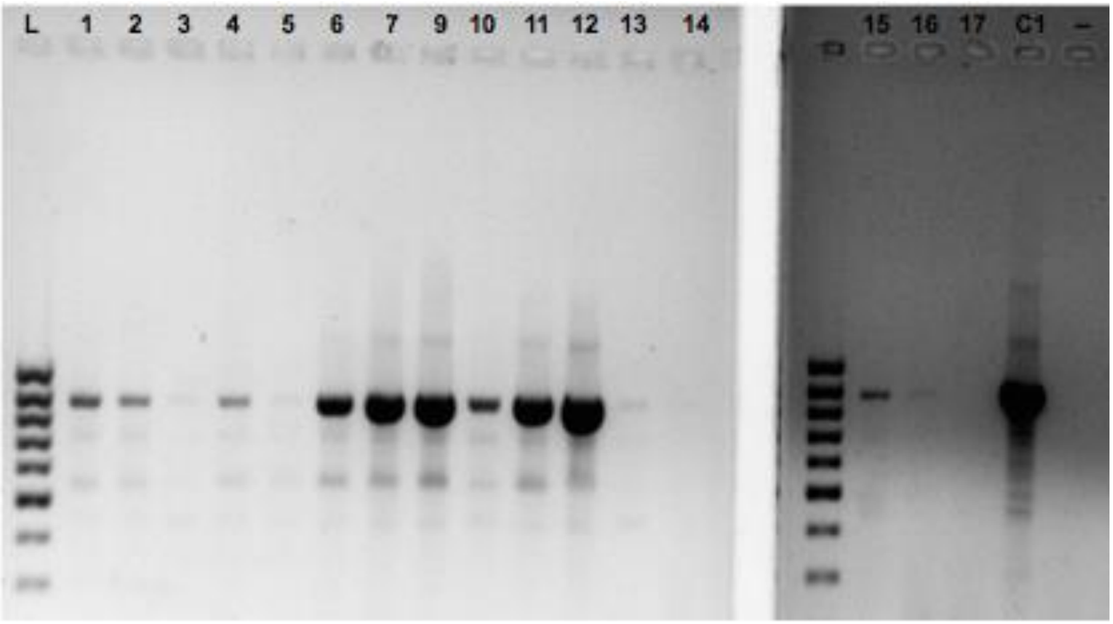


CLO-12 Replicate 1LR:

A

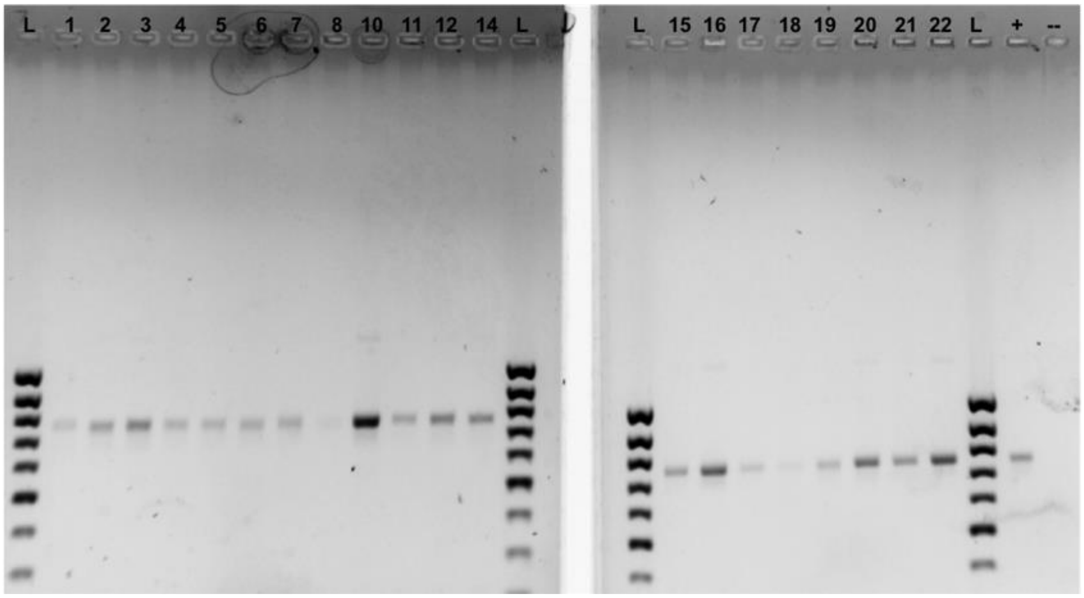


B

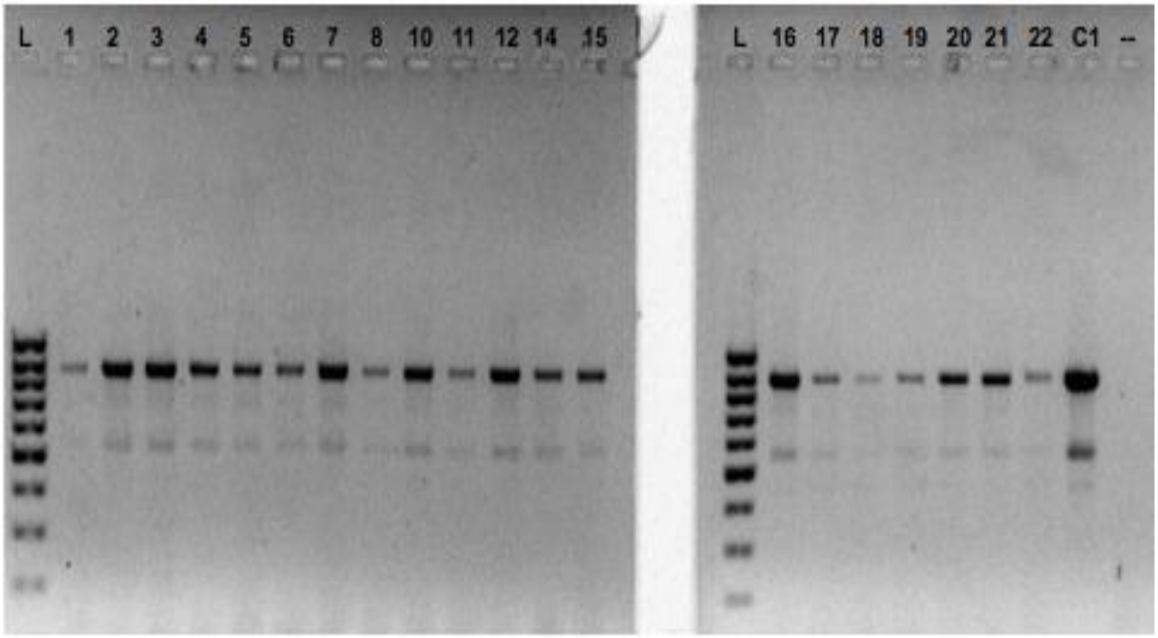


CLO-12 Replicate 2LR:

A

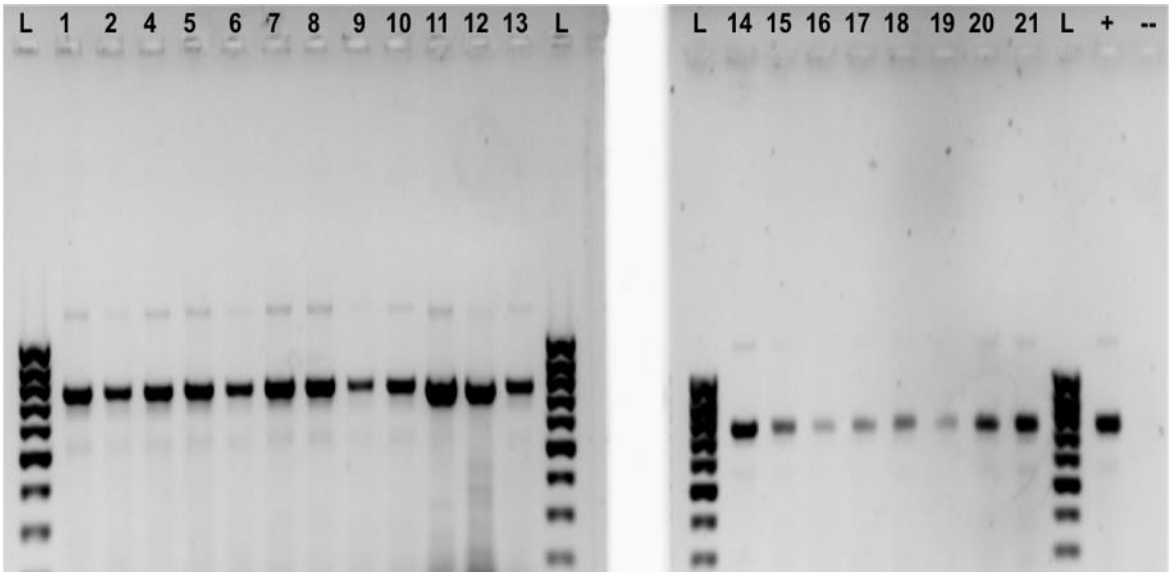


B

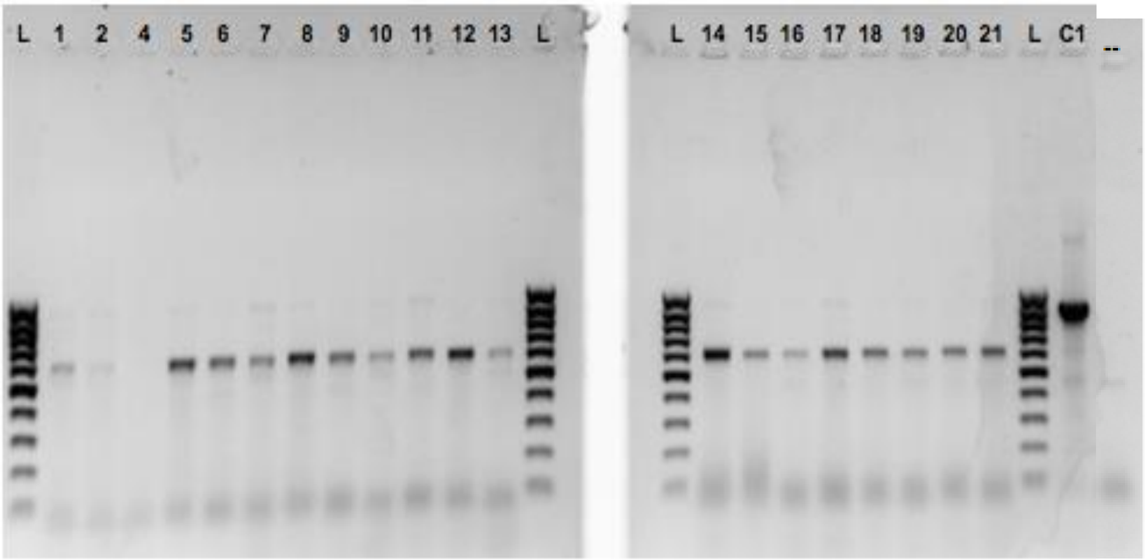


CLOI-3Sub1 Replicate 1LR:

A

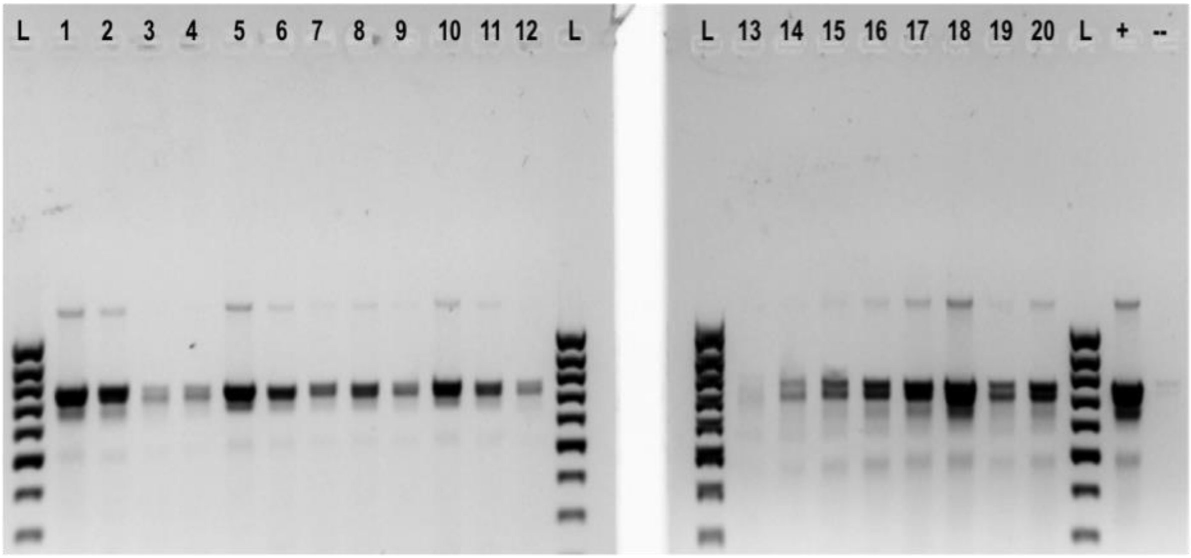


B

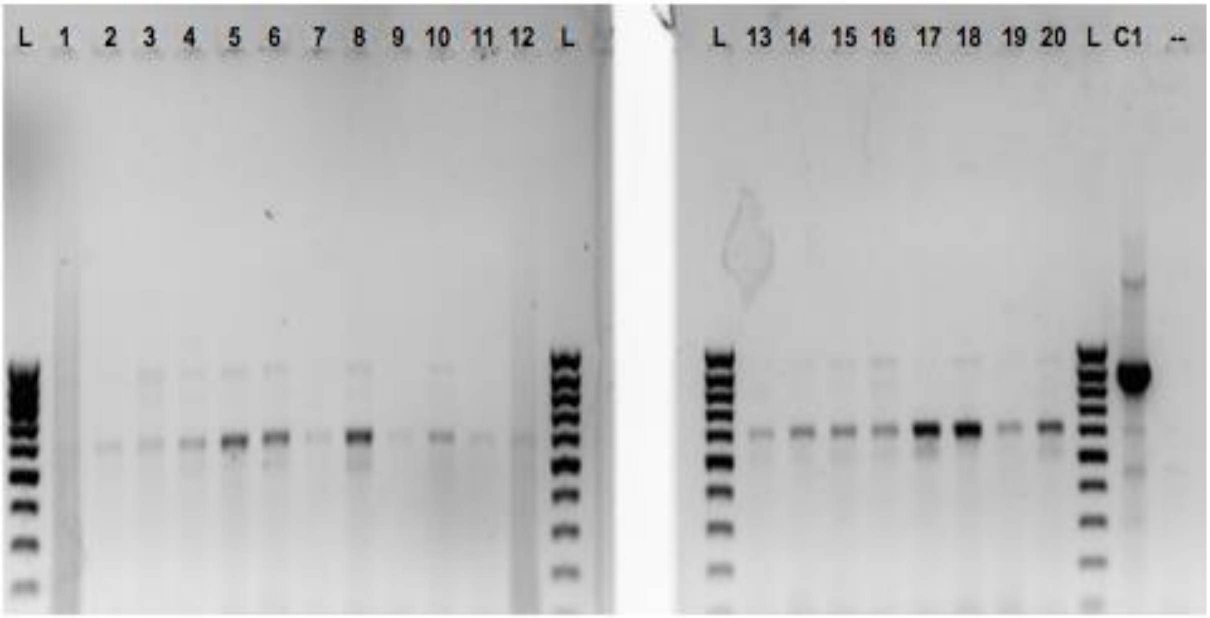


CLOI-3Sub2 Replicate 1LR:

A

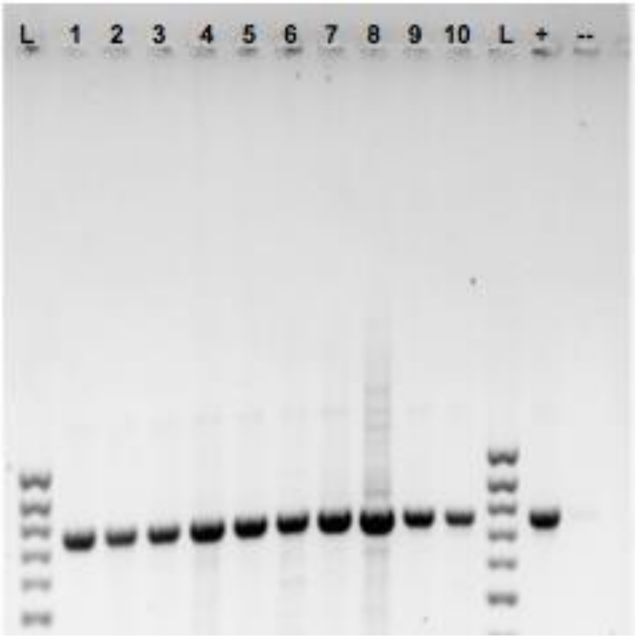


B

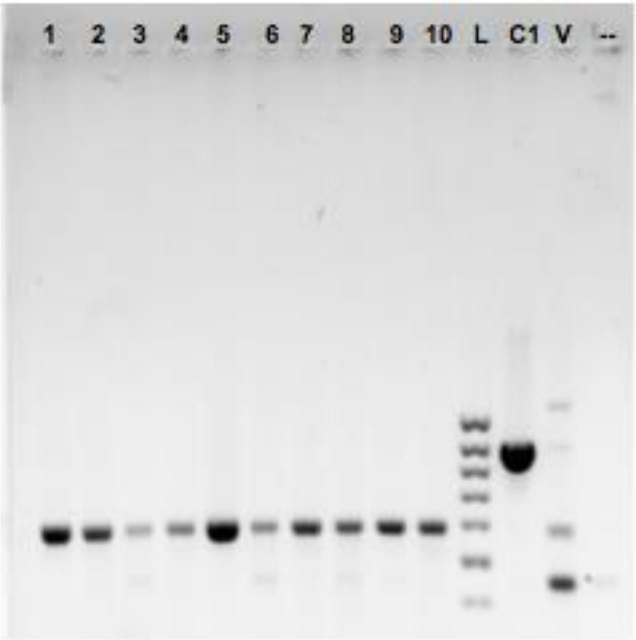


CLOI-3Sub3 Replicate 1LR:

A



B



Cloning Outlines for Selected Potential Insulators

A: DNA Cloning of Long Sequences (~450bp)

Long sequences (~450bp) are difficult to clone because of the presence of internal restriction sites (see **Figure 13, Batool Gandorah, 2012**). We had hypothesized that they were formed by the ligation of three 150bp fragments, each beginning with an *Eco*RI site and ending with a *Bam*HI site, prior to ligation into the vector pB31.

Hence, in order to mediate cloning of long sequences, we investigated the strategy of adding new restriction sites by PCR. *Mun*I and *Bgl*II restriction enzymes with the restriction sites were respectively investigated. When these sites are cut they produce “sticky ends” that are complimentary to *Eco*RI and *Bam*HI sites in pB31. To validate this approach we used the sequence CLOIII-35 as follows:

1. The insert in CLOIII-35 was amplified by PCR using the primer pair (where **bold** indicates the important restriction sites).

CLO-F-*Bgl*II 5'-CGTTCT**AGATCT**CAGAGGATCCGAGACAAGC-3' (31mer)

CLO-R-*Mun*I 5’-AGTTCT**CAATTG**TCGACGAATTCAGGAGGAGG-3' (30mer)

2. The PCR fragment was cloned into pGEM-T using T/A cloning to amplify the amount of CLOIII-35 DNA and sequenced to confirm the fidelity of the cloning.

3. Finally, the CLOIII-35 insert was separated from pGEM-T by digestion with *Bgl*II and *Mun*I and ligated into pB31 that had been digested with *Bam*HI and *Eco*RI. Again sequencing was used to confirm the fidelity of the cloning.

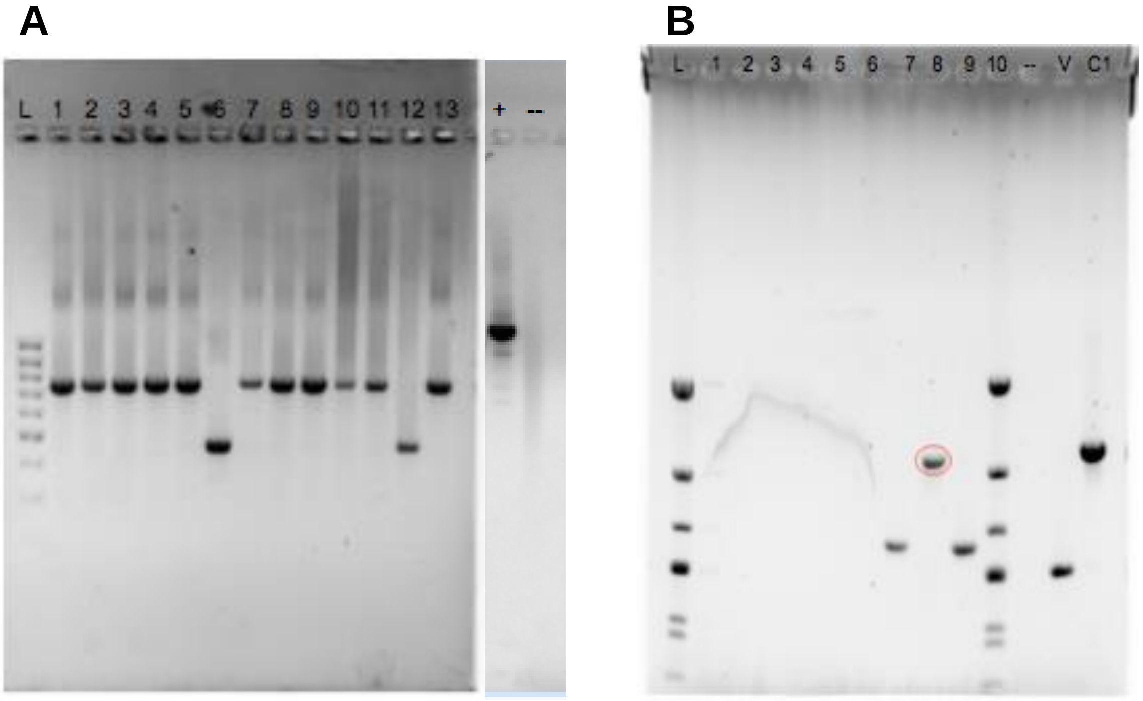
Figure 4. Example of agarose gel electrophoresis showing PCR amplification of CLOIII-35 (511bp) in two respective plasmid vectors. This sequence was first cloned into pGEM-T to amplify the DNA (**A**). The sequence was subsequently cloned into pB31 to identify insulator function in CLOIII-35 (**B** – Used with permission of Dr. Douglas A. Johnson).

A: + = positive control (5S genomic DNA)
-- = negative control (No DNA)
L = Fermentas MassRuler Low Range DNA Ladder

B: + = positive control (CLOI-3 DNA in pB31 (~900bp) labeled '**C1**', and pB31 (only vector ~450bp) labeled '**V**'.

-- = negative control (No DNA)

L & lane 10 = plasmid pφX174 cut with HaeIII. the fragment sizes (in base pairs) are 1,353; 815; 587; 458/449/434; 398; 267; 174)



B: Subclones Generated from Candidate Insulator CLOI-3 (refer to Figure 17)

Three subclones on CLOI-3 of approximately equal size were generated using traditional methods.

i. Restriction digestion of CLOI-3 (cloned in pB31) with *EcoRI* followed by self-ligation results in the loss of CLOI-3 sequences corresponding to CLOI-3Sub1 plus CLOI-3Sub2. The resulting clone, CLOI-3Sub3, has CLOI-3 sequences in the same orientation as CLOI-3;

ii. PCR amplification of CLOI-3 with primers CLO-F-BPX and CLO-R-ESX generates sub-fragments of different sizes, which were gel purified and cloned into pGEM-T Easy. Following screening and sequencing, suitable subclones were chosen. For cloning into pB31, the chosen subclone was digested with *EcoRI* plus *BamHI* and ligated into pB31 digested with the same enzymes. CLOI-3Sub1 has CLOI-3 sequences in the same orientation as CLOI-3 while CLOI-3Sub2 has CLOI-3 sequences in the opposite orientation. In addition, CLOI-3Sub2 contains a single mutation indicated by an “X” in **Figure 17** where A was changed to a G.

Summary of Staining Results for Transgenic *Arabidopsis thaliana*

Table 1. GUS staining results of putative insulator sequences. GUS staining causes blue coloration of the plant tissue and was scored based on the intensity of staining.
Strong – 3, Medium – 2, and Weak – 1, as described in **section 2.4.4** (Gudynaite-Savitch, 2009). Stained tissue is highlighted in red for visual representation. NS = No Sample.

Construct	**Plant #**	**Flower**	**Leaf**	**Silique**	**Silique 2**
pB31: pCAM 1300-35S46-*GUS*	1	3	3	3	NS
	2	0	0	0	NS
	3	3	0	2	NS
	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	NS
	12	0	2	1	NS
	13	0	1	0	NS
	14	0	2	2	NS
	15	0	1	0	NS
	16	0	3	2	NS
	19	0	1	1	NS
	20	0	1	1	NS
	21	0	1	1	NS
	22	0	0	0	NS
	23	0	1	1	NS
	24	0	0	0	NS
	25	0	0	0	NS
	26	0	2	3	NS
	27	0	1	1	NS
	28	2	3	3	NS
	29	0	1	0	NS
	30	0	0	0	NS
	31	1	3	3	NS
	32	0	1	0	NS
	33	3	3	3	NS
	34	0	1	1	NS
	35	0	1	0	NS
	36	0	2	2	NS

	37	1	3	3	NS
	38	1	3	3	NS
	39	2	3	3	NS
	40	0	0	0	NS
	41	1	3	3	NS
	42	0	0	0	NS
	43	0	3	0	NS
	44	0	3	3	NS
	45	0	3	3	NS
TOTAL n= 42		**12**	**30**	**27**	

Construct	**Plant #**	**Flower**	**Leaf**	**Silique**	**Silique 2**
CLOI-3/pB31 = 438bp	1	0	0	0	0
First Replicate	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
	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
	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
	52	0	0	0	0
	53	0	0	0	0
TOTAL n= 51		**0**	**0**	**0**	**0**

Construct	Plant #	Flower	Leaf	Silique	Silique 2
CLOI-3Sub1/pB31 = ~150bp	1	0	1	0	1
First Replicate	2	0	0	0	1
	4	0	0	0	0
	5	0	0	0	0
	6	0	0	0	0
	7	0	0	0	0
	8	0	1	0	2
	9	1	1	1	1
	10	1	1	1	0
	11	0	1	1	1
	12	0	0	0	0
	13	1	0	0	1
	14	0	0	0	0
	15	0	1	0	0
	16	0	0	0	0
	17	0	0	0	0
	18	0	0	0	0
	19	0	1	1	1
	20	0	1	1	1
	21	1	2	1	1
	22	0	0	0	0
	23	0	1	1	1
	24	0	0	0	0
	25	0	1	1	1
	26	0	0	0	0
	27	0	1	0	1
	28	1	3	1	2
	29	0	1	1	1
	30	0	0	0	0
	31	0	1	0	1
	32	0	0	0	1
	33	0	1	0	1
	34	0	1	0	0

	35	0	2	1	2
	36	0	0	0	0
	37	0	1	0	1
	38	0	3	3	1
	39	0	0	0	0
	40	2	3	2	2
	41	0	3	1	1
	42	0	0	0	0
	43	0	2	1	1
	44	2	3	3	2
	45	0	3	1	2
	46	0	1	0	0
	47	0	1	1	1
	48	0	1	0	0
	49	0	2	1	1
	50	0	1	1	1
	51	0	0	0	0
	52	1	2	1	1
	53	0	2	1	1
	54	0	1	0	1
	55	1	3	2	2
	56	0	0	0	0
	58	0	2	1	1
	59	0	0	0	0
	60	0	0	0	0
	61	0	0	0	0
	62	0	1	1	1
	63	0	0	0	0
	64	0	1	0	0
	65	0	1	1	0
	66	0	2	1	1
	67	0	2	1	1
	68	0	2	1	1
	69	0	0	0	0

	70	0	2	1	1
	71	0	0	0	0
	72	1	3	2	1
	73	0	3	0	1
	74	1	3	2	2
	75	0	0	0	0
	76	1	1	0	2
	77	0	1	1	1
	78	0	0	0	0
	79	0	0	0	0
	80	0	0	0	0
	81	3	3	1	3
	82	0	2	1	1
	83	0	0	0	0
	89	0	0	0	0
TOTAL n= 82		**13**	**49**	**35**	**45**

Construct	**Plant #**	**Flower**	**Leaf**	**Silique**	**Silique 2**
CLOI-3Sub2/pB31 = ~150bp	1	0	1	1	1
First Replicate	2	0	0	0	0
	3	0	1	0	0
	4	0	1	0	1
	5	0	1	1	1
	6	0	1	1	1
	7	0	1	1	2
	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	1	1	1
	17	0	0	0	0
	18	0	0	0	NS
	19	3	3	1	1
	20	0	0	0	0
	21	0	0	0	0
	22	0	1	1	1
	23	0	0	0	0
	24	0	1	1	1
	25	0	1	1	1
	26	0	0	0	0
	27	1	3	1	1
	28	0	1	1	2
	29	0	1	1	1
	30	0	1	1	1
	31	0	2	2	1
	32	0	2	2	2
	33	0	0	0	0
	34	0	0	0	0
	35	0	0	0	0
	36	0	1	1	0
	37	0	1	1	0
	38	0	0	0	0
	39	0	0	0	0
TOTAL n= 39		**2**	**19**	**17**	**16**

Construct	**Plant #**	**Flower**	**Leaf**	**Silique**	**Silique 2**
CLOI-3Sub3/pB31 = ~150bp	1	0	0	0	NS
(1st Replicate)	2	0	0	0	NS
	3	0	1	1	NS
	4	0	0	0	0
	5	0	0	0	0
	6	0	1	1	1

	7	1	2	3	3
	8	0	1	1	2
	9	0	0	0	0
	10	0	0	0	NS
TOTAL n= 10		**1**	**4**	**4**	

Construct	**Plant #**	**Flower**	**Leaf**	**Silique**	**Silique 2**
CLOI-3Sub3/pB31 = ~150bp	1	0	1	0	0
(2nd Replicate)	2	0	2	1	0
	3	0	0	0	0
	4	0	1	1	0
	5	0	0	0	0
	6	0	0	0	2
	7	0	1	1	1
	8	0	2	1	1
	9	0	1	0	0
	10	0	1	0	0
	11	0	0	0	0
	12	0	2	1	0
	13	0	0	0	0
	14	0	0	0	0
	15	0	0	0	0
	16	0	3	1	1
	17	0	0	0	0
	18	0	0	0	0
	19	0	0	0	0
	20	0	0	0	0
	21	0	1	0	0
	23	0	1	0	0
	24	0	1	0	0
	25	0	0	0	0
	26	0	0	0	0
	27	0	0	0	0

	28	0	2	0	0
	29	0	0	0	0
	30	0	0	0	0
	31	0	1	0	0
	32	0	1	0	0
	33	0	1	0	1
	34	0	2	1	1
	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	2	0	1
	41	0	1	0	0
	43	0	1	0	0
	44	0	2	0	1
	45	0	0	0	0
	46	0	1	0	0
	47	0	1	0	0
	48	0	1	0	0
	49	0	0	0	0
	51	0	0	0	0
	52	0	0	0	0
	53	0	1	0	0
	54	0	1	1	0
	55	0	1	0	0
	56	0	0	0	0
	57	0	0	0	0
	58	0	1	0	0
	59	0	0	0	0
	60	0	1	0	0
	63	0	0	0	0
	66	0	0	0	0
	69	0	0	0	0

	71	0	0	0	0
	72	0	0	0	0
	73	0	1	0	0
	74	0	1	0	0
	75	0	0	0	0
	76	0	1	0	0
	77	0	0	0	0
	78	0	1	0	0
	79	0	0	0	0
	80	0	0	0	0
	81	0	2	1	2
	82	0	0	0	0
	85	0	0	0	0
	87	0	1	1	1
	88	0	0	0	0
TOTAL n= 75		**0**	**35**	**10**	**10**

Construct	Plant #	Flower	Leaf	Silique	Silique 2
CLOII-10/pB31 = 154bp	1	0	0	NS	NS
First Replicate	2	0	0	NS	NS
	3	0	0	NS	NS
	4	0	0	NS	NS
	5	0	0	NS	NS
	6	0	0	NS	NS
	7	0	0	NS	NS
	8	0	0	NS	NS
	10	0	0	NS	NS
	11	0	0	NS	NS
	12	0	0	NS	NS
	13	0	0	NS	NS
	14	0	0	NS	NS
TOTAL n= 14		**0**	**0**	**NS**	**NS**

Construct	Plant #	Flower	Leaf	Silique	Silique 2
CLOII-10/pB31 = 154bp	1	0	0	0	0
Second Replicate	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
	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 n= 20		**0**	**0**	**0**	**0**

Construct	Plant #	Flower	Leaf	Silique	Silique 2
CLOII-12/pB31 = 427bp	1	0	0	0	0
First Replicate	3	0	0	0	1
	4	0	0	0	0
	5	0	0	0	0
	6	0	0	0	0
	7	0	0	0	0
	9	0	0	0	0
	10	0	0	0	0

	11	0	0	0	0
	12	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
TOTAL n= 17		**0**	**0**	**0**	**1**

Construct	**Plant #**	**Flower**	**Leaf**	**Silique**	**Silique 2**
CLOII-12/pB31 = 427bp	1	0	0	0	0
Second Replicate	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
	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
	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
TOTAL n= 36		**0**	**0**	**0**	**0**

Table 2. GUS staining results of unsuccessful insulator sequences (A). GUS staining results of CLOI-3 and its subclones (B); graphical representation is present in Figures 18-21. GUS staining causes blue coloration of the plant tissue and was scored based on the intensity of staining. Strong – 3, Medium – 2, and Weak – 1, as described in **section 2.4.4** (Gudynaite-Savitch, 2009). This analysis was used to compare staining trends in different tissue types. By comparing sequences that do not have insulator activity with potential insulator candidates such as our subclones that do have GUS staining will allow us to determine patterns of tissue specificity, as there is variability in the numbers. Further analysis can be found in the Discussion section.

A

Inserts in pB31 vector	N	Flower	Leaf	Silique
CLOII-3	14	50%	36%	50%
CLOII-7	30	37%	37%	17%
CLOIII-4	28	64%	14%	0%
CLOIII-17	30	83%	27%	1%
CLOIII-22	36	72%	28%	1%
CLOIII-27	13	100%	23%	15%
CLOIII-43	13	23%	31%	38%
CLOIII-53	15	87%	93%	87%
CLOIII-55	24	96%	10%	1%
CLOIII-57	16	1%	41%	50%
CLOIII-63	7	29%	29%	71%
CLOIII-74	6	50%	50%	50%

B

Inserts in pB31 vector	N	Flower	Leaf	Silique
CLOI-3	51	0%	0%	0%
CLOI-3Sub1	82	16%	60%	55%
CLOI-3Sub2	39	0%	45%	44%
CLOI-3Sub3	10	10%	40%	40%
CLOI-3Sub3	75	0%	47%	13%

The Search for Known Repressor Sequences in Plants

Database: PLACE (http://www.dna.affrc.go.jp/PLACE/)

Date Searched: February 19, 2014

Second Database: PlantCARE
(http://bioinformatics.psb.ugent.be/webtools/plantcare/html/)

Date Searched: April 9, 2014

Repressor Name: WBOXNTERF3

Target Sequence: TGAC(C/T)

Found in Sequences: CLOI-3, CLOI-3Sub3, CLOII-10, CLOII-12, CLOI-2 CLOII-3, CLOIII-4, CLOIII-22, CLOIII-27, CLOIII-52, CLOIII-53, CLOIII-55, CLOIII-57, CLOIII-58, CLOIII-63, CLOIII-74, CLOIII-80

Description: “W box” is a binding site of tobacco WRKY proteins in wound-activated transcription of the ERF3 gene .

Reference: Nishiuchi, T., Shinshi, H., Suzuki, K. (2004). Rapid and transient activation of transcription of the ERF3 gene by wounding in tobacco leaves: Possible involvement of NtWRKY’s and autorepression. *J Biol Chem, 279*, 55355-55361.

Database: PLACE (http://www.dna.affrc.go.jp/PLACE/)

Date Searched: February 19, 2014

Second Database: PlantCARE
(http://bioinformatics.psb.ugent.be/webtools/plantcare/html/)

Date Searched: April 9, 2014

Repressor Name: WRKY71OS

Target Sequence: TGAC

Found in Sequences: CLOI-3, CLOI-3Sub3, CLOII-10, CLOII-12, CLOI-2 CLOII-3, CLOIII-4, CLOIII-22, CLOIII-27, CLOIII-52, CLOIII-53, CLOIII-55, CLOIII-57, CLOIII-58, CLOIII-63, CLOIII-74, CLOIII-80

Description: A core of TGAC-containing W-box, binding site of rice OsWRKY71 proteins, a transcriptional repressor of the gibberellin signaling pathway.

Reference: Zhang Z.L., Xie Z., Zou X., Casaretto J., Ho T.H., Shen Q.J. (2004). A rice WRKY gene encodes a transcriptional repressor of the RT gibberellin signaling pathway in aleurone cells. *Plant Physiology, 134*, 1500-1513.

Database: PLACE (http://www.dna.affrc.go.jp/PLACE/)

Date Searched: February 19, 2014

Second Database: PlantCARE
(http://bioinformatics.psb.ugent.be/webtools/plantcare/html/)

Date Searched: April 9, 2014

Repressor Name: CACGTGMOTIF

Target Sequence: CACGTG

Found in Sequences: CLOII-12, CLOII-3, CLOIII-52

Description: ‘G-box’ binding site of Arabidopsis GBF4. Roseus G-box binding factor 1 (CrBGF1) and 1 (CrGBF2) can act as transcriptional repressors of the Str promoter via direct interaction with the G-box.

Reference: Chandrasekharan, M.B., Bishop, K.J., Hall, T.C. (2003). Module-specific regulation of the beta-phaseolin promoter during embryogenesis. *Plant J., 33*, 853-866.

Busk, R.A., Pages, M. (1998). Regulation of abscisic acid-induced transcription. *Plant Mol Biol, 37*, 425-435.

Hudson, M.E., Quail, P.H. (2003). Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. *Plant Physiol., 133*, 1605-1616.

Results from the database PLACE for each sequence searched

The PLACE database (http://www.dna.affrc.go.jp/PLACE/) was searched using the sequence of interest, resulting in a list of matches in the output as shown below under each sequence searched (CLOI-3, CLOII-10, CLOII-12 and CLOIII-78). Each individual match in the output was opened and the description given by the database was searched in regards to the type of cis-regulatory element. The output was narrowed down to the descriptions that indicated possible repressor activity. At this point the references provided by the database were read to confirm the repressor activity. The criteria used to choose these target sequences are that they are known repressor sites that have been identified in plants and match the result of the second database. Such matches are indicated in black bold.
The same approach was used for the second database PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Results from PlantCARE that match the PLACE output are indicated in red bold. Sites that were only found in PlantCARE are listed below the PLACE output and were also manually searched within each sequence to verify since discrepancies were found. All results are explained in **section 3.4** and elaborated on in the Discussion section.

CLOI-3:

```
RESULTS OF YOUR SIGNAL SCAN SEARCH REQUEST
438 base pairs Signal Database File: user.dat
SITE #  Factor or Site Name          Loc.(Str.)   Signal Sequence
_____________________________________________________________________________
S000392 ABRELATERD1                 site    41 (+) ACGTG
S000414 ABREMOTIFAOSOSEM            site    40 (+) TACGTGTC
S000299 ABRERATCAL                  site   365 (-) MACGYGB
S000507 ACGTABREMOTIFA2OSEM         site    41 (+) ACGTGKC
S000394 ACGTABREMOTIFAOSOSEM        site    40 (+) TACGTGTC
S000281 ACGTATERD1                  site    41 (+) ACGT
S000415 ACGTATERD1                  site    41 (-) ACGT
S000415 ARFAT                       site   142 (+) TGTCTC
S000270 ARFAT                       site   427 (+) TGTCTC
S000270 ARFAT                       site   154 (-) TGTCTC
S000270 BOXIINTPATPB                site   167 (+) ATAGAA
S000296 CAATBOX1                    site   313 (+) CAAT
S000028 CAATBOX1                    site    77 (-) CAAT
S000028 CACTFTPPCA1                 site   326 (+) YACT
S000449 CACTFTPPCA1                 site   163 (-) YACT
S000449 CACTFTPPCA1                 site   191 (-) YACT
S000449 CGCGBOXAT                   site   216 (+) VCGCGB
S000501 CGCGBOXAT                   site   396 (+) VCGCGB
S000501 CGCGBOXAT                   site   365 (+) VCGCGB
S000501 CGCGBOXAT                   site   216 (-) VCGCGB
S000501 CGCGBOXAT                   site   396 (-) VCGCGB
S000501 CGCGBOXAT                   site   365 (-) VCGCGB
S000501 CURECORECR                  site   417 (+) GTAC
S000493 CURECORECR                  site   417 (-) GTAC
S000493 DOFCOREZM                   site    95 (+) AAAG
S000265 DOFCOREZM                   site   171 (+) AAAG
S000265 DOFCOREZM                   site   267 (+) AAAG
S000265 DOFCOREZM                   site    47 (-) AAAG
S000265 DOFCOREZM                   site   128 (-) AAAG
S000265 DOFCOREZM                   site   390 (-) AAAG
```

```
S000265 E2FCONSENSUS                site    350 (+) WTTSSCSS
S000476 EBOXBNNAPA                  site    114 (+) CANNTG
S000144 EBOXBNNAPA                  site    114 (-) CANNTG
S000144 EECCRCAH1                   site    376 (+) GANTTNC
S000494 ELRECOREPCRP1               site    310 (-) TTGACC
S000142 GADOWNAT                    site     41 (+) ACGTGTC
S000438 GATABOX                     site     85 (-) GATA
S000039 GATABOX                     site    371 (-) GATA
S000039 GT1CONSENSUS                site    263 (+) GRWAAW
S000198 GT1CONSENSUS                site    264 (+) GRWAAW
S000198 GT1CONSENSUS                site    348 (-) GRWAAW
S000198 GT1CONSENSUS                site     49 (-) GRWAAW
S000198 GT1CONSENSUS                site    349 (-) GRWAAW
S000198 GT1CONSENSUS                site    391 (-) GRWAAW
S000198 GT1GMSCAM4                  site    264 (+) GAAAAA
S000453 GTGANTG10                   site    208 (-) GTGA
S000378 GTGANTG10                   site    343 (-) GTGA
S000378 IBOXCORE                    site    370 (-) GATAA
S000199 LTRECOREATCOR15             site    270 (-) CCGAC
S000153 MYB2AT                      site    131 (+) TAACTG
S000177 MYB2AT                      site    181 (+) TAACTG
S000177 MYB2CONSENSUSAT             site    131 (+) YAACKG
S000409 MYB2CONSENSUSAT             site    181 (+) YAACKG
S000409 MYBCORE                     site    134 (+) CNGTTR
S000176 MYBCORE                     site    131 (-) CNGTTR
S000176 MYBCORE                     site    181 (-) CNGTTR
S000176 MYCCONSENSUSAT              site    114 (+) CANNTG
S000407 MYCCONSENSUSAT              site    114 (-) CANNTG
S000407 NODCON1GM                   site    388 (-) AAAGAT
S000461 NTBBF1ARROLB                site     94 (-) ACTTTA
S000273 OSE1ROOTNODULE              site    388 (-) AAAGAT
S000467 P1BS                        site    230 (+) GNATATNC
S000459 P1BS                        site    230 (-) GNATATNC
S000459 POLLEN1LELAT52              site    169 (+) AGAAA
S000245 POLLEN1LELAT52              site     90 (-) AGAAA
S000245 PRECONSCRHSP70A             site    303 (-) SCGAYNRNNNNNNNNNNNNNNNHD
S000506 QELEMENTZMZM13              site    243 (+) AGGTCA
S000254 QELEMENTZMZM13              site    309 (+) AGGTCA
S000254 RAV1AAT                     site    135 (-) CAACA
S000314 RHERPATEXPA7                site    208 (+) KCACGW
S000512 ROOTMOTIFTAPOX1             site    231 (-) ATATT
S000098 S1FBOXSORPS1L21             site    177 (+) ATGGTA
S000223 SEBFCONSSTPR10A             site    141 (+) YTGTCWC
S000391 SEBFCONSSTPR10A             site    426 (+) YTGTCWC
S000391 SEBFCONSSTPR10A             site    154 (-) YTGTCWC
S000391 SEF3MOTIFGM                 site    330 (+) AACCCA
S000115 SORLIP1AT                   site     72 (-) GCCAC
S000482 SURECOREATSULTR11           site    154 (+) GAGAC
S000499 SURECOREATSULTR11           site    143 (-) GAGAC
S000499 SURECOREATSULTR11           site    428 (-) GAGAC
S000499 TAAAGSTKST1                 site     94 (+) TAAAG
S000387 TAAAGSTKST1                 site    128 (-) TAAAG
S000387 TAAAGSTKST1                 site    390 (-) TAAAG
S000387 WBOXATNPR1                  site    311 (-) TTGAC
S000390 WBOXHVISO1                  site    375 (+) TGACT
S000442 WBOXNTCHN48                 site    374 (+) CTGACY
S000508 WBOXNTCHN48                 site    244 (-) CTGACY
```

```
S000508 WBOXNTERF3                 site     375  (+)  TGACY
S000457 WBOXNTERF3                 site     244  (-)  TGACY
S000457 WBOXNTERF3                 site     310  (-)  TGACY
S000457 WRKY71OS                   site     375  (+)  TGAC
S000447 WRKY71OS                   site     245  (-)  TGAC
S000447 WRKY71OS                   site     311  (-)  TGAC
```

Results on search of PlantCARE that were additional to those found in PLACE:

Name of Motif:	Signal Sequence:
TATA-box	TATAAAT

CLOII-10:

```
RESULTS OF YOUR SIGNAL SCAN SEARCH REQUEST
154 base pairs Signal Database File: user.dat
SITE #   Factor or Site Name          Loc.(Str.)    Signal Sequence
__________________________________________________________________________
S000270 CAATBOX1                   site      27  (-)  CAAT
S000028 CATATGGMSAUR               site      16  (+)  CATATG
S000370 CATATGGMSAUR               site      16  (-)  CATATG
S000370 CGACGOSAMY3                site     123  (-)  CGACG
S000205 DOFCOREZM                  site      43  (-)  AAAG
S000265 DOFCOREZM                  site      89  (-)  AAAG
S000265 DPBFCOREDCDC3              site      76  (-)  ACACNNG
S000292 EBOXBNNAPA                 site      16  (+)  CANNTG
S000144 EBOXBNNAPA                 site      16  (-)  CANNTG
S000144 ELRECOREPCRP1              site      28  (+)  TTGACC
S000142 GATABOX                    site      82  (-)  GATA
S000039 MYBPZM                     site     130  (+)  CCWACC
S000179 MYCCONSENSUSAT             site      16  (+)  CANNTG
S000407 MYCCONSENSUSAT             site      16  (-)  CANNTG
S000407 PRECONSCRHSP70A            site     105  (-)  SCGAYNRNNNNNNNNNNNNNNNHD
S000506 SEBFCONSSTPR10A            site     142  (+)  YTGTCWC
S000391 SURECOREATSULTR11          site     144  (-)  GAGAC
S000499 WBOXATNPR1                 site      28  (+)  TTGAC
S000390 WBOXNTERF3                 site      29  (+)  TGACY
S000457 WRKY71OS                   site      29  (+)  TGAC
```

Results on search of PlantCARE that were additional to those found in PLACE:

Name of Motif:	Signal Sequence:
None	None

CLOII-12:

```
RESULTS OF YOUR SIGNAL SCAN SEARCH REQUEST
427 base pairs Signal Database File: user.dat
SITE #   Factor or Site Name          Loc.(Str.)    Signal Sequence
_________________________________________________________________________
S000392 -10PEHVPSBD               site     254  (-)  TATTCT
S000414 ABRELATERD1               site      105  (+)  ACGTG
S000414 ABRELATERD1                site     386  (+)  ACGTG
S000414 ABRELATERD1                site     104  (-)  ACGTG
```

```
S000414 ABRERATCAL                  site    104 (+) MACGYGB
S000507 ABRERATCAL                  site    103 (-) MACGYGB
S000507 ACGTABOX                    site     59 (+) TACGTA
S000130 ACGTABOX                    site     59 (-) TACGTA
S000130 ACGTATERD1                  site     60 (+) ACGT
S000415 ACGTATERD1                  site    105 (+) ACGT
S000415 ACGTATERD1                  site    386 (+) ACGT
S000415 ACGTATERD1                  site     60 (-) ACGT
S000415 ACGTATERD1                  site    105 (-) ACGT
S000415 ACGTATERD1                  site    386 (-) ACGT
S000415 ARFAT                       site    133 (+) TGTCTC
S000270 ARFAT                       site    416 (+) TGTCTC
S000270 ARFAT                       site    145 (-) TGTCTC
S000270 ARR1AT                      site    397 (+) NGATT
S000454 ARR1AT                      site    324 (+) NGATT
S000454 ARR1AT                      site     71 (-) NGATT
S000454 ARR1AT                      site    259 (-) NGATT
S000454 CAATBOX1                    site     54 (-) CAAT
S000028 CAATBOX1                    site    322 (-) CAAT
S000028 CACGTGMOTIF                 site    104 (+) CACGTG
S000042 CACGTGMOTIF                 site    104 (-) CACGTG
S000042 CACTFTPPCA1                 site    362 (+) YACT
S000449 CACTFTPPCA1                 site    116 (-) YACT
S000449 CAREOSREP1                  site     96 (+) CAACTC
S000421 CGCGBOXAT                   site    269 (+) VCGCGB
S000501 CGCGBOXAT                   site    269 (-) VCGCGB
S000501 CIACADIANLELHC              site    325 (-) CAANNNNATC
S000252 CMSRE1IBSPOA                site    122 (+) TGGACGG
S000511 CURECORECR                  site     15 (+) GTAC
S000493 CURECORECR                  site    275 (+) GTAC
S000493 CURECORECR                  site     15 (-) GTAC
S000493 CURECORECR                  site    275 (-) GTAC
S000493 DOFCOREZM                   site     19 (-) AAAG
S000265 DOFCOREZM                   site    110 (-) AAAG
S000265 DOFCOREZM                   site    356 (-) AAAG
S000265 DOFCOREZM                   site    374 (-) AAAG
S000265 EBOXBNNAPA                  site    104 (+) CANNTG
S000144 EBOXBNNAPA                  site    181 (+) CANNTG
S000144 EBOXBNNAPA                  site    188 (+) CANNTG
S000144 EBOXBNNAPA                  site    104 (-) CANNTG
S000144 EBOXBNNAPA                  site    181 (-) CANNTG
S000144 EBOXBNNAPA                  site    188 (-) CANNTG
S000144 EECCRCAH1                   site     45 (+) GANTTNC
S000494 EECCRCAH1                   site    325 (+) GANTTNC
S000494 EECCRCAH1                   site    398 (+) GANTTNC
S000494 GATABOX                     site    351 (-) GATA
S000039 GT1CONSENSUS                site    326 (-) GRWAAW
S000198 GT1CONSENSUS                site    399 (-) GRWAAW
S000198 GTGANTG10                   site    159 (+) GTGA
S000378 GTGANTG10                   site    316 (-) GTGA
S000378 GTGANTG10                   site    359 (-) GTGA
S000378 IRO2OS                      site    103 (-) CACGTGG
S000505 MYBCORE                     site    215 (+) CNGTTR
S000176 MYBPZM                      site    216 (-) CCWACC
S000179 MYCCONSENSUSAT              site    104 (+) CANNTG
S000407 MYCCONSENSUSAT              site    181 (+) CANNTG
S000407 MYCCONSENSUSAT              site    188 (+) CANNTG
```

```
S000407 MYCCONSENSUSAT            site    104 (-) CANNTG
S000407 MYCCONSENSUSAT            site    181 (-) CANNTG
S000407 MYCCONSENSUSAT            site    188 (-) CANNTG
S000407 NODCON2GM                 site    354 (+) CTCTT
S000462 NODCON2GM                 site    364 (+) CTCTT
S000462 OSE2ROOTNODULE            site    354 (+) CTCTT
S000468 OSE2ROOTNODULE            site    364 (+) CTCTT
S000468 PALBOXAPC                 site    123 (-) CCGTCC
S000137 POLASIG3                  site    256 (+) AATAAT
S000088 POLLEN1LELAT52            site     20 (-) AGAAA
S000245 POLLEN1LELAT52            site    307 (-) AGAAA
S000245 POLLEN1LELAT52            site    328 (-) AGAAA
S000245 PYRIMIDINEBOXOSRAMY1A     site    373 (+) CCTTTT
S000259 RHERPATEXPA7              site    105 (-) KCACGW
S000512 RHERPATEXPA7              site    386 (-) KCACGW
S000512 ROOTMOTIFTAPOX1           site     52 (+) ATATT
S000098 ROOTMOTIFTAPOX1           site    320 (+) ATATT
S000098 SEBFCONSSTPR10A           site    132 (+) YTGTCWC
S000391 SEBFCONSSTPR10A           site    415 (+) YTGTCWC
S000391 SEBFCONSSTPR10A           site    145 (-) YTGTCWC
S000391 SORLIP1AT                 site    336 (+) GCCAC
S000482 SURECOREATSULTR11         site    145 (+) GAGAC
S000499 SURECOREATSULTR11         site    134 (-) GAGAC
S000499 SURECOREATSULTR11         site    417 (-) GAGAC
S000499 SV40COREENHAN             site    196 (+) GTGGWWHG
S000123 UPRMOTIFIIAT              site     89 (+) CCNNNNNNNNNNNNCCACG
S000426 WBOXNTERF3                site    160 (+) TGACY
S000547 WRKY71OS                  site    160 (+) TGAC
```

Results on search of PlantCARE that were additional to those found in PLACE:

Name of Motif:	Signal Sequence:
Unnamed__11	TCCACATAGA
Unnamed__3	CGTGG
Unnamed__4	CCACGTGG

CLOIII-78:

```
RESULTS OF YOUR SIGNAL SCAN SEARCH REQUEST
452 base pairs Signal Database File: user.dat
SITE #  Factor or Site Name          Loc.(Str.)   Signal Sequence
_________________________________________________________________________
S000414 ABRELATERD1               site    413 (+) ACGTG
S000414 ABRELATERD1               site    370 (-) ACGTG
S000414 ABRERATCAL                site    406 (-) MACGYGB
S000507 ACGTATERD1                site    194 (+) ACGT
S000415 ACGTATERD1                site    371 (+) ACGT
S000415 ACGTATERD1                site    413 (+) ACGT
S000415 ACGTATERD1                site    194 (-) ACGT
S000415 ACGTATERD1                site    371 (-) ACGT
S000415 ACGTATERD1                site    413 (-) ACGT
S000415 ACGTCBOX                  site    193 (+) GACGTC
S000131 ACGTCBOX                  site    193 (-) GACGTC
S000131 ARFAT                     site    131 (+) TGTCTC
S000270 ARFAT                     site    422 (+) TGTCTC
S000270 ARFAT                     site    143 (-) TGTCTC
```

S000270	ARR1AT	site	24	(+)	NGATT
S000454	ARR1AT	site	51	(-)	NGATT
S000454	ARR1AT	site	185	(-)	NGATT
S000454	ASF1MOTIFCAMV	site	411	(+)	TGACG
S000024	BOXCPSAS1	site	326	(+)	CTCCCAC
S000226	BS1EGCCR	site	322	(-)	AGCGGG
S000352	BS1EGCCR	site	342	(-)	AGCGGG
S000352	**CAATBOX1**	**site**	**50**	**(+)**	**CAAT**
S000028	CACTFTPPCA1	site	330	(+)	YACT
S000449	CACTFTPPCA1	site	69	(-)	YACT
S000449	CGACGOSAMY3	site	195	(-)	CGACG
S000205	CGACGOSAMY3	site	208	(-)	CGACG
S000205	CGCGBOXAT	site	254	(+)	VCGCGB
S000501	CGCGBOXAT	site	406	(+)	VCGCGB
S000501	CGCGBOXAT	site	254	(-)	VCGCGB
S000501	CGCGBOXAT	site	406	(-)	VCGCGB
S000501	CURECORECR	site	367	(+)	GTAC
S000493	CURECORECR	site	367	(-)	GTAC
S000493	DOFCOREZM	site	386	(-)	AAAG
S000265	E2FCONSENSUS	site	317	(+)	WTTSSCSS
S000476	E2FCONSENSUS	site	387	(+)	WTTSSCSS
S000476	EBOXBNNAPA	site	80	(+)	CANNTG
S000144	EBOXBNNAPA	site	80	(-)	CANNTG
S000144	GATABOX	site	403	(-)	GATA
S000039	**GT1CORE**	**site**	**261**	**(+)**	**GGTTAA**
S000125	GTGANTG10	site	410	(+)	GTGA
S000378	GTGANTG10	site	415	(+)	GTGA
S000378	GTGANTG10	site	187	(-)	GTGA
S000378	HEXMOTIFTAH3H4	site	411	(-)	ACGTCA
S000053	MYB1AT	site	374	(+)	WAACCA
S000408	MYB2CONSENSUSAT	site	121	(-)	YAACKG
S000409	MYBCORE	site	121	(+)	CNGTTR
S000176	MYBCORE	site	238	(-)	CNGTTR
S000176	MYBCOREATCYCB1	site	121	(-)	AACGG
S000502	MYBST1	site	403	(-)	GGATA
S000180	MYCCONSENSUSAT	site	80	(+)	CANNTG
S000407	MYCCONSENSUSAT	site	80	(-)	CANNTG
S000407	NODCON2GM	site	74	(+)	CTCTT
S000462	NODCON2GM	site	336	(+)	CTCTT
S000462	OSE2ROOTNODULE	site	74	(+)	CTCTT
S000468	OSE2ROOTNODULE	site	336	(+)	CTCTT
S000468	**PALBOXAPC**	**site**	**393**	**(+)**	**CCGTCC**
S000137	RAV1AAT	site	84	(-)	CAACA
S000314	RBCSCONSENSUS	site	22	(-)	AATCCAA
S000127	REBETALGLHCB21	site	403	(-)	CGGATA
S000363	RHERPATEXPA7	site	413	(-)	KCACGW
S000512	RHERPATEXPA7	site	197	(-)	KCACGW
S000512	RYREPEATBNNAPA	site	30	(+)	CATGCA
S000264	SEBFCONSSTPR10A	site	130	(+)	YTGTCWC
S000391	SEBFCONSSTPR10A	site	421	(+)	YTGTCWC
S000391	SEBFCONSSTPR10A	site	143	(-)	YTGTCWC
S000391	SURECOREATSULTR11	site	143	(+)	GAGAC
S000499	SURECOREATSULTR11	site	170	(+)	GAGAC
S000499	SURECOREATSULTR11	site	98	(-)	GAGAC
S000499	SURECOREATSULTR11	site	132	(-)	GAGAC
S000499	SURECOREATSULTR11	site	220	(-)	GAGAC
S000499	SURECOREATSULTR11	site	423	(-)	GAGAC

```
S000499 TATABOX5                  site    176 (-) TTATTT
S000203 TGACGTVMAMY               site    411 (+) TGACGT
S000377 WRKY71OS                  site    411 (+) TGAC
```

Results on search of PlantCARE that were additional to those found in PLACE:

Name of Motif:	Signal Sequence:
TCCACCT-motif	TCCACCT
TCCC-motif	TCTCCT
TCT-motif	AATCT