



**UnB**

Universidade de Brasília

Instituto de Ciências Biológicas

Departamento de Biologia Celular

RENATA HENRIQUE SANTANA

**ISOLAMENTO E CARACTERIZAÇÃO DE PROMOTORES  
ÓRGÃO-ESPECÍFICOS DE PLANTAS DE SOJA (*GLYCINE  
MAX*)**

RENATA HENRIQUE SANTANA

**ISOLAMENTO E CARACTERIZAÇÃO DE PROMOTORES  
ÓRGÃO-ESPECÍFICOS DE PLANTAS DE SOJA (*GLYCINE  
MAX*)**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação do Departamento de Biologia Celular do Instituto de Biologia da Universidade de Brasília, como requisito parcial à obtenção de título de Mestre em Biologia Molecular.

Orientadora: Dr.<sup>a</sup> Vera Tavares de Campos Carneiro

Co-Orientadora: Dr.<sup>a</sup> Leila Maria Gomes Barros

RENATA HENRIQUE SANTANA

**ISOLAMENTO E CARACTERIZAÇÃO DE PROMOTORES  
ÓRGÃO-ESPECÍFICOS DE PLANTAS DE SOJA (*GLYCINE  
MAX*)**

Dissertação submetida à comissão examinadora como requisito parcial à obtenção do grau de Mestre em Biologia Molecular.

Brasília, 16 de março de 2012

**BANCA EXAMINADORA**

Dr.<sup>a</sup> Vera Tavares de Campos Carneiro

Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF

Presidente – Orientadora

Dr.<sup>a</sup> Lucília Helena Marcellino

Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF

Membro efetivo

Dr. Ricardo Vilela Abdelnoor

Embrapa Soja, Londrina-PR

Membro efetivo

Dr. Júlio Carlyle M. Rodrigues

Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF

Suplente

*A minha família,  
principalmente aos meus pais,  
pelo exemplo, apoio, carinho e  
amor incondicionais. Sem eles  
nada disso seria possível.*

*Ao Igor, pela força,  
alegria, companhia,  
compreensão, amor e  
carinho.*

## **Agradecimentos**

Agradeço a Deus por me guiar e iluminar sempre meu caminho.

A todos da minha família por todo o apoio durante esse período.

À Dr.<sup>a</sup> Leila M. G. Barros, por todas as oportunidades oferecidas, pelos ensinamentos e por sempre acreditar em mim. Pela atenção, paciência, entusiasmo, amizade e carinho. Agradeço de coração por tudo!

À Dr.<sup>a</sup> Juliana D. de Almeida, por todos os ensinamentos, pela grande ajuda no meu desenvolvimento no ambiente científico e pela oportunidade. Obrigada por sempre acreditar no meu potencial e por todo o incentivo.

À Dr.<sup>a</sup> Vera T.C.Carneiro pela orientação, por acreditar em mim e no meu trabalho, pelos seus conhecimentos, pela contribuição e apoio. Principalmente agradeço pela paciência, tranquilidade, atenção e carinho.

À Michelle G. Cotta, colega de laboratório e grande amiga. Pelas conversas, pela amizade, pelo incentivo e apoio. Por ter sempre me ajudado, inicialmente me recomendando para o estágio no laboratório e depois por todo o auxílio.

À minha irmã, pela ajuda na revisão do texto e ao meu irmão pelos momentos de descontração e risadas.

Aos amigos do colégio que sempre estiveram ao meu lado, pela compreensão, apoio, momentos de alegria e pela torcida.

A todos do laboratório de Apomixia, por toda ajuda, contribuições e disponibilidade.

À D. Isabel e ao Adailton por toda a ajuda com o preparo do material do laboratório.

Aos estagiários Henrique, Marcos e Thayanne pelo auxílio.

À Dr.<sup>a</sup> Lucília H. Marcelino, e ao Dr. Júlio C. M. Rodrigues, por aceitarem participar da banca de defesa e pelas contribuições feitas ao trabalho.

Ao Dr. Ricardo V. Abdelnoor pelo apoio necessário para realização desse trabalho e pela participação da banca de defesa.

Ao Dr. Felipe R. da Silva por montar o banco de ESTs de soja da Embrapa Recursos Genéticos e Biotecnologia.

Ao departamento de Biologia Celular do IB da UnB pela oportunidade de participar do programa de pós-graduação.

À CAPES pelo suporte financeiro.

À Embrapa Recursos Genéticos e Biotecnologia pelo espaço para a execução do trabalho.

A todos que, de alguma forma, contribuíram para a realização desse trabalho!

*"É do buscar e não do achar que nasce o  
que eu não conhecia."*

Clarice Lispector

## Resumo

A planta de soja (*Glycine max* [L.] Merr) destaca-se mundialmente pela diversidade de produtos que proporciona para uso animal e humano. No Brasil, a soja é o produto que mais gera divisas cambiais atualmente. No entanto, a produtividade dessa cultura é afetada por diversos fatores bióticos e abióticos. No melhoramento da soja, a engenharia genética tem sido utilizada como ferramenta para geração de novos cultivares capazes de superar essas adversidades. Nessa técnica, o uso de promotores órgão-específicos em detrimento de promotores constitutivos para expressar transgenes restringe sua expressão temporal e espacialmente, aumentando a biossegurança e estabilidade do sistema. Neste trabalho utilizamos dados transcriptômicos e genômicos de domínio público para identificar genes órgão-específicos de soja, isolar e caracterizar suas regiões promotoras. Os genes *GmSulfT1* e *GmCit1* foram identificados por meio de análises *in silico* em três bancos de ESTs de soja: GenoSoja, Unigene e da Embrapa Recursos Genéticos e Biotecnologia. Esses genes foram selecionados para validação biológica por apresentarem o perfil de EST órgão-específico em pelo menos dois dos bancos citados e pelo ineditismo. Seus perfis de expressão foram avaliados por RT-PCR semiquantitativa e *Northern blot*. O *GmSulfT1* apresentou expressão preferencial em raiz e semente e o *GmCit1*, por sua vez, em folha de plantas de soja. As regiões promotoras desses genes foram identificadas no genoma da soja cv. Williams 82 e analisadas *in silico* para identificação de motivos putativos de elementos *cis* regulatórios. Iniciadores desenhados baseados nessas sequências foram capazes de amplificar fragmentos dessas regiões do genoma da soja cv. Conquista por PCR. Os promotores derivados de deleções da região promotora de *GmSulfT1*, PSulfT0,5 e PCit0,4, PCit0,8 e PCit1,9, de *GmCit1*, foram clonados a montante do gene repórter *gus* em vetor binário utilizando o sistema Gateway<sup>®</sup>. Plantas de fumo (*Nicotiana tabacum*) foram transformadas para caracterização *in vivo* dos promotores. No ensaio histoquímico, o gene *gus* sob regulação dos promotores PSulfT0,5, PCit0,8 e PCit1,9 apresentou altos níveis de expressão em raízes e folhas, enquanto o PCit0,4 preferencialmente em folha. PCit0,4 isolado no presente trabalho poderá ser utilizado futuramente para conferir expressão preferencial em folhas de transgênicos. O caráter de alta expressão de *gus* em mais de um órgão atribuído por PSulfT0,5, PCit0,8 e PCit1,9, pode, da mesma forma, ser explorado para expressão de transgenes em plantas.

Palavras-chave: promotores de soja, promotores órgão-específicos, caracterização de promotores, regulação gênica, elementos *cis*.

## Abstract

The soybean plant (*Glycine max* [L.] Merr) is a worldwide important crop due to the diversity of products that it provides for animals and humans. In Brazil, the soybean is currently the commodity that generates most of its foreign-exchange reserves. However, the productivity of this crop is reduced by several biotic and abiotic factors. In soybean breeding, genetic engineering has been successfully applied to generating new cultivars able to overcome these adversities. In this technique, the use of organ-specific instead of constitutive promoters restricts the transgene expression temporally and spatially, increasing biosafety and the stability of the system. In this work we used transcriptomic and genomic data of public domain to identify soybean organ-specific genes, isolate and characterize their promoter region. Genes *GmSulfT1* and *GmCit1*, were *in silico* identified in three ESTs databases, GenoSoja, Unigene and from Embrapa Genetic Resources and Biotechnology. These genes were selected for further analysis due to their organ specific EST profile in at least two of the mentioned database, and novelty criteria. Their expression profiles were biologically validated by semiquantitative RT-PCR and *Northern blot* assays. *GmSulfT1* was preferentially expressed in roots and seeds and *GmCit1* in leaves of soybean plants. The promoter regions of these two genes were identified in the soybean cv. Williams 82 genome sequence and *in silico* analyzed to detect putative motifs of *cis* regulatory elements. Primers designed based in those sequences were able to amplify fragments of the promoter regions of *GmSulfT1* and of *GmCit1* by PCR with soybean cv. Conquista genomic DNA. The deleted promoters derived from *GmSulfT1* promoter region, PSulfT0.5, and from *GmCit1*, PCit0.4, PCit0.8 and PCit1.9, were fused with the reporter gene *gus* in a binary vector using the Gateway<sup>®</sup> System. Tobacco plants were transformed in order to *in vivo* characterizing these promoters. Histochemical assay showed high *gus* expression in roots and leaves under control of PSulfT0.5, PCit0.8 and PCit1.9, and under control of PCit0.4, more in leaves. PCit0.4, isolated in the present work, can be used hereafter to activate transgene expression preferentially in leaves. The high expression levels in more than one organ obtained using the other promoters isolated in this work can also be explored to express transgenes in plants.

Keywords: soybean promoters, organ specific promoters, promoter characterization, gene regulation, *cis* elements.



## Lista de Ilustrações

<b>Figura 1.</b> Regulação gênica em eucariotos.-----	20
<b>Figura 2.</b> Elementos regulatórios de um gene eucariótico da classe II. -----	22
<b>Figura 3.</b> Possibilidade combinatória de proteínas envolvidas no controle da expressão gênica. -----	23
<b>Figura 4.</b> Representação esquemática de um estimulossoma ( <i>enhanceosome</i> ).-----	23
<b>Figura 5.</b> Motivos do promotor mínimo ( <i>core promoter</i> ). -----	26
<b>Figura 6.</b> Ação dos isoladores em genomas eucarióticos. -----	27
<b>Figura 7.</b> Representação esquemática do vetor pENTR <sup>TM</sup> . -----	43
<b>Figura 8.</b> Representação esquemática do vetor pMDC162.-----	43
<b>Figura 9.</b> Perfil de expressão do <i>contig</i> 8703 baseado em ESTs. -----	54
<b>Figura 10.</b> Análise comparativa da sequência do <i>contig</i> 8703 com o genoma da soja. -----	55
<b>Figura 11.</b> Anotação funcional do <i>locus</i> Glyma13g26070 correspondente ao <i>contig</i> 8703 no Phytozome.-----	55
<b>Figura 12.</b> Perfil de expressão do <i>contig</i> 18151 baseado em ESTs. -----	57
<b>Figura 13.</b> Análise comparativa da sequência do <i>contig</i> 18151 com o genoma da soja.-----	59
<b>Figura 14.</b> Anotação funcional dos <i>loci</i> que alinham com o <i>contig</i> 18151 no Phytozome. -----	60
<b>Figura 15.</b> Perfil de expressão do gene <i>GmSulft1</i> .-----	61
<b>Figura 16.</b> Perfil de expressão do gene <i>GmCit1</i> .-----	61
<b>Figura 17.</b> Sequência genômica da região promotora e codificadora de <i>GmSulft1</i> de <i>G. max.</i> cv. Williams 82. -----	63
<b>Figura 18.</b> Sequência genômica da região promotora e codificadora de <i>GmCit1</i> de <i>G. max.</i> cv. Williams 82. -----	64
<b>Figura 19.</b> Fragmentos amplificados da região promotora dos genes <i>GmSulft1</i> e <i>GmCit1</i> . -	65
<b>Figura 20.</b> Fragmentos de restrição gerados do vetor de entrada pENTR <sup>TM</sup> (VE) e dos vetores binários pMDC162 (VB)-----	66
<b>Figura 21.</b> Hibridização de colônias de <i>Agrobacterium tumefaciens</i> transformadas com os vetores binários. -----	67
<b>Figura 22.</b> Análise <i>in silico</i> e funcional da região promotora do <i>GmSulft1</i> . -----	70
<b>Figura 23.</b> Análise <i>in silico</i> e funcional da região promotora do <i>GmCit1</i> .-----	72

## Lista de Tabelas

<b>Tabela 1.</b> Sequências dos iniciadores específicos utilizados nas reações de RT-PCR semiquantitativa. -----	40
<b>Tabela 2.</b> Sequências de iniciadores utilizados na clonagem dos promotores e deleções 5' dos genes <i>GmSulfT1</i> e <i>GmCit1</i> .-----	44
<b>Tabela 3.</b> Níveis de atividade da GUS encontrada nas plantas de <i>Nicotiana tabacum</i> transformadas. -----	68

## Lista de abreviaturas

° C	graus Celsius
μF	microfarad
μg	micrograma
μm	micrometro
μL	microlitro
	Ohm
BAP	6-benzilaminopurina
BLAST	<i>basic local alignment search tool</i>
cDNA	ácido desoxirribonucléico complementar
cm	centímetro
CTAB	brometo de cetiltrimetilamônio
DEPC	dietilpirocarbonato
DNA	ácido desoxirribonucléico
dNTP	desoxirribonucleotídeo trifosfatado
EDTA	ácido etileno diaminotetracético
EST	Expressed Sequence Tag
g	grama
GUS	-glucuronidase
kb	quilobase
kV	quilovolt
L	litro
LB	Luria-Bertani

mg	miligrama
mL	mililitros
mm	milímetro
mM	milimolar
MOPS	Ácido 3-(N-morfolino) propanosulfônico
mRNA	RNA mensageiro
MS	Murashige e Skoog
pb	pares de base
PCR	reação de polimerase em cadeia
RNA	ácido ribonucléico
rpm	rotações por minuto
RT	transcriptase reversa
SDS	dodecil sulfato de sódio
TM	melting temperature
TRIS	tri (hidroximetil) aminometano
TSS	sítio de início da transcrição ( <i>transcription start site</i> )
U	unidade
UTR	região não traduzida ( <i>untranslated region</i> )
X-Gluc	5- bromo- 4- cloro- 3- indolil- - D- glucoronídeo

# SUMÁRIO

1.	<i>Introdução</i>	15
1.1.	A soja	16
1.1.1.	Taxonomia	16
1.1.2.	Cultura de soja	17
1.2.	Engenharia genética de plantas	18
1.3.	Regulação da transcrição em eucariotos	19
1.3.1.	Elementos <i>trans</i> -regulatórios	22
1.3.2.	Elementos <i>cis</i> -regulatórios	24
1.3.2.1.	Caracterização de elementos cis para expressão gênica	28
1.4.	Promotores constitutivos	28
1.5.	Promotores órgão-específicos	30
1.5.1.	Promotores órgão-específicos de soja	31
2.	<i>Objetivos</i>	33
2.1.	Objetivo Geral	34
2.2.	Objetivos específicos	34
3.	<i>Materiais e Métodos</i>	35
3.1.	Identificação de <i>contigs</i> órgão-específicos	36
3.2.	Análises virtuais dos <i>contigs</i> órgão-específicos	36
3.2.1.	Anotação funcional e mapeamento	37
3.3.	Desenho dos iniciadores para RT-PCR	37
3.4.	Material Vegetal	37
3.5.	Extração de DNA	38
3.6.	Extração de RNA	39
3.7.	Reação de RT-PCR semiquantitativa	39
3.8.	Northern blot	40
3.9.	Isolamento dos promotores	41
3.10.	Análise <i>in silico</i> das regiões promotoras	42

3.11.	Clonagem dos promotores -----	42
3.11.1.	Extração de DNA plasmidial -----	45
3.11.2.	Recombinação LR da tecnologia Gateway® -----	46
3.12.	Transformação de <i>Agrobacterium tumefaciens</i> -----	47
3.13.	Hibridização de colônias -----	47
3.14.	Cultura de tecidos-----	48
3.14.1.	Transformação de plantas de <i>Nicotiana tabacum</i> -----	48
3.14.1.1.	Co-Cultura -----	48
3.14.1.2.	Seleção das plantas transformadas -----	49
3.14.1.3.	Aclimação das plantas transformadas-----	50
3.15.	Ensaio Histoquímico-----	50
4.	<i>Resultados</i> -----	52
4.1.	Análises <i>in silico</i> -----	53
4.2.	Análise da expressão dos genes <i>GmSulFT1</i> e <i>GmCit1</i> -----	60
4.3.	Isolamento dos promotores dos genes <i>GmSulFT1</i> e <i>GmCit1</i> -----	62
4.4.	Análises <i>in silico</i> e funcional dos promotores isolados de <i>GmSulFT1</i> -----	67
4.5.	Análises <i>in silico</i> e funcional dos promotores isolados de <i>GmCit1</i> -----	71
5.	<i>Discussão</i> -----	73
6.	<i>Conclusão</i> -----	82
7.	<i>Perspectivas</i> -----	84
8.	<i>Referências Bibliográficas</i> -----	86

## **1. Introdução**

## 1.1. A soja

### 1.1.1. Taxonomia

Reino: Plantae

Subreino: Tracheophyta

Superdivisão: Spermatophyta

Divisão: Magnoliophyta

Classe: Magnoliopsida

Subclasse: Rosidae

Ordem: Fabales

Família: Fabaceae

Subfamília: Papilionoideae

Gênero: *Glycine* Willd.

Espécie: *Glycine Max* [L.] Merr

O genoma da soja (*Glycine max* [L.] Merr) ( $2n = 40$ ) foi originado a partir de vários ciclos de duplicação (paleopoliploidia) ocorridos nos últimos 50 milhões de anos (Shoemaker *et al.*, 2006). A planta de soja destaca-se pela diversidade de produtos que proporciona para uso humano e animal. A semente é constituída por óleo (18%) e proteína (38%), sendo que o primeiro é o seu principal produto, consumido como óleo comestível ou como insumo industrial (Hartman *et al.*, 2011). A proteína, cuja qualidade se iguala àquelas de fonte animal, é purificada para uso na alimentação humana e animal. As sementes de soja podem ser utilizadas também para produção de biocombustível e são fontes de nutracêuticos, produtos alimentícios que proporcionam benefícios à saúde, como lecitinas e bioflavonóides (Hartman *et al.*, 2011).



### 1.1.2. Cultura de soja

No Brasil, a soja possui grande importância econômica, pois a exportação do complexo da planta, constituído por grão, farelo e óleo, tem o maior peso na balança comercial, tornando-se o produto que mais gera divisas cambiais atualmente (Ministério da Agricultura Pecuária e Abastecimento, 2011). No cenário mundial, o país é o segundo maior exportador dessa *commodity*, segundo os dados econômicos da safra 2010/ 2011 (Embrapa Soja, 2011).

Estima-se que a produção e consumo da soja aumentem à medida que a população mundial cresce, devido à sua importância tanto na alimentação humana e animal quanto industrial e aplicações farmacêuticas (Hartman *et al.*, 2011). No entanto, para o aumento ser eficiente e sustentável, é necessário contornar diversos fatores que afetam a produção negativamente. Seca, alagamento, congelamento, disponibilidade de nutrientes no solo, salinidade e fotoperíodo são alguns dos fatores abióticos que afetam o cultivo de soja. Entre os fatores bióticos estão pragas como insetos e microrganismos que causam doenças como a ferrugem asiática da soja (*Phakopsora pachyrhizi*) e infecção da raiz por nematóides (*Heterodera glycines*) (Hartman *et al.*, 2011).

As estratégias que podem ser utilizadas para contornar as perdas por doenças, pragas e estresses abióticos são o uso de pesticidas, fertilizantes, irrigação ou o desenvolvimento de variedades de plantas resistentes (Hartman *et al.*, 2011). Porém, os insumos agrícolas podem ameaçar a saúde humana e o meio ambiente ao contaminar lençóis de água, o solo e acumular no produto final de consumo, ou seja, nos grãos (Matson *et al.*, 1997). Além disso, a disponibilidade de água e os altos custos podem limitar o sistema de irrigação da cultura, tornando-o em alguns casos inviável (Hartman *et al.*, 2011).

A simbiose entre raízes de leguminosas com microrganismos fixadores de nitrogênio tem sido foco de muitas pesquisas visando o desenvolvimento de novos cultivares onde esse sistema seja mais eficiente e, assim, auxilie ainda mais no enriquecimento do solo e, conseqüentemente, reduza a aplicação de fertilizantes (Keyser, 1992; Salvagiotti *et al.*, 2008).

A engenharia genética é uma ferramenta importante para a produção de novos cultivares de soja que possam superar as adversidades da cultura. Adicionalmente, essa técnica contribui com o aumento da produtividade auxiliando o melhoramento, pois amplia a

base genética de culturas ao transpor barreiras genéticas pela introdução de características encontradas em organismos filogeneticamente distantes (Singh e Hymowitz, 1999).

A transgenia permite a inserção de características que podem beneficiar as plantas e seus produtos (Singh *et al.*, 2006) ou o desenvolvimento de plantas sem características indesejáveis. Em diversos países essa tecnologia já está sendo utilizada para aumentar a produção agrônômica, sendo o Brasil o segundo país com maior área plantada de culturas geneticamente modificadas: 21,4 milhões de hectares, onde 16,2 milhões de hectares são de soja Roundup Ready (RR), atualmente o único plantado comercialmente no país (James, 2009). Dos cultivares de soja aprovados para plantio no Brasil, o Roundup Ready, Cultivance, Liberty Link<sup>TM</sup> possuem tolerância a herbicidas, enquanto o cultivar Intacta RR2 PRO, além de ser tolerante a herbicida, produz a proteína inseticida Cry1A (Ministério da Agricultura Pecuária e Abastecimento, 2011).

O desenvolvimento da pesquisa na área genômica e o recente sequenciamento do genoma da soja (Schmutz *et al.*, 2010) tornarão mais rápida e direcionada a criação de novas variedades, uma vez que o genoma provê informações sobre a expressão gênica, estrutura, desenvolvimento e evolução dos organismos. Desta forma, é possível por meio da engenharia genética vislumbrar o aumento da produção da soja sem expansão da área plantada, manipulando-se rotas metabólicas para aumentar a eficiência fotossintética, a concentração de carbono nos seus tecidos de reserva, a eficiência da fixação do nitrogênio e influenciar na fase reprodutiva da espécie (Ainsworth *et al.*, 2012).

## **1.2. Engenharia genética de plantas**

A transgenia consiste na inserção de um ou mais genes capazes de conferir ao organismo uma característica desejável. Denomina-se transgene a sequência de nucleotídeos contendo uma região promotora, uma região codificadora e uma região terminadora inserida em um genoma hospedeiro (Visarada *et al.*, 2009). Os transgenes podem ser oriundos de organismos semelhantes ou filogeneticamente distantes do hospedeiro (Singh *et al.*, 2006). Os dois métodos mais utilizados para inserir genes em plantas são a biobalística, no qual a planta é bombardeada por partículas de ouro ou de tungstênio cobertas pelo DNA; e via *Agrobacterium sp.*, uma bactéria de solo que é capaz de transferir um segmento de seu DNA para plantas por meio do plasmídeo *Ti* (*tumor inducing*) (Singh *et al.*, 2006).

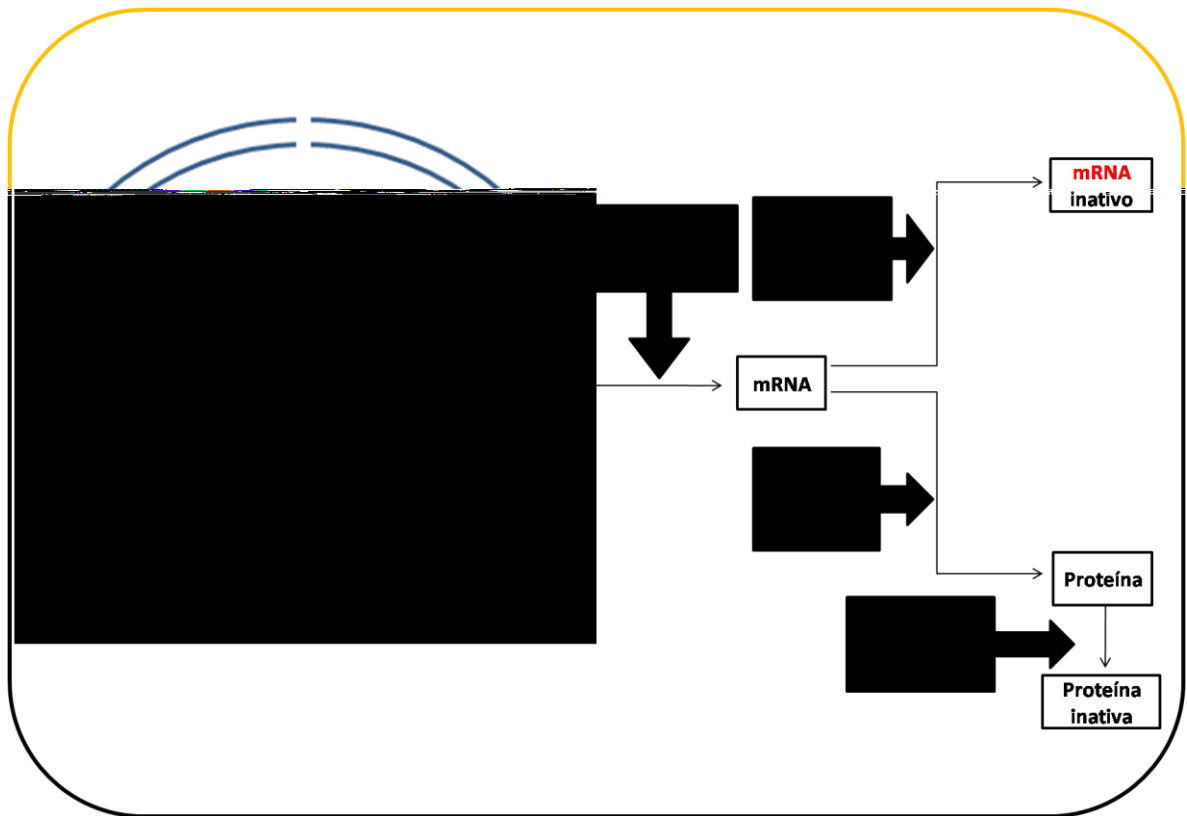
A regulação da expressão do transgene será feita, em maior parte, pelo promotor, uma vez que a transcrição é o primeiro processo de regulação gênica. A expressão do transgene, porém, não é uniforme em todas as plantas geradas sob as mesmas condições, pois ele está sujeito a outros mecanismos de regulação endógenos da planta. A escolha de um promotor adequado para regular o transgene pode diminuir essa variabilidade de expressão e aumentar a eficiência da técnica (Cammue *et al.*, 2005).

Apesar de proporcionar diversos benefícios para a agricultura ao aumentar a produção de culturas, diminuir o uso de agrotóxicos e os custos, a cultura de organismos geneticamente modificados ainda levanta questões quanto à segurança ecológica e toxicológica (Singh *et al.*, 2006). Uma medida que pode ser utilizada para diminuir preocupações com a biossegurança de plantas GM é o uso de promotores, sequências regulatórias a montante (*upstream*) da região codificadora, responsáveis pelo controle preciso dos transgenes, ou seja, promotores que limitam a expressão desses a determinado órgão ou período (Potenza *et al.*, 2004).

Atualmente, existem diversos promotores isolados utilizados para regular transgenes em plantas transformadas: promotores constitutivos, órgão/tecido/célula-específicos, induzíveis e promotores sintéticos. A escolha do promotor a ser utilizado depende do objetivo final da transformação, seja ela para estudo de expressão gênica e desenvolvimento da planta, ou uso comercial (Potenza *et al.*, 2004).

### **1.3. Regulação da transcrição em eucariotos**

As células respondem a sinais externos por meio da expressão gênica, ou seja, quais genes serão ativados e quais serão desativados para gerar um conjunto de proteínas em resposta a determinado estímulo. A regulação da expressão gênica é o mecanismo utilizado pelas células para modular essa resposta. Em eucariotos essa regulação pode ocorrer nas diversas etapas do processo de decodificação da informação contida na sequência nucleotídica de DNA para gerar uma proteína (Alberts *et al.*, 2002) (Figura 1).



**Figura 1.** Regulação gênica em eucariotos. No esquema estão indicadas as etapas do processo que levam à produção de proteínas que são sujeitos a mecanismos de regulação em uma célula eucariótica. Fonte: adaptado de Alberts et al (2002).

A enzima responsável pela transcrição de genes codificadores de proteína em eucariotos é a RNA polimerase II. Nessas células existe ainda a RNA polimerase I, localizada no nucléolo, que sintetiza o precursor do rRNA, e a RNA polimerase III que atua fora do nucléolo e transcreve genes que codificam tRNAs, rRNA 5S e outras moléculas de pequenos RNAs (Lodish *et al.*, 2000). Em plantas, existem ainda as RNA polimerases IV e V que participam da síntese de pequenos RNAs envolvidos no silenciamento gênico (Haag e Pikaard, 2011).

A transcrição pela RNA polimerase II requer fatores gerais de transcrição associados ao promotor e a estrutura do DNA nuclear eucariótico empacotado em cromatina configura outro mecanismo de regulação gênica (Kornberg, 2007).

Existem dois modos de iniciação da transcrição usados para classificar o tipo de promotor: de iniciação focada, quando a transcrição inicia-se sempre em um único nucleotídeo ou em um grupo deles e de iniciação dispersa, o qual possui vários sítios de

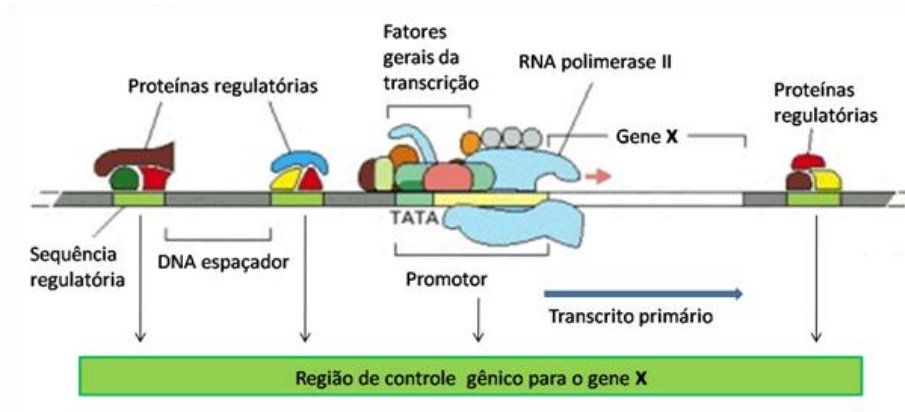
iniciação fracos dentro de uma região de 50 a 100 nucleotídeos. Genes regulados espacialmente e temporalmente tendem a apresentar promotores com iniciação focada e genes constitutivos tendem a apresentar iniciação dispersa (Juven-Gershon e Kadonaga, 2010). Promotores podem apresentar ainda os dois modos de iniciação, ou seja, contêm múltiplos sítios de iniciação fracos e um sítio forte (Juven-Gershon e Kadonaga, 2010).

Nos eucariotos a maioria dos genes é regulada individualmente contendo a montante uma região promotora e a jusante um sítio de terminação da transcrição e os transcritos são monocistrônicos. No entanto, alguns genes são transcritos de forma policistrônicas, como os genes do complexo de histocompatibilidade (MHC) em animais. Em *Arabidopsis thaliana* foram identificados dois grupos de genes similares a operons que são necessários para a síntese de dois tipos diferentes de triterpenos (Field *et al.*, 2011). Os mRNAs policistrônicos podem ser separados em duas classes: dicistrônicos ou policistrônicos. Nos mRNAs dicistrônicos, que é transportado para o citoplasma, o mRNA contém sítios internos de entrada dos ribossomos (IRES) ou alguma maneira de reiniciação da tradução a partir do códon de parada. Já os pré-mRNA policistrônicos são processados ainda no núcleo por *splice* (retirada de íntrons da molécula) alternativo, os quais resultam em mRNAs monocistrônicos que são transportados para o citoplasma e lá traduzidos (Blumenthal, 2004).

Em relação à disposição dos genes nos cromossomos dos organismos eucariotos são descritas três formas de arranjo. Existem genes localizados na mesma fita que podem compartilhar ou não as regiões 3' (terminador) de um com a 5' (promotor) do outro; genes localizados em fitas diferentes os quais podem ser independentes ou compartilhar as porções 5' ou as porções 3' e genes sobrepostos codificados na mesma fita ou na fita complementar (Takai *et al.*, 2004; Wei *et al.*, 2011).

Para que a transcrição ocorra em células eucarióticas é necessário que uma proteína de ativação gênica se ligue ao DNA e recrute um complexo responsável pelo remodelamento da cromatina, tornando a região acessível às enzimas de modificação de histonas. A modificação das histonas permitirá a ligação de outras proteínas ativadoras. Essas proteínas adicionais recrutarão os fatores gerais de transcrição (TFIIB, TFIIF, TFIIE, TFIIH, TFIIA, TFIID) e a RNA polimerase para o promotor, formando o chamado complexo de pré-iniciação, que será rearranjado após a ligação de outras proteínas de ativação e dará início à transcrição (Kornberg, 2007). A disposição desses elementos na região de regulação de um gene eucariótico codificador de proteína está ilustrada na Figura 2. Nesse contexto, na regulação

gênica há participação de dois tipos de elementos regulatórios: elementos *trans* (proteínas ativadoras e fatores gerais de transcrição e elementos *cis* (sequências reconhecidas pelas proteínas regulatórias).



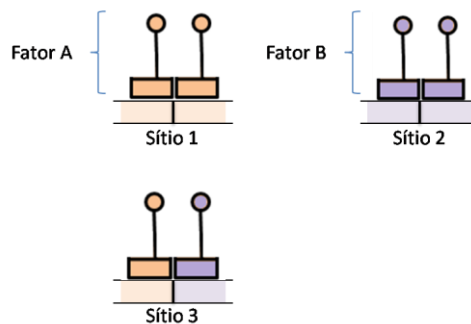
**Figura 2.** Elementos regulatórios de um gene eucariótico da classe II. Os elementos *cis* são sítios de ligação no DNA para as proteínas regulatórias (elementos *trans*). Fonte: Alberts et al. (2002).

### 1.3.1. Elementos *trans*-regulatórios

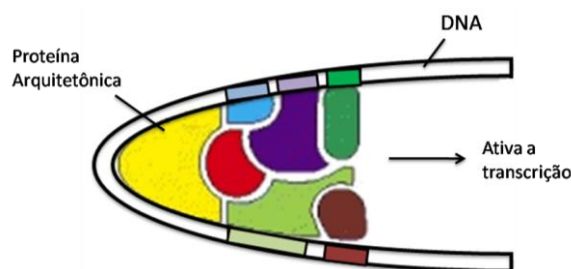
Os elementos *trans*-regulatórios ou fatores de transcrição são proteínas que possuem domínios capazes de reconhecer sequências específicas de DNA e se ligarem a elas. Tais elementos influenciam diferentes passos do início da transcrição, desde o remodelamento da cromatina até a ligação do complexo de pré-iniciação ao promotor (Alberts *et al.*, 2002).

As proteínas regulatórias ativadoras e repressoras da transcrição (fatores de transcrição) possuem, geralmente, estrutura modular com domínios distintos. No modelo mais simples, um domínio é responsável pelo reconhecimento da sequência do DNA e outro, o domínio de ativação ou repressão, acelera ou diminui a taxa de início da transcrição (Lodish *et al.*, 2000). Os fatores de transcrição são classificados de acordo com o domínio de ligação ao DNA (Lodish *et al.*, 2000), fatores que possuem estruturas de reconhecimento de DNA semelhantes pertencem à mesma família. Existem ainda as proteínas regulatórias co-ativadoras ou co-repressoras, que não possuem o domínio de reconhecimento de DNA, mas se ligam a proteínas que o reconhecem (Alberts *et al.*, 2002)

As proteínas responsáveis pela regulação podem formar fatores heterodiméricos de regulação da transcrição (Figura 3) (Lodish *et al.*, 2000). Essa regulação combinatória aumenta o repertório de controle da transcrição, assim, não é necessário que a célula sintetize um fator de transcrição específico para cada gene, o que resultaria em um gasto energético muito grande. Outro mecanismo de controle combinatório é o estimulossoma (*enhanceosome*) (Figura 4), no qual proteínas arquitetonas ligam-se ao DNA dobrando-o e promovendo a associação entre os fatores de regulação gênica que podem estar localizados a milhares de pares de bases distantes do sítio de início da transcrição (TSS) (Alberts *et al.*, 2002).



**Figura 3.** Possibilidade combinatória de proteínas envolvidas no controle da expressão gênica. No esquema, dois fatores de transcrição (A e B) podem regular a transcrição individualmente ou podem combinar-se entre si para formar uma terceira unidade de regulação. Fonte: Adaptado de Lodish et al (2000).



**Figura 4.** Representação esquemática de um estimulossoma (*enhanceosome*). A proteína arquitetônica, em amarelo, dobra a fita de DNA para permitir a interação entre as proteínas regulatórias situadas distantes umas das outras. Fonte: Alberts et al. (2002).

As proteínas de regulação gênica atuam sinergicamente, dessa forma, a taxa de transcrição de um gene dependerá de quais fatores de transcrição estão ligados às sequências regulatórias (Alberts *et al.*, 2002). Uma vez que proteínas diferentes são expressas em locais diferentes e em estágios de desenvolvimento diferentes (Lodish *et al.*, 2000), a expressão de determinado gene dependerá dos fatores de transcrição presentes na célula.

### 1.3.2. Elementos *cis*-regulatórios

De acordo com a localização, existem três classes de elementos *cis*-regulatórios em eucariotos nos quais as proteínas *trans*-regulatórias se ligam (Maston *et al.*, 2006).

Promotor mínimo (*core promoter*): região de ligação da RNA polimerase II e fatores gerais de transcrição próximos ao TSS;

Elementos *cis* proximais: sequências próximas ao promotor que auxiliam a ligação da RNA polimerase II e modulam sua atividade;

Elementos independentes de distância (estimuladores, silenciadores e isoladores): podem atuar a distâncias grandes do TSS e influenciam a taxa de transcrição.

O promotor mínimo contém elementos regulatórios essenciais para que a transcrição ocorra e possuem sequências consenso que direcionam a RNA polimerase II para que a transcrição seja iniciada de maneira correta (Juven-Gershon e Kadonaga, 2010). Existem diversos motivos não universais presentes no promotor mínimo que se combinam entre si para formar diversos tipos de promotores mínimos (Juven-Gershon e Kadonaga, 2010).

O TATA Box é um elemento simétrico cuja sequência consenso está localizada geralmente de 25 a 30 pb a montante do TSS, em alguns organismos, no entanto, esse elemento pode estar de 40 a 120 pb a montante do TSS (Smale e Kadonaga, 2003). A prevalência do elemento TATA, ou seja, da sequência consenso TATAAA ou de sequências consenso contendo apenas um nucleotídeo diferente, nos promotores de genes que codificam proteínas varia entre os organismos. Por exemplo, em drosófila de 205 promotores mínimos analisados o TATA Box estava presente em 43% deles (Smale e Kadonaga, 2003); em mamíferos 10-15% dos promotores o possuem (Juven-Gershon e Kadonaga, 2010); em plantas a prevalência do TATA Box é de 20-30% (Yamamoto *et al.*, 2011). Esse elemento é reconhecido pelo complexo de multiproteínas TFIID, que contém a proteína TATA *binding protein* (TBP) e os fatores associados a ela (TAFs – *TBP associated factors*) (Smale e Kadonaga, 2003).



A TBP liga-se ao TATA Box e essa ligação, segundo Irani et al (2010), se dá fortemente a uma fita do DNA auxiliando a separação das duas fitas para promover a transcrição. A ligação da TBP ao TATA box provoca uma dobra no DNA (Lodish *et al.*, 2000), permitindo a interação com o TFIIB que, por sua vez, posiciona a polimerase no promotor (Kornberg, 1999).

Muitos genes codificadores de proteínas, no entanto, não possuem TATA Box canônico (Smale, 1997). Os promotores sem TATA (*TATA-less promoters*) não possuem especificidade intrínseca pelo TFIID, fator necessário para o início da transcrição por meio da RNA polimerase II, tanto em promotores com TATA box quanto naqueles sem TATA (Pugh e Tjian, 1991).

Segundo Pugh e Tjian (1991), o recrutamento do TFIID ao promotor sem TATA é mediado por uma proteína ativadora (Sp1 ou similar a Sp1) que se liga ao DNA e é reconhecida por um fator mediador do TFIID, resultando na ligação do TFIID ao sítio onde o complexo de pré-iniciação é formado (Pugh e Tjian, 1991). O reconhecimento do promotor na ausência de TATA box pelo TFIID pode também ocorrer via TAFs (fatores associados à TBP) presentes no complexo (Juven-Gershon *et al.*, 2008) e também por meio da TBP. A TBP pode interagir também com sequências ricas em A/T que podem funcionar como TATA Box (Smale e Kadonaga, 2003) ou com sítios de ligação de TBP não canônicos, porém funcionais (Zenzie-Gregory *et al.*, 1993).

Inicialmente promotores sem TATA eram relacionados a genes constitutivos, em plantas, no entanto, foi constatado que essa classe de promotor está relacionada a genes da fotossíntese (Nakamura *et al.*, 2002). Alguns promotores sem TATA possuem um elemento iniciador (Inr) que especifica e se sobrepõe ao sítio de início da transcrição (Zenzie-Gregory *et al.*, 1993).

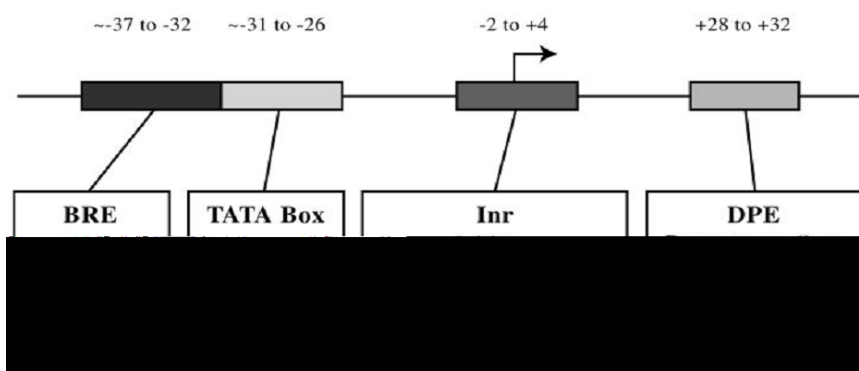
O iniciador é outro motivo *cis*-regulatório presente no promotor mínimo de alguns genes. Esse elemento é caracterizado pela presença de uma adenosina no sítio de início da transcrição (+1), uma citosina na posição -1 e algumas pirimidinas ao redor desses nucleotídeos (Smale e Kadonaga, 2003). Funcionalmente o Inr é semelhante ao TATA Box, pois é igualmente reconhecido pelo TFIID e pode compensar a ausência desse elemento em promotores sem TATA. Porém, quando esses dois elementos estão presentes no mesmo promotor separados por 25 pb, eles funcionam sinergisticamente (Smale e Kadonaga, 2003).

Este fato pode ser explicado pela ligação cooperativa do TFIID ao TATA Box e ao Inr (Smale e Kadonaga, 2003). A proteína TBP sozinha é incapaz de interagir com o Inr, mas na presença dos outros elementos do complexo TFIID, a atividade foi detectada (Smale e Kadonaga, 2003). Além do TFIID, a RNA polimerase II também reconhece o Inr e a partir dessa evidência supõe-se que o TFIID reconhece o Inr durante a formação do complexo de iniciação e a preferência da polimerase a esse elemento contribui para seu posicionamento adequado no complexo de iniciação (Smale e Kadonaga, 2003).

O DPE (*downstream core promoter element*) está localizado na posição +28 a +33 em relação ao TSS (+1) em promotores que contém o Inr. Esses dois elementos agem em conjunto, uma vez que o fator TFIID se liga cooperativamente aos dois (Juven-Gershon e Kadonaga, 2010). Ao contrário do TATA box, o DPE não possui atividade sem o Inr, pois funcionam como unidade regulatória única (Smale e Kadonaga, 2003). O espaçamento Inr-DPE é essencial para a ligação do TFIID a esses elementos, já que essa capacidade é perdida com a adição ou deleção de apenas uma base nesse intervalo (Smale e Kadonaga, 2003).

O BRE (*TFIIB recognition element*) é o único elemento caracterizado do promotor mínimo que é reconhecido por outro fator além do TFIID. Ele pode estar localizado logo a montante (BREu) ou logo a jusante (BREd) do TATA-box (Buchel *et al.* 1999 ) e pode aumentar ou diminuir as taxas basais de transcrição (Juven-Gershon e Kadonaga, 2010). Plantas e leveduras não possuem esse elemento regulatório (Smale e Kadonaga, 2003).

A localização e as sequências consenso dos diferentes elementos do promotor mínimo estão ilustradas na Figura 5.



**Figura 5.** Motivos do promotor mínimo (core promoter). O esquema mostra a localização dos motivos no DNA e os quadros mostram as sequências consensos. Fonte: Smale e Kadonaga (2003).



### **1.3.2.1. Caracterização de elementos cis para expressão gênica**

A contribuição dos elementos *cis* para a expressão global do gene pode ser caracterizada por diferentes técnicas. A geração de deleções na extremidade 5' da região promotora fornece informações sobre a contribuição dos elementos da sequência para a regulação da transcrição gênica. O mesmo ocorre quando há deleção de um único elemento *cis* ou mutação sítio dirigida, inserção ou deleção de nucleotídeos na região promotora (Gowik *et al.*, 2004; Kim *et al.*, 2006).

Tais análises permitem correlacionar sequências de DNA específicas e seu papel na mudança dos níveis de transcrição produzida. A abordagem desconstrutiva de deleções na extremidade 5' foi utilizada para caracterizar os domínios do promotor do oncogene *rolA* de *Agrobacterium rhizogenes* que são responsáveis pelas alterações morfológicas relacionadas com a síndrome de raízes em cabeleira na planta infectada (Carneiro e Vilaine, 1993). Neste trabalho, a remoção dos domínios A, B ou C do promotor *rolA* provocou o aparecimento de novos padrões de expressão dos transcritos *rolA* em folha, caule e raiz e, conseqüentemente, alteração dos fenótipos da síndrome nesses órgãos.

Outra abordagem utilizada para caracterização de diversas regiões promotoras, ou domínios, é a análise de ganho ou perda de função, na qual ocorre fusão desses domínios individuais com promotores mínimos, como a utilizada por Benfey *et al.* (1989) para caracterização dos domínios que são responsáveis pela expressão tecido-específica do promotor CaMV 35S. No estudo citado, o autor constatou que o domínio A confere a expressão raiz-específica em plantas de tabaco e o domínio B é responsável pela expressão forte em cotilédones e no tecido vascular do hipocótilo. Juntos, esses dois domínios, além de conferirem expressão mais forte nesses tecidos, dirigiram a expressão em células adicionais (Benfey *et al.*, 1989).

## **1.4. Promotores constitutivos**

Transgenes sob a regulação de promotores constitutivos são ativos na maioria das células durante todo o ciclo celular embora o nível de expressão dependa do tipo celular (Twyman *et al.*, 2003). Esse tipo de promotor é utilizado quando se deseja obter grande quantidade de proteína, como na produção de biofármacos em plantas (Twyman *et al.*, 2003).

O promotor mais utilizado nessa categoria devido à facilidade de obtenção (Potenza *et al.*, 2004) é o CaMV 35S, o promotor do RNA 35S do vírus causador do mosaico da couve-flor, que é ativo na maioria dos tecidos e fases do desenvolvimento dos vegetais (Odell *et al.*, 1985). Outro fator que contribui para a ampla utilização desse promotor é a caracterização detalhada dos subdomínios que o compõem, assim é possível a manipulação destes para modular sua atividade temporal e espacial (Benfey *et al.*, 1990; Kay *et al.*, 1987). A utilização de promotores de origem viral na engenharia de plantas possui limitações no que diz respeito ao silenciamento transcricional do gene sob sua regulação, pois as células vegetais podem reconhecer essa sequência como estranha e promover sua metilação, excisão ou remodelamento da cromatina na região onde ela foi inserida (Potenza *et al.*, 2004).

Tal problema pode ser contornado com a utilização de promotores de origem vegetal capazes de atribuir expressão constitutiva ao gene. Os mais utilizados são os promotores dos genes constitutivos da ubiquitina, actina e citocromo c (Cominelli e Tonelli, 2010). A seleção desses promotores é devida à conservação da sequência desses genes entre espécies de plantas e à presença desses genes em quase todas as células vegetais (Potenza *et al.*, 2004). No entanto, esses promotores também não estão isentos dos mecanismos de silenciamento transcricional, visto que a planta pode reconhecer promotores provenientes de espécies distantes como estranha, semelhante ao que ocorre com promotores de origem viral (Potenza *et al.*, 2004).

Promotores ativos em determinados tecidos de dicotiledôneas podem não o ser em monocotiledôneas ou ser ativos em diferentes tecidos ou em níveis diferentes (Potenza *et al.*, 2004). Dessa forma, é de interesse que promotores constitutivos sejam isolados das mesmas espécies, ou de espécies próximas, nas quais serão utilizados para expressar transgenes.

A expressão, em todos os tecidos vegetais, de genes somente expressos em situações de estresse, apesar de aumentar a resistência da planta, pode causar retardamento no crescimento ou ocasionar problemas no desenvolvimento (Chen *et al.*, 2009; Cominelli e Tonelli, 2010). Além disso, proteínas inseticidas expressas constitutivamente em culturas podem aumentar a resistência dos insetos a elas, anulando o efeito positivo do uso de plantas transgênicas com essas características (Potenza *et al.*, 2004).

Além dos efeitos negativos que a expressão constitutiva de transgenes pode causar no desenvolvimento da planta, os promotores constitutivos utilizados na engenharia genética não

são ubíquos e, portanto, incapazes de ativar a expressão de transgene a níveis desejáveis em determinados tecidos e órgãos (Cominelli e Tonelli, 2010; Twyman, 2003). Assim, faz-se necessário o emprego de promotores órgão-específicos, ou promotores que tenham atividade preferencial em um tecido ou órgão. O emprego de sequências regulatórias induzíveis e/ou órgão específicas é feito com o objetivo de diminuir esses efeitos negativos da expressão constitutiva.

### **1.5. Promotores órgão-específicos**

O controle preciso da expressão do transgene por promotores órgão ou tecido específicos traz, entre outras vantagens, o aumento da biossegurança (Daniell, 2002) e também a redução de preocupações com o uso comercial dos transgênicos por parte do público. No caso de toxinas para pragas é possível limitar sua expressão somente ao órgão alvo desses organismos, impedindo a presença das toxinas no produto que será consumido pela população e também em outros órgãos que são utilizados na alimentação de outros animais, diminuindo, então, a probabilidade de que organismos não alvos sejam afetados (Potenza *et al.*, 2004).

Diversos promotores órgão/tecido específicos já foram isolados visando aplicação na produção de plantas transgênicas:

- Promotores específicos de frutos podem ser utilizados para expressar no órgão vacinas e micronutrientes, bem como alterar seu conteúdo nutritivo, aroma e sabor (Potenza *et al.*, 2004);
- Promotores semente-específicos são utilizados para aumentar seu teor nutritivo ou o conteúdo de algum composto de interesse industrial, como óleo e compostos fenólicos e também para melhorar a qualidade do grão (Potenza *et al.*, 2004);
- Promotores específicos de tecidos verdes são úteis na análise da fotossíntese e no caso de folhas, para aumentar a tolerância a estresse hídrico (Cominelli e Tonelli, 2010; Twyman, 2003);
- Promotores específicos de flores são usados para agregação de valor comercial como a mudança de cor e aroma (Potenza *et al.*, 2004);
- Promotores específicos de pólen ou *tapetum* (antera) são utilizados para causar esterilidade masculina da planta (Daniell, 2002);

- Promotores específicos de raiz são desejáveis para modular entrada de nutrientes, para aumentar a tolerância da planta a estresse hídrico ou a solos pobres e desenvolver estratégias de bioremediação (Ghanem *et al.*, 2011; Twyman, 2003).

Além desses usos, promotores específicos podem dirigir a expressão de genes que conferem resistência a patógenos de maneira dirigida (Twyman, 2003). Na produção de biofármacos, o uso de promotores órgão-específicos é importante para expressar o gene naqueles órgãos que são capazes de produzir a proteína de forma apropriada (Twyman *et al.*, 2003).

Apesar de existir diversos promotores de origem viral, microbiana e vegetal capazes de dirigir a expressão órgão-específica em plantas, ainda é desejável que esses promotores tenham origem na mesma espécie de planta ou em espécies filogeneticamente próximas, pois os sistemas regulatórios são únicos, podendo não agir da maneira prevista em espécies heterólogas distantes (Tyagi, 2001) ou ainda, estarem sujeitas ao silenciamento (Cammue *et al.*, 2005).

### **1.5.1. Promotores órgão-específicos de soja**

Alguns promotores de soja capazes de dirigir a expressão gênica órgão-específica já foram isolados e caracterizados biologicamente. Strömvik *et al.* (1999) isolaram o promotor do gene *Msg* de soja que possui atividade preferencialmente em vagem sem ativar a expressão em semente, podendo ser utilizado na geração de plantas resistentes a doenças que atacam a vagem sem afetar a semente. O promotor isolado do gene *SRSP1* (rubisco) de soja (Quandt *et al.*, 1992) foi capaz de ativar expressão específica do gene *arsC* de *Escherichia coli*, cujo produto catalisa a remediação de arsênico, em folhas de arábida (Dhankher *et al.*, 2002) e dirigir a expressão de *gus* a tecidos fotossinteticamente ativos em macieira (Gittins *et al.*, 2000).

Os promotores dos genes das subunidades da  $\alpha$ -conglucina (promotores SSU) de soja regulam a expressão no embrião em estágios de maturação diferentes. O promotor do gene da subunidade  $\beta$  da  $\alpha$ -conglucina foi utilizado no sistema de expressão heteróloga do hormônio do crescimento humano (hGH) em sementes de soja (Cunha *et al.*, 2011). A região promotora do gene de lecitina de semente de soja *Le1* também é capaz de regular a expressão específica nesse órgão (Saeed *et al.*, 2008). Preiszner *et al.* (2001) e Komatsu *et al.* (2011) relataram que os promotores dos genes raiz-específicos *Adh1* (álcool desidrogenase) e *Adh2* são os

responsáveis pelo padrão de expressão específico desses genes e são induzíveis por estresse como anoxia e alagamento, respectivamente.

O promotor do gene da isoflavona sintase (IFS) também é capaz de ativar a expressão raiz específica em plantas de soja, além de ser responsiva a hormônios e à nodulação (Subramanian *et al.*, 2004). A expressão gênica específica em nódulos formados nas raízes de plantas de soja após infecção com rizóbio também foi alcançada com o uso promotores dos genes ENOD2 (Lauridsen *et al.*, 1993) e do gene leghemoglobina lbc3 de soja (Ramlov *et al.*, 1993).

O número de promotores de soja caracterizados para aplicações biotecnológicas tende a aumentar com o avanço de pesquisas na área genômica. A disponibilidade das sequências e anotações genômicas de soja e de outras culturas de valor agrônômico juntamente com recente avanço no sequenciamento de alto desempenho do transcriptoma facilitaram a identificação de genes diferencialmente expressos (Priest *et al.*, 2009; Schmutz *et al.*, 2010; Tran e Mochida, 2010). Dessa forma, é possível prever e isolar regiões regulatórias de genes diferencialmente expressos sob estresse, entre os órgãos e envolvidos em vias metabólicas para aplicação na engenharia genética de soja (Priest *et al.*, 2009; Tran e Mochida, 2010). O estudo feito por Hernandez-Garcia *et al.* (2010) proporciona um panorama desse novo cenário e de como dados virtuais podem ser utilizados para isolamento de promotores em larga escala. Os autores identificaram membros das famílias gênicas de ubiquitina e ERF (*Ethylene Response Factor*) no genoma da soja e isolaram a região promotora de dez membros de cada família, que foram capazes de ativar expressão constitutiva no ensaio de expressão transiente. Apesar de facilitar a identificação e isolamento de promotores, dados obtidos por meio de análises de bioinformática precisam ser validados biologicamente.



## **2. Objetivos**

## 2.1. Objetivo Geral

Isolar e caracterizar promotores de plantas de soja responsáveis pela expressão gênica específica em raiz, folha ou semente.

## 2.2. Objetivos específicos

- Identificar *in silico* genes de soja órgão-específicos utilizando diferentes bancos de dados;
- Avaliar a expressão dos genes pré-selecionados em diversos órgãos de planta de soja;
- Isolar fragmentos da região promotora, dos genes selecionados, a partir da extremidade 5’;
- Clonar os fragmentos das regiões promotoras obtidos em vetores binários a montante do gene repórter *gus* (*uidA*);
- Transformar plantas de fumo (*Nicotiana tabacum*) com os cassetes de expressão;
- Caracterizar os promotores isolados por meio de avaliação da atividade da proteína GUS.

### **3. Materiais e Métodos**

### 3.1. Identificação de *contigs* órgão-específicos

Os contrastes virtuais (*Northern* eletrônico) para identificação de sequências específicas e abundantes de raiz, folha, semente e vagem foram realizados no banco de ESTs (*expressed sequence tags*) de soja (<https://alanine.cenargen.embrapa.br/Soja001>) da Embrapa Recursos Genéticos e Biotecnologia. O banco reúne *contigs* formados a partir de ESTs disponíveis em domínio público (Shoemaker *et al.*, 2002) e possui uma ferramenta para realização de *Northern* virtual que utiliza o Teste Exato de Fisher para determinar a significância estatística ( $P = 0,05$ ) dos resultados. Por meio dessa ferramenta é possível obter a frequência de ESTs que formam um *contig* (sequência de um transcrito putativo). As bibliotecas com sequências de cDNA provenientes de um determinado órgão foram agrupadas e contrastadas com as sequências das bibliotecas de cDNA dos demais. Assim, foram realizados quatro contrastes entre os grupos raiz versus (*vs.*) não raiz, folha *vs.* não folha, vagem *vs.* não vagem e semente *vs.* não semente. Bibliotecas de cDNA de tecidos não diferenciados e de plântulas não foram inseridas em nenhum grupo.

### 3.2. Análises virtuais dos *contigs* órgão-específicos

Os *contigs* resultantes do *Northern* eletrônico foram comparados por meio de BLASTn (Altschul *et al.*, 1997) com as sequências presentes no banco de dados do genoma estrutural da soja, construído por sequências obtidas durante a execução do projeto Genoma da Soja (GenoSoja) (<http://bioinfo03.ibi.unicamp.br/soja/>). Tal banco utiliza o teste de significância Audic-Claverie (Audic e Claverie, 1997) para determinar a frequência dos ESTs dentro das diversas bibliotecas. Dessa maneira, os *contigs* obtidos pelo *Northern* eletrônico realizado no banco de dados da Embrapa Recursos Genéticos e Biotecnologia que possuíam sequências correspondentes no banco GenoSoja foram novamente analisados quanto ao seu perfil de expressão.

Os *contigs* selecionados foram comparados ao banco de sequências não redundantes do NCBI (<http://www.ncbi.nlm.nih.gov>) utilizando o BLASTn (Altschul *et al.*, 1997). Desta forma, foram encontrados os transcritos (CDS) correspondentes (100% de identidade de sequência) aos *contigs* nesse banco. Os transcritos idênticos aos *contigs* foram também analisados no Unigene (<http://www.ncbi.nlm.nih.gov/UniGene/>) quanto ao perfil de expressão.

As bibliotecas de cDNA de tecidos não diferenciados foram desconsideradas nas análises.

### 3.2.1. Anotação funcional e mapeamento

Aqueles *contigs* que apresentaram expressão preferencial em algum órgão em pelo menos dois dos três bancos (<https://alanine.cenargen.embrapa.br/Soja001>; <http://bioinfo03.ibi.unicamp.br/soja> e <http://www.ncbi.nlm.nih.gov/UniGene/>) foram alinhados por meio do BLASTn (Altschul *et al.*, 1997) ao genoma da soja cv. Williams 82 (Schmutz *et al.*, 2010) disponível no banco genômico Phytozome (<http://www.phytozome.net/>). O banco fornece informações quanto ao mapeamento dos transcritos no genoma por meio do *Gbrowse* e anotação funcional destes com dados obtidos nas plataformas de anotação: *Pfam*, *Panther*, *KOG*, *GO*. Assim, a sequência do gene correspondente ao *contig* (identidade >90 % no alinhamento) foi localizada no genoma e sua anotação funcional obtida.

### 3.3. Desenho dos iniciadores para RT-PCR

A interface da *web* Primer3 (Rozen e Skaletsky, 2000) foi utilizada para o desenho dos iniciadores (*primers*) a serem usados nas análises de expressão dos genes escolhidos. Os parâmetros determinados para o desenho dos iniciadores foram: (1) tamanho de 18-21 pb (pares de bases), (2) amplicom de 390-450 pb, T<sub>m</sub> de 57 a 63 °C. Todos os iniciadores foram desenhados em regiões dos *contigs* com sequência considerada de boa qualidade.

Para o isolamento de fragmentos da região promotora dos genes, os parâmetros foram determinados de acordo com as especificações indicadas pelo sistema de clonagem Gateway (Invitrogen™). O iniciador anti-senso (*reverse*) foi desenhado na região de -1 até -20 a -33 pb a montante (*upstream*) do sítio de iniciação da tradução. Dessa forma, os iniciadores senso tiveram os parâmetros determinados para que o T<sub>m</sub> (*melting temperature*) e tamanho tivessem valores próximos àqueles dos iniciadores anti-senso. Foi adicionado aos iniciadores senso a sequência CACC, necessária para a clonagem no vetor pENTR™ (Invitrogen™).

### 3.4. Material Vegetal

Raízes, folhas jovens (não expandidas) e folhas maduras (expandidas e saudáveis), vagens e sementes de plantas de soja (*Glycine max* cv. Conquista), cultivadas em casa de

vegetação da Embrapa Recursos Genéticas e Biotecnologia, foram coletadas em diferentes estágios de desenvolvimento. As vagens foram coletadas de plantas nos estágios reprodutivos R4, R5 e R6 e as sementes nos estágios R5, R6 e R7 (Fehr e Caviness, 1977). Tais estágios foram determinados baseados no florescimento, desenvolvimento das vagens e semente e maturação da planta, segundo Fehr e Caviness (1977). No estágio R4, a vagem possui 2 cm. Nos demais, as vagens são identificadas de acordo com o desenvolvimento das sementes. No estágio R5 a semente possui 3 mm, no R6 a semente verde completa a cavidade da vagem e no R7 é o início da maturidade, no qual a vagem possui a cor amarronzada e a semente já atingiu seu tamanho final, porém não a cor. Foram pesados dois gramas de cada órgão, congelados imediatamente em nitrogênio líquido e armazenados a -80 °C. As sementes e vagens foram separadas antes do congelamento.

### **3.5. Extração de DNA**

O DNA genômico de soja foi extraído pelo método CTAB (Doyle e Doyle, 1987) modificado. Primeiramente, 2,5 g de folhas jovens congeladas em N<sub>2</sub> líquido foram maceradas. O pó obtido foi transferido para um tubo de propileno e ressuspenso em 20 mL do tampão de extração CTAB (NaCl 1,4 M; Tris-HCl 100mM; CTAB 2%; EDTA 0.05M; PVP 1%; -mercaptoetanol 1%) pré-aquecido a 65 °C. As amostras foram incubadas a essa temperatura por uma hora. Em seguida, foi adicionado um volume de fenol ácido: clorofórmio: álcool isoamílico (25:24:1) e os tubos foram agitados manualmente por inversão três vezes. O material foi centrifugado por dez minutos a 6869 × g. A fase aquosa foi transferida para um novo tubo contendo um volume de clorofórmio: álcool isoamílico (24:1) e novamente centrifugada nas mesmas condições. A fase superior foi, então, transferida cuidadosamente para um novo tubo, no qual foi adicionado um volume de isopropanol e misturado por inversão vagarosamente. Em seguida o tubo foi centrifugado por 30 minutos a 8694 × g. O sobrenadante foi descartado e ao sedimento foram adicionados 10 mL de etanol 70% e centrifugado por cinco minutos a 6869 × g. O etanol 70% foi descartado por inversão do tubo e o sedimento foi deixado secar a temperatura ambiente. Por fim, o DNA foi ressuspenso em água livre de DNases e incubado com 2 µL de RNase A [100 µg/mL] por 30 minutos a 37 °C. A quantificação do DNA e determinação de sua pureza foram realizadas por espectrofotometria (NanoDrop DN-1000) nos comprimentos de onda de 260, 280, 260/280, 260/230nm.

### 3.6. Extração de RNA

O RNA total de raiz, folha jovem, folha madura, vagem nos estágios R4, R5 e R6 e semente nos estágios R5, R6 e R7 foi extraído individualmente. As amostras foram maceradas em N<sub>2</sub> líquido até a obtenção de um pó fino. O pó fino resultante foi transferido para uma solução contendo 9 mL de tampão de extração NTES (NaCl 0.1 M; Tris-HCl 0.01 M; EDTA 0,5 M pH8; SDS 1%; H<sub>2</sub>O DEPC 0.1%) e 6 mL de fenol pH 8: clorofórmio: álcool isoamílico (25:24:1). Os tubos foram agitados vigorosamente por 15 minutos e em seguida centrifugados a 10733 × g por dez minutos. A fase aquosa foi transferida para tubos novos nos quais foram adicionados 1/10 do volume do sobrenadante de NaAc 3M pH4,5 e dois volumes de etanol absoluto. As amostras foram mantidas por duas horas a -20 °C e depois centrifugada a 10733 × g por dez minutos. O sobrenadante foi descartado e o precipitado dissolvido em 2,5 mL de água estéril tratada com DEPC 0,1%. A precipitação do RNA foi feita com LiCl 4M a 4 °C por aproximadamente 15 horas. Em seguida, o material foi centrifugado a 10733 × g por 15 minutos. Por fim, o sedimento foi lavado em etanol 70% e ressuspendido em 200 µL de água tratada com DEPC 0,1%.

As análises de pureza, quantificação e integridade do material foram feitas por espectrofotometria (NanoDrop DN-1000) e eletroforese em gel de agarose 1,5 % em condições desnaturantes (formaldeído). Todas as soluções utilizadas, exceto Tris e solventes orgânicos, foram tratadas com DEPC 0.1%.

### 3.7. Reação de RT-PCR semiquantitativa

O RNA total extraído foi tratado com a *DNase I Amplification Grade* (Invitrogen™). Amostras de 5 µg de RNA foram tratadas em reação contendo 10X *Dnase I Reaction Buffer*, 5 unidades de *DNase I Amplification Grade* (1U/ µL) e água tratada com DEPC em volume final de 10 µL. A reação foi incubada a 25 °C por 15 minutos. A enzima foi inativada pela adição de EDTA para uma concentração de 12.5 mM seguido de aquecimento a 65 °C por dez minutos.

A primeira fita do cDNA foi sintetizada por transcrição reversa a partir do RNA de raiz, folha, vagem e semente de planta de soja utilizando a enzima *SuperScript III Reverse Transcriptase* (Invitrogen™) conforme protocolo do fabricante. Em tubos de microcentrífuga foram adicionados inicialmente 300 ng de Oligo(dT), 2 µg de RNA de um dos órgãos tratados

com DNase e 0.8 mM de cada dNTP (desoxirribonucleotídeo trifosfatado). A reação foi mantida a 65 °C durante cinco minutos. Em seguida, adicionou-se *First-strand buffer* 1X, DTT 5 mM e 200 unidades da transcriptase reversa. A síntese ocorreu a 50 °C por uma hora. Esse procedimento foi realizado para a síntese de cDNA a partir do RNA total extraído de cada órgão. Os produtos das reações foram diluídos 20 vezes e 5 µL foram utilizados nas reações de PCR.

Para a reação de RT-PCR semiquantitativa, utilizou-se, além do par de iniciador específico (Tabela 1) para o transcrito alvo, outro par de iniciador para amplificar um fragmento do gene de actina, gene expresso constitutivamente. Todas as PCRs foram montadas em duplicatas de reação para cada cDNA de raiz, folha, vagem ou semente contendo os seguintes reagentes: 1X do tampão para PCR, 0.4 mM de cada dNTP (Invitrogen™), 0.4 µM de cada iniciador citado acima, 1.5 unidades de *Taq polymerase* (Invitrogen™), 5 µL de cDNA diluído, 3 mM de MgCl<sub>2</sub> e água estéril (Milli-Q), no volume final de 25 µL. As condições da PCR foram: 94 °C por um minuto seguido pelos ciclos de 94 °C por 30 segundos, 55-57°C por 30 segundos e 68 °C por dois minutos. Os produtos de PCR (15 µL) foram separados em gel de agarose 2% corado com brometo de etídio e visualizados sob luz ultravioleta (UV). O número de ciclos de PCR foi otimizado para assegurar que as reações de amplificação fossem paradas na fase exponencial de amplificação do produto.

**Tabela 1.** Sequências dos iniciadores específicos utilizados nas reações de RT-PCR semiquantitativa.

Gene	Iniciador	Sequência	Tamanho do fragmento a ser amplificado
<i>GmSulfT1</i>	RzGmSdF3	5'-GTGGCCACAATTCCAAAATC-3'	423 pb
	RzGmSdR4	5'-CCAAGTTGGACCAAACCCTA-3'	
<i>GmCit1</i>	FoGm1SF1	5'-TTACAGAAGCACACCACAGG-3'	419 pb
	FoGm1SR1	5'-GACTCGAAGAAGGACAAACC-3'	

### 3.8. Northern blot

O RNA foi fracionado em gel de agarose 1,5% em condições desnaturantes (formaldeído) e tampão MOPS (MOPS 0,2 M, AcNa 50 mM, EDTA 10mM). Foram aplicados por poço do gel, 20 µg de RNA total dissolvido em tampão de amostra (ficol 30%, EDTA 0,5M pH 8,0, azul de bromo fenol 0,025%, formamida 30,1%, glicerol 2% e brometo



de etídio 0,1%). Após a eletroforese o RNA total de soja foi transferido a vácuo para membrana de náilon (Hybond – N, Amersham Bioscience). O tampão de transferência utilizado foi o SSPE 10X (NaCl 1,5M; NaH<sub>2</sub>PO<sub>4</sub> 0,1M; Na<sub>2</sub>·EDTA·2H<sub>2</sub>O 10mM). A transferência foi realizada durante quatro horas à pressão de 5 mm Hg. Ao final, a membrana foi incubada por cinco minutos em SSPE 2X e o RNA foi fixado à membrana por exposição à luz UV (UV Stratalinker 1800 - Stratagene) durante 30 segundos.

As sondas dos *Northern blots* foram feitas com os mesmos fragmentos obtidos na RT-PCR, cujos produtos possuíam aproximadamente 400 pb (Tabela1). Os fragmentos foram purificados utilizando o kit *Wizard<sup>®</sup> SV Gel and Clean Up System* (Promega). Cinquenta ng de cada fragmento foram desnaturados por cinco minutos a 95-100 °C e incubados no gelo por mais cinco minutos. Em seguida, o fragmento desnaturado foi adicionado no kit de marcação *Ready to Go* (Amersham Bioscience) juntamente com 5 µL de dCTP -P<sup>32</sup> (50 µCi), conforme especificações do fabricante. A reação foi incubada a 37 °C por 40 minutos. Após o período, a sonda foi desnaturada por cinco minutos a 95-100 °C e imediatamente colocada no gelo por dez minutos. Em seguida, a sonda foi adicionada à membrana contendo o RNA previamente pré-hibridizada com o tampão de hibridização *ULTRAHyb Ultrasensitive* (Applied Biosystems) a 42 °C por quatro horas. A hibridização ocorreu durante a noite à mesma temperatura.

A membrana foi lavada a 42 °C duas vezes por 15 minutos com a solução de lavagem 2X (SSC 2X; SDS 0,1%) e duas vezes com a solução de lavagem 0,1 X (SSC 0,1X; SDS 0,1%). Em seguida, foi exposta ao Imaging Plate (IP BAS – SR 2040) por aproximadamente quatro horas, momento em que a radioatividade presente na membrana foi capturada e fotodocumentada pelo equipamento FLA 3000 (Fujifilm).

### **3.9. Isolamento dos promotores**

No Phytozome (<http://www.phytozome.net>) foi possível obter a sequência da região a montante (*upstream*) das sequências codificadoras (CDS) dos genes selecionados para isolamento de suas regiões promotoras. Assim, as regiões promotoras foram obtidas por meio do mapeamento dos genes selecionados no *Gbrowse* (<http://www.phytozome.net/cgi-bin/gbrowse/soybean>) do genoma da soja cv. Williams 82 (Schmutz *et al.*, 2010). Assim, os 3000 pb a montante da extremidade 5' da CDS dos genes foram utilizados para o desenho de iniciadores. Foram amplificados fragmentos com iniciadores desenhados especificamente para gerar fragmentos de tamanhos distintos, mas sempre contendo a região 5' UTR do transcrito e

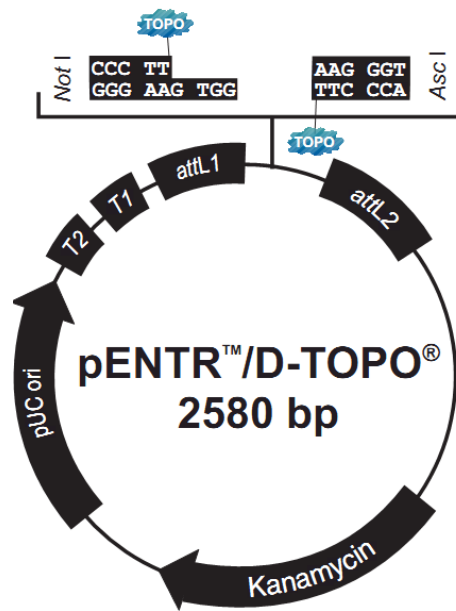
o início da região promotora. O tamanho dos fragmentos dependeu das sequências favoráveis ao desenho dos iniciadores.

### **3.10. Análise *in silico* das regiões promotoras**

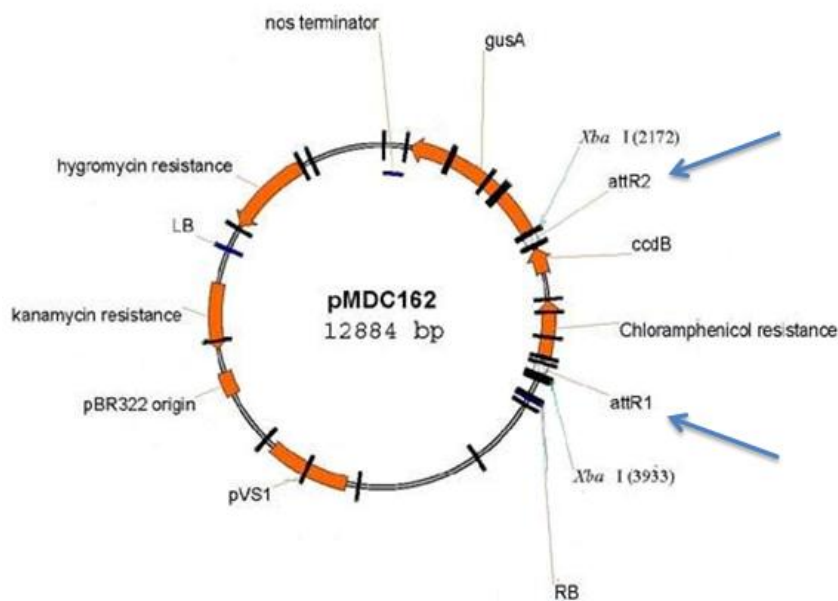
Os elementos *cis* regulatórios putativos foram obtidos pelas análises das sequências promotoras nos bancos de motivos regulatórios *PLACE* (<http://www.dna.affrc.go.jp/PLACE/>) (Higo *et al.*, 1998) e *RegSite Plant DB* (<http://linux1.softberry.com/berry.phtml?topic=tssp&group=programs&subgroup=promoter>) (Softberry, Inc.) por meio das ferramentas TSSP (predição de promotores de plantas) e NSITE- PL (reconhecimento de motivos regulatórios de plantas).

### **3.11. Clonagem dos promotores**

Os fragmentos da região promotora foram clonados em vetor binário por meio do sistema Gateway<sup>®</sup>, baseado na recombinação sítio-específica do bacteriófago  $\lambda$ . Esse sistema consiste em transferir um fragmento de DNA inserido em um vetor de entrada para um vetor destino por meio de recombinação sítio-específica. Assim, os sítios attL1 e attL2 que flanqueiam a região do DNA a ser transferido no vetor de entrada recombinam-se, respectivamente, com os sítios attR1 e attR2 presentes no vetor destino e que flanqueiam o gene letal *ccdB*. Após a reação, o vetor de entrada irá conter o gene letal e o vetor destino conterá o fragmento de DNA de interesse. Foi utilizado como vetor de entrada o plasmídeo pENTR<sup>™</sup> (Figura 7) e como vetor de destino o plasmídeo binário pMDC162 (Figura 8) específico para uso em *Agrobacterium*.



**Figura 7.** Representação esquemática do vetor pENTR™. O sítio em que o fragmento de DNA é ligado, de modo que seja flanqueado pelos sítios de recombinação attL1 e attL2, está em destaque no esquema. Fonte: Invitrogen™ (2006).



**Figura 8.** Representação esquemática do vetor pMDC162. A figura mostra as regiões codificadoras que compõem o vetor, seu mapa de restrição e os sítios de recombinação attR1 e attR2. A recombinação entre os sítios attL1 e attL2 do vetor de entrada com os sítios attR1 e attR2 (indicados por setas) do vetor destino resultará na inserção do fragmento de DNA de interesse e na excisão do gene mortal *ccdB*. Fonte: Curtis e Grossniklaus (2003).

Os iniciadores desenhados a partir da sequência genômica (Tabela 2) foram utilizados para amplificar os diversos fragmentos da região promotora. A enzima *Platinum Pfx DNA Polymerase* (Invitrogen™) foi utilizada para catalisar a reação, no seguinte sistema: tampão da enzima 1X, 0.3 mM de cada dNTP, MgSO<sub>4</sub> 2 mM, 0.2 μM de cada iniciador, 500 ng de DNA genômico e 0,5 U de *Pfx* no volume final de 25 μL. O mesmo iniciador anti-senso foi combinado com todos os iniciadores senso do respectivo promotor para gerar diferentes fragmentos a partir da região 5' (Tabela 2). A PCR ocorreu nos seguintes parâmetros: 94 °C por três minutos, 94 °C por 30 segundos, 53°C por 30 segundos e 68 °C por dois minutos.

**Tabela 2.** Sequências de iniciadores utilizados na clonagem dos promotores e deleções 5' dos genes *GmSulfT1* e *GmCit1*.

Região promotora do gene	Iniciador	Sequência	Tamanho do produto
<i>GmSulfT1</i>	RzGm1P1.8F2	5'-CACCACGTGTATGGTACTTTGAAC-3'	1868 pb
	RzGm1P1.4F1	5'-CACCAGGCATATCTTTGCATATCG-3'	1466 pb
	RzGm1P0.9F1	5'-CACCGGACAATGCAGGGATTATG-3'	985 pb
	RzGm1P0.8F1	5'-CACCTCCATCTCTTTGTTCTTTAC-3'	810 pb
	RzGm1P0.5F1	5'-CACCGGATTGGAGTTGCTTACTTG-3'	505 pb
	RzGm1PrR1 <sup>r</sup>	5'-TAGAGTGCTAATGGTCGATC-3'	
<i>GmCit1</i>	FoGm1P2.3F1	5'-CACCAGTACTACCAAATTGATTAGAAATTAGC-3'	2329 pb
	FoGm1P1.9F1	5'-CACCTATATTAGGATTAGGATTAGGATTTTGG-3'	1940 pb
	FoGm1P1.5F1	5'-CACCCATATTTCTACAAAAGTTTGCTTGG-3'	1575 pb
	FoGm1P1.2F1	5'-CACCTATAAATTAATAATTAGTCGTAACAGTGG-3'	1249 pb
	FoGm1P0.8F1	5'-CACCACTATTATTAGGACTTCTAGATTTTATGG-3'	822 pb
	FoGm1P0.4F1	5'-CACCTAACTCAACTAACAAAATGAATTGATGG-3'	413 pb
	FoGm1PrR1 <sup>r</sup>	5'-TATTTTCTGAAATGAGATAACATTACC-3'	

r: iniciadores anti-senso.

Os amplicons foram ligados ao vetor pENTR™ /D-TOPO® (Invitrogen™) em reação com volume final de 6 μL contendo de 0.5 a 4μL de produto de PCR, 1 μL de solução salina e 1 μL do vetor TOPO®. A reação foi incubada por dez minutos à temperatura ambiente e subsequentemente utilizada na transformação por choque térmico de células competentes de *Escherichia coli* One Shot® TOP10 (Invitrogen™), conforme as especificações do fabricante. Inicialmente, foram adicionados 2 μL da reação de ligação ao tubo de microcentrífuga contendo as células em estoque. Incubou-se o sistema no gelo por cinco minutos e, em seguida, a 42 °C por 30 segundos. Imediatamente após o choque térmico, adicionou-se ao

tubo de microcentrífuga 250  $\mu$ L de meio SOC (Sambrook e Russell, 2001) e a cultura foi incubada a 37 °C por uma hora. As células foram espalhadas em placas de *Petri* com meio LB-ágar (Sambrook e Russell, 2001) sólido e canamicina [50  $\mu$ g/mL] e incubadas a 37 °C por 14 horas.

A construção primária de fragmento da região promotora ligada ao vetor pENTR™ constitui o vetor de entrada e as células transformadas com ela são os clones de entrada.

### **3.11.1. Extração de DNA plasmidial**

As colônias contendo os vetores pENTR™, onde foram clonadas as sequências promotoras putativas, foram inoculadas em tubos de ensaio com 3 mL de meio LB (Sambrook e Russell, 2001) líquido seletivo (50  $\mu$ g/mL de canamicina) e mantidas em incubadora por 16 horas, com agitação de 200 rpm e temperatura de 37 °C. Alíquotas de 1,5 mL foram transferidas para tubo de microcentrífuga e centrifugadas por três minutos a 13792  $\times$  g (rcf). O sobrenadante foi descartado e adicionado o restante da cultura, ou seja, 1,5 mL. Após a centrifugação e a remoção do sobrenadante, o precipitado foi ressuscitado em 100  $\mu$ L de TE (Tris-HCl 10 mM pH 8; EDTA 1mM). Em seguida, adicionou-se 200  $\mu$ L de solução NaSE (NaOH 0,2 M; SDS 1%; EDTA 10 mM) e os tubos foram levemente agitados. Após cinco minutos à temperatura ambiente, foram adicionados 30  $\mu$ L de KAc 5 M pH 4,8. As amostras foram incubadas por cinco minutos no gelo e então centrifugadas por cinco minutos a 13 400  $\times$  g, 4 °C. Os sobrenadantes foram transferidos para tubos novos contendo 2  $\mu$ L de RNase A [10 mg/mL] e deixados por 20 minutos a 37 °C em banho-maria. As soluções foram, então, lentamente homogeneizadas com 450  $\mu$ L de LiCl 5M, incubadas por duas horas a -20 °C e centrifugadas por dez minutos a 13 400  $\times$  g, 4 °C. Novamente, os sobrenadantes foram transferidos para novos tubos contendo meio volume de isopropanol, deixados por cinco minutos à temperatura ambiente e submetidos à centrifugação por 15 minutos a 13792  $\times$  g (rcf). Os precipitados foram lavados com 400  $\mu$ L de etanol 70% em centrifugação a 13 400  $\times$  g a 4 °C por três minutos. Após a secagem do material, o DNA foi ressuscitado em 40  $\mu$ L de água deionizada.

Para confirmação da clonagem dos fragmentos de DNA, os plasmídeos foram digeridos com as enzimas *NotI* e *EcoRV* (GIBCO BRL™) em reação de 20  $\mu$ L contendo o tampão de reação 1X, três unidades de cada enzima e 5  $\mu$ L do DNA extraído dos clones de

entrada. As digestões ocorreram durante uma hora a 37 °C em banho-maria e foram analisadas por eletroforese em gel de agarose 1% corado com brometo de etídio.

### **3.11.2. Recombinação LR da tecnologia Gateway®**

Uma reação de recombinação sítio específica foi montada para que o fragmento da região promotora do vetor de entrada fosse transferido para o vetor binário, ou vetor destino.

O vetor destino utilizado foi o pMDC 162 (Curtis e Grossniklaus 2003) (Figura 8), doado pela Universidade de Zurique -Suíça.

A reação foi montada em um volume final de 8 µL com aproximadamente 150 ng do vetor de entrada linearizado com a enzima *EcoRV* ou *pvuI* (GIBCO BRL™) (no caso de a região promotora conter o sítio de reconhecimento da *EcoRV*), 150 ng do vetor destino pMDC162 e se necessário completada com tampão TE, pH8. Os componentes foram brevemente agitados e centrifugados e, então foram adicionados 2 µL de Gateway® LR Clonase™ II Enzyme Mix (Invitrogen™) conforme protocolo do fabricante. Novamente as amostras foram brevemente agitadas e centrifugadas. Em seguida, as reações foram incubadas por cinco horas a 25 °C. Para finalizar a reação adicionou-se ao tubo 1 µL de proteinase K e a reação foi incubada por dez minutos a 37 °C. O produto de interesse dessa reação de recombinação é o vetor binário pMDC162 contendo a região promotora a montante do gene *gus*.

Para gerar os clones de expressão (células transformadas com o vetor de expressão), foi feita a transformação por choque térmico das células quimicamente competentes OmniMAX™ (Invitrogen™) com o produto da reação LR. Em um tubo de microcentrífuga contendo 100 µL de células competentes foram adicionados 5 µL da reação LR e a mistura foi incubada por 30 minutos em gelo. Após esse período, foi dado o choque térmico por 90 segundos a 42 °C e adicionados 500 µL de meio SOC (Sambrook e Russell, 2001). A cultura foi incubada por uma hora a 37 °C. As células foram, então, precipitadas por centrifugação a 13792 × g por 1 minuto, uma parte do meio sobrenadante foi descartada e as células foram ressuspensas no restante de meio ( 100 µL). As células foram espalhadas em placas de *Petri* com meio LB (Sambrook e Russell, 2001) sólido e canamicina [50 µg/mL] e incubadas a 37 °C por 14 horas. As colônias resistentes foram inoculadas em 3 mL de meio LB para isolamento do plasmídeo conforme protocolo descrito acima. O vetor de expressão (vetor destino pMDC162 contendo os diferentes fragmentos da região promotora), por sua vez, foi

digerido com a enzima *XbaI* em 30  $\mu\text{L}$  de reação com tampão de reação 1X, 20 unidades de enzima e 20  $\mu\text{L}$  dos vetores pMDC162. Igualmente ao vetor de entrada as digestões do pMDC162 ocorreram durante uma hora a 37 °C e foram analisadas por eletroforese em gel de agarose 1% corado com brometo de etídio.

### **3.12. Transformação de *Agrobacterium tumefaciens***

Os plasmídeos extraídos dos clones positivos foram inseridos em células competentes de *Agrobacterium tumefaciens* da estirpe GV3101 por meio de eletroporação. Em um tubo para microcentrífuga contendo 40  $\mu\text{L}$  de células foi adicionado 1  $\mu\text{L}$  (50-300 ng/ $\mu\text{L}$ ) de plasmídeo e a mistura transferida para uma cuveta de 0,2 cm. Em seguida as células foram submetidas a pulso elétrico de 25  $\mu\text{F}$ , 2,5 kV, 200 , 5,5 segundos e logo em seguida foi adicionado 1 mL de meio SOC (Sambrook e Russell, 2001) à cubeta de eletroporação. A cultura foi transferida para novo tubo de microcentrífuga e incubada por 60 minutos a 28 °C para que as células pudessem recompor sua membrana. Após esse período, 30 e 100  $\mu\text{L}$  da cultura foram espalhadas em duas placas de *Petri* contendo meio LB-ágar (Sambrook e Russell, 2001) com canamicina [100  $\mu\text{g}/\text{mL}$ ], gentamicina [50  $\mu\text{g}/\text{mL}$ ] e rifampicina [100  $\mu\text{g}/\text{mL}$ ] e incubadas por aproximadamente 48 horas a 28 °C.

### **3.13. Hibridização de colônias**

Para confirmar se as células de *Agrobacterium tumefaciens* transformadas com os vetores de expressão continham de fato o plasmídeo, foi realizado o ensaio de hibridização de colônias. Primeiramente, colônias isoladas foram repicadas em membrana de náilon (Hybond – N, Amersham Bioscience) depositada em cima de uma placa de *Petri* com meio LB-ágar, canamicina [100  $\mu\text{g}/\text{mL}$ ], gentamicina [50  $\mu\text{g}/\text{mL}$ ] e rifampicina [100  $\mu\text{g}/\text{mL}$ ] e incubada por 15 horas a 28 °C. As mesmas colônias foram também repicadas para uma placa semelhante contendo o mesmo meio LB-ágar e os mesmos antibióticos sem, no entanto, conter a membrana. As colônias foram posicionadas na placa da mesma forma que na membrana de modo a possibilitar o resgate das colônias positivas. Após a incubação, as placas contendo membrana foram incubadas por 30 minutos a 4 °C para facilitar a retirada da membrana sem o meio de cultura.

As membranas foram colocadas em uma superfície plana, em cima de um filme de PVC, onde foram adicionadas as soluções. Sempre ao trocar de solução a membrana foi

secada rapidamente em papel filtro. Inicialmente a membrana foi colocada com as colônias voltadas para cima em 2 mL de NaOH 0,5 M durante três minutos por duas vezes para a lise das células. Em seguida a membrana foi colocada sobre 2 mL de Tris 1 M pH 7,4 por dois minutos e então em uma solução de NaCl 1,5 M e Tris 0,5 M pH 7,4 por três minutos. As membranas foram cobertas com 2 mL de NaCl 0,3M por três minutos e depois incubadas na solução SET 2X (Tris 60 mM pH 8,0, EDTA 2 mM pH 8,0, NaCl 0,3 M) por aproximadamente cinco minutos. Em um funil coberto com discos de papel filtro a membrana foi lavada duas vezes com SET 2X e uma vez com etanol 70% gelado. Após a secagem da membrana, o DNA das colônias foi fixado à membrana por exposição à luz UV (UV Stratalinker 1800 - Stratagene) durante 30 segundos.

As sondas utilizadas na hibridização da membrana foram os menores fragmentos (Tabela 2) da região promotora dos genes selecionados amplificados por PCR. O preparo e marcação da sonda com dCTP -P<sup>32</sup>, a pré-hibridização da membrana, hibridização da sonda à membrana e a lavagem desta foram feitas de forma semelhante ao do ensaio de *Northern blot* descrito acima.

Após o processo de lavagem, a membrana foi exposta ao Imaging Plate (IP BAS – SR 2040) por aproximadamente duas horas, momento em que a radioatividade presente na membrana foi capturada e fotodocumentada pelo equipamento FLA 3000 (Fujifilm).

### **3.14. Cultura de tecidos**

#### **3.14.1. Transformação de plantas de *Nicotiana tabacum***

A transformação de *Nicotiana tabacum* foi feita segundo Barros, 1989, com modificações.

##### **3.14.1.1. Co-Cultura**

As colônias contendo os vetores binário pMDC 162 que contêm as regiões promotoras putativas a montante do gene *gus* foram inoculadas em 5 mL de meio LB (Sambrook e Russell, 2001), canamicina [100 µg/mL], gentamicina [50 µg/mL] e rifampicina [100 µg/mL] e incubadas a 180 rpm por aproximadamente 24 horas a 28 °C. Cinquenta microlitros desse pré-inóculo foram colocados em 50 mL de meio LB e novamente incubados a 180 rpm por aproximadas 15 horas a 28 °C.



Folhas jovens (terceira a partir do ápice) de plantas de *Nicotiana tabacum* foram coletadas de plantas com 55 e 80 dias cultivadas em casa de vegetação e imediatamente colocadas em água. Em capela de fluxo laminar vertical, as folhas foram lavadas com 1 litro de álcool 50% e em seguida desinfestadas em solução de hipoclorito de sódio 2% por 20 minutos. Após esse período as folhas foram lavadas três vezes em água Milli-Q estéril e mantidas em um béquer com água esterilizada enquanto não eram manipuladas.

Paralelamente, a cultura bacteriana foi distribuída em duas placas de *Petri* e deixada na capela. Em placa de *Petri* de 130 mm x 15 mm contendo papéis-filtro estéreis, a folha foi dividida ao meio, retirou-se a nervura central e as bordas com auxílio de lâmina de bisturi e então, o restante foi cortado em quadrados de 0,6 x 0,6 cm aproximadamente. À medida que eram cortados, os explantes eram mergulhados na cultura de *A. tumefaciens* até que as duas placas de *Petri* de 90 mm x 15 mm contendo a cultura bacteriana ficassem com toda a superfície coberta com os explantes foliares. Os controles negativos foram mergulhados em meio LB sem bactérias. Em seguida, os explantes foram transferidos para papel filtro a fim de retirar o excesso da cultura bacteriana e colocados em placas de *Petri* de 90 mm x 15 mm contendo o meio MS (SIGMA) (Murashige e Skoog, 1962) pH 5,6-5,8 com sacarose 3%, 6-benzilaminopurina (SIGMA) (1 mg/mL) e ágar (purificado para cultura de tecidos – SIGMA) 0,3%, com a superfície adaxial voltada para o meio. As placas foram fechadas com filme de PVC e colocadas em sala de cultura climatizada (28 °C), no escuro por dois dias.

#### **3.14.1.2. Seleção das plantas transformadas**

O vetor de expressão utilizado para a transformação de tabaco possui o gene *hpt*, cujo produto, a proteína higromicina fosfotransferase, confere à planta resistência à higromicina. Assim, após a co-cultura de dois dias, para selecionar os explantes transformados e eliminar as agrobactérias, os explantes foram transferidos para placas de *Petri* contendo meio MS (SIGMA) (Murashige e Skoog, 1962) pH 5,6-5,8 com sacarose 3%, 6-benzilaminopurina (SIGMA) (1 mg/mL) e ágar 0,3%, adicionado de cefotaxima [500 µg/mL] e higromicina [200 µg/mL]. As placas foram fechadas com filme de PVC e incubadas em sala de cultura climatizada (28 °C), sob fotoperíodo controlado de 16 horas de luz e oito horas de escuro. Nas placas com os controles negativos não foi adicionado higromicina. A cada duas semanas, período que os antibióticos começam a perder o efeito, os explantes eram transferidos para placas novas.

Após um mês da transferência para o meio seletivo os brotos regenerados e transformados foram transferidos para tubos de ensaio com meio MS (SIGMA) (Murashige e Skoog, 1962) pH 5,6-5,8; sacarose 2%; ágar 0,3%, contendo cefotaxima [300 µg/mL] e higromicina [200 µg/mL]. Os tubos com os brotos foram selados e incubados em sala de cultura nas mesmas condições anteriores. Os brotos não transformados (controle negativo) foram transferidos para os tubos contendo o mesmo meio de cultura usado para os brotos transformados, porém, sem a adição dos antibióticos.

#### **3.14.1.3. Aclimação das plantas transformadas**

As plantas que enraizaram nos tubos de ensaio foram transferidas para sacos pequenos com terra adubada quimicamente e úmida. Após a lavagem da raiz com água para retirar o meio de cultura, a planta foi colocada na terra e coberta com saco de plástico transparente. Durante esta etapa foram coletados segmentos de raiz e folha para a realização do teste histoquímico para detecção da atividade da enzima repórter Gus. As plantas foram mantidas em casa de vegetação com temperatura e fotoperíodo naturais. Os sacos plásticos transparentes foram abertos nas extremidades de modo progressivo a partir da primeira semana para permitir a aclimação gradativa das plantas às condições da casa de vegetação e ao final de duas semanas os sacos foram completamente retirados.

### **3.15. Ensaio Histoquímico**

Para caracterizar a atividade dos fragmentos dos promotores por meio da regulação da expressão do gene repórter *gus* foi realizado ensaio histoquímico segundo McCabe (1988). Os segmentos dos ápices radiculares e de folha, coletados durante a transferência das plantas do tubo de ensaio para a terra, foram incubados em solução contendo o substrato X-Gluc (5-bromo- 4- cloro- 3- indolil- - D- glucoronídeo) na concentração de 2 mM, ou seja: 100 mg X-Gluc foram dissolvidos em 2 mL de DMSO e adicionados em uma solução contendo EDTA 10 mM, NaH<sub>2</sub>PO<sub>4</sub> 100 mM, K<sub>4</sub>Fe(CN)<sub>6</sub> · 3H<sub>2</sub>O 0,5mM, Triton X-100 0,1%, ácido ascórbico 1% e água para completar o volume de 200 mL. O pH final da solução foi ajustado para pH 7,0 com NaOH 10 M e a solução finalmente filtrada em filtro estéril Millex<sup>®</sup> (membrana Millipore com poro de 45 µM) e armazenada a -20 °C. Os segmentos das raízes e folhas foram colocados em placas de ELISA de poços contendo 200 µL da solução e incubados em estufa a 37 °C por 18 horas. Após esse período a solução foi retirada com ajuda de pipeta automática e adicionado etanol 70% para retirada da clorofila e melhor visualização

do produto final da reação, o índigo *blue*. O etanol foi trocado várias vezes até a completa retirada da clorofila. O resultado do ensaio foi visualizado na lupa SteREO Discovery.V8 (Zeiss) e as imagens capturadas.

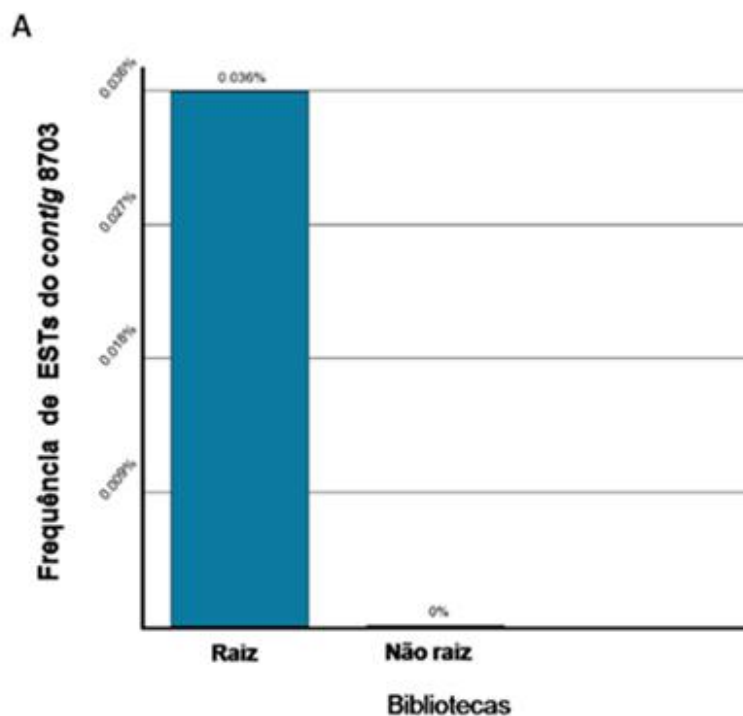
## **4. Resultados**

Transcritos com expressão preferencial em determinado órgão de *G.max* foram identificados por análises *in silico*. O *Northern* eletrônico realizado no banco de dados da Embrapa Recursos Genéticos e Biotecnologia (<https://alanine.cenargen.embrapa.br/Soja001/>) apontou 19 *contigs* específicos de raiz, sete *contigs* específicos de folha e 13 *contigs* específicos de semente, e nenhum *contig* específico de vagem. Após comparação desses *contigs* as sequências de outros dois bancos de dados: Unigene do banco de sequências NCBI (<http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Gma.6643>) e o banco do projeto Genoma da Soja - GenoSoja - (<http://bioinfo03.ibi.unicamp.br/soja/>), foram encontradas as sequências correspondentes (idênticas) e analisada a expressão das ESTs (*expressed sequence tags*) que as compõem. Os *contigs* que apresentaram perfil de expressão preferencial em algum órgão como resultado das análises em pelo menos dois dos três bancos de dados e cujas sequências ainda não foram estudadas (inéditas) foram selecionados como candidatos para o isolamento das respectivas regiões promotoras. Assim, dos 39 *contigs* iniciais foram selecionados três *contigs* com expressão preferencial em raiz, um em folha e três em semente para o isolamento dos seus respectivos promotores.

São apresentados nesta dissertação os resultados das análises de sequência de dois dos *contigs* selecionados dos quais foram isoladas as respectivas regiões promotoras: 8703, com perfil de expressão baseado em ESTs preferencial em raiz e 18151 com expressão preferencial em folha.

#### **4.1. Análises *in silico***

O *contig* 8703 foi selecionado no banco de ESTs de soja da Embrapa Recursos Genéticos e Biotecnologia por ser constituído apenas de sequências oriundas de bibliotecas de raiz. (Figura 9-A). O alinhamento desse *contig* a sequências do banco de dados UniGene (NCBI) por meio do BLAST (*Basic local alignment search tool*; Altschul et al. 1997) mostrou que ele corresponde ao gene Gma 6643. Análises de expressão no UniGene (*EST profile*) do Gma 6643 mostraram perfil de expressão preferencial em raiz, com 288 transcritos por milhão (TPM) e reduzida expressão em cotilédone (33TPM) e caule (48TPM); sem expressão nos outros órgãos analisados como pode ser verificado na Figura 9-B.



**B**

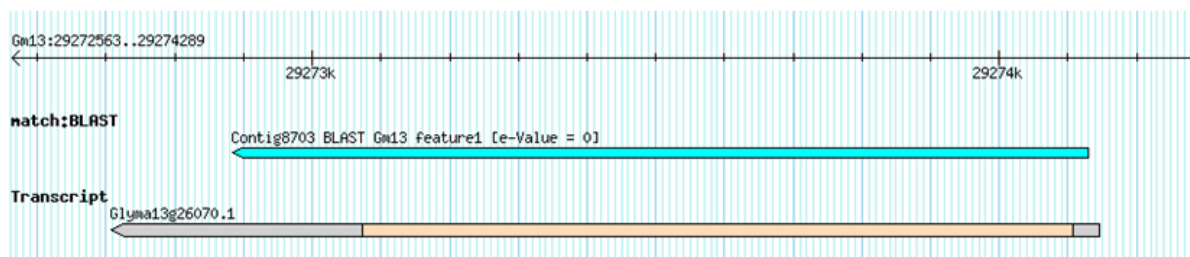
**Perfil de EST**

Gma.6643 - Clone JCVI-FLGm-21M5 unknown mRNA

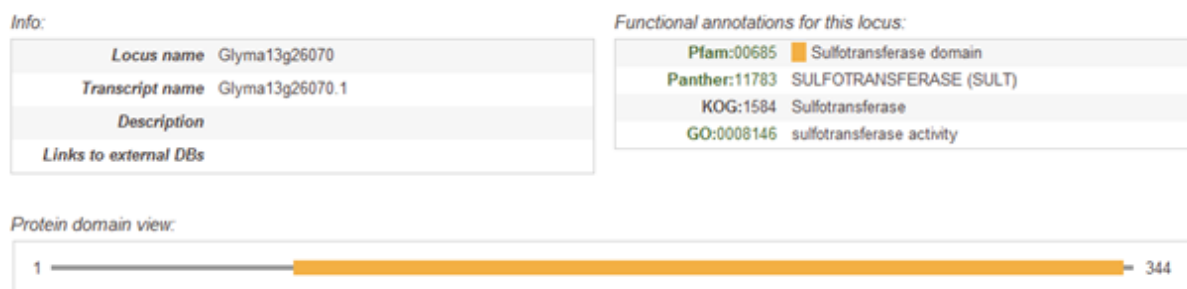
Biblioteca	TPM	Intensidade da mancha	Frequência relativa
Cotilédone	33	●	1/29970
Epicótilo	0		0/2678
Flor	0		0/16526
Hipocótilo	0		0/22531
Folha	0		0/44946
Meristema	0		0/5880
Vagem	0		0/8623
Raiz	288	●	13/44984
Tegumento	0		0/10663
Embrião somático	0		0/13049
Caule	48	●	1/20546
Broto vegetativo	0		0/2631

**Figura 9.** Perfil de expressão do *contig* 8703 baseado na frequência relativa de ESTs (ESTs do *contig* / total de ESTs da biblioteca). (A) frequência das ESTs que formam o *contig* 8703 provenientes de bibliotecas de raiz e não raiz (folha, flor, semente e vagem) de soja no banco montado pela Embrapa Recursos Genéticos e Biotecnologia (<https://alanine.cenargen.embrapa.br/Soja001/>); (B) perfil de expressão do gene Gma 6643 (idêntico ao 8703) originadas das respectivas bibliotecas no UniGene do banco de sequências do NCBI (<http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Gma.6643>), indicando o nome das bibliotecas, o número de transcritos por milhão (TPM); a intensidade estimada da mancha baseada em TPM e a frequência relativa.

Embora não tenha sido encontrada sequência semelhante ao *contig* 8703 no banco GenoSoja, notou-se que ocorre alinhamento com *e-value* igual a zero e 100% de identidade no genoma da soja (Phytozome, <http://www.phytozome.net/>) com a sequência do transcrito Glyma13g26070.1 no cromossomo Gm13 (Figura 10). A anotação funcional do *locus* Glyma13g26070 indica atividade de sulfotransferase (Figura 11), mais especificamente de flavonol-sulfotransferase de acordo com o CDD (*conserved domain database*) do NCBI (dados não mostrados). Diante dos resultados, o gene correspondente ao *contig* 8703 foi nomeado *GmSulfT1* (*Glycine max* Sulfotransferase 1).



**Figura 10.** Análise comparativa da sequência do *contig* 8703 de 1043 pb com o genoma da soja no Phytozome (<http://www.phytozome.net/>). O transcrito Glyma13g26070.1 do cromossomo Gm13 que alinhou com o *contig* 8703 (em azul) com identidade 100% está indicado em amarelo.



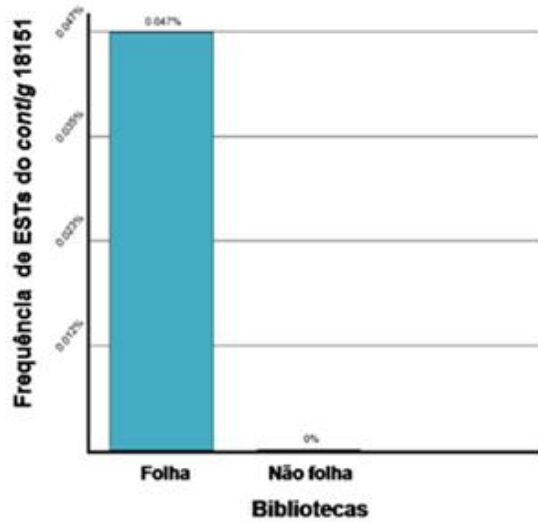
**Figura 11.** Anotação funcional do *locus* Glyma13g26070 correspondente ao *contig* 8703 no banco de dados do Phytozome (<http://www.phytozome.net/>).

O segundo *contig* do qual foi isolada a região promotora é o 18151, cujas ESTs são exclusivamente originárias de bibliotecas de folha, de acordo com o *Northern* eletrônico

realizado no banco da Embrapa Recursos Genéticos e Biotecnologia (Figura 12-A). O gene Gma 12822, correspondente a esse *contig* no NCBI, também apresentou no UniGene o perfil de expressão baseado em EST preferencial em folha e em menor grau em cotilédone (Figura 12-B), bem como o *contig* 24764 (idêntico ao 18151) do banco do projeto GenoSoja apresentou maior expressão em folhas não expandidas e ápices caulinares de plântulas com duas semanas, seguida de tecido foliar senescente de plantas maduras e folhas totalmente expandidas (Figura 12-C).



A



B

Perfil de EST

Gma.12822 - LOC100306224: Hypothetical protein LOC100306224

Biblioteca	TPM	Intensidade da mancha	Frequência relativa
Cotilédone	0		0/29971
Epicótilo	0		0/2677
Flor	0		0/16526
Hipocótilo	44		1/22534
Folha	311		14/44946
Meristema	0		0/5880
Vagem	0		0/8621
Raiz	0		0/44991
Tegumento	0		0/10663
Embrião somático	0		0/13049
Caule	0		0/20546
Broto vegetativo	0		0/2631

C

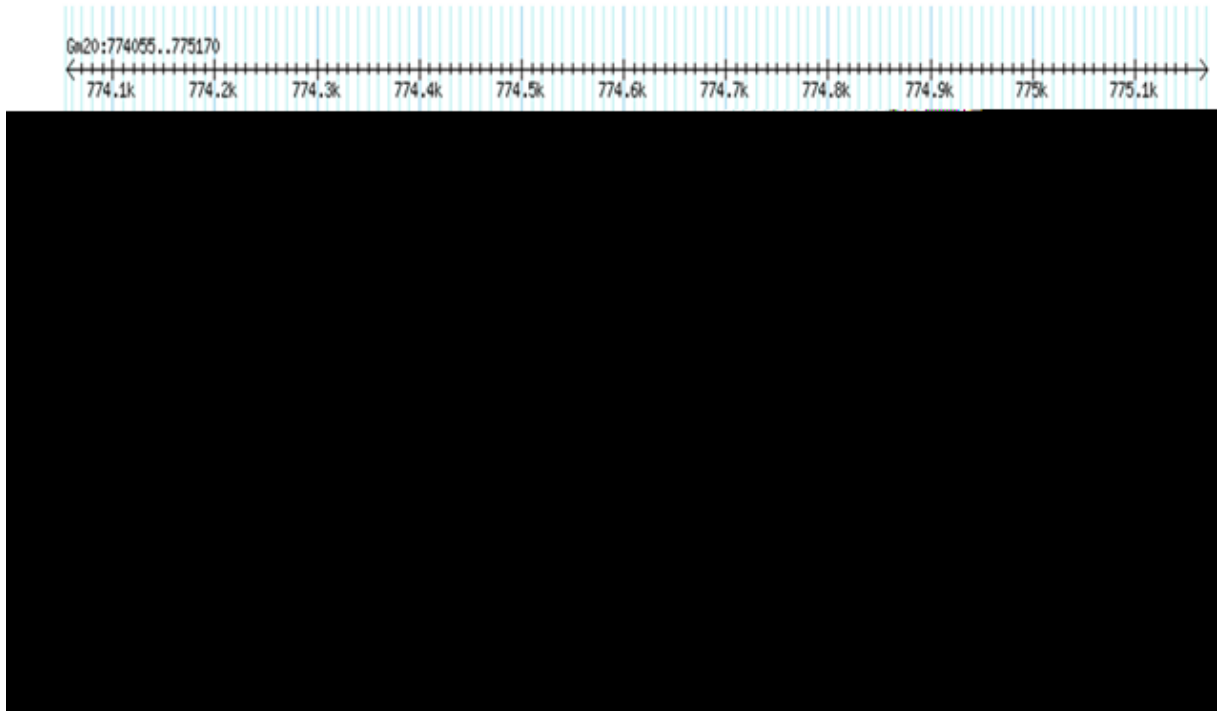
Electronic Northern - Contigs analysis									
Contigs	Expression HO3	Expression LOS	Expression LO6	Expression SH2	Expression LO3	Expression UK1	Expression S10	Expression LO1	Expression LO8
Contig24764	6.05044	24.0713	7.66693	3.59759	20.5843	2.34901	7.26728	22.5004	5.91273

**Figura 12.** Perfil de expressão do *contig* 18151 baseado na frequência relativa de ESTs (*expressed sequence tags*) (ESTs do *contig* / total de ESTs da biblioteca). (A) frequência das ESTs que compõem o *contig* 18151 de bibliotecas de folha e não folha (raiz, flor, semente e vagem) de soja no banco da Embrapa Recursos Genéticos e Biotecnologia (<https://alanine.cenargen.embrapa.br/Soja001/>); (B) frequência de ESTs do gene Gma 12822

(idêntico ao 18151) no UniGene do banco de sequências do NCBI (<http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Gma.12822>), indicando o nome das bibliotecas, o número de transcritos por milhão (TPM); a intensidade estimada da mancha baseada em TPM e a frequência relativa; (C) frequência relativa de ESTs que formam o *contig* 24764 (idêntico ao 18151) no banco de sequências do projeto GenoSoja (<http://bioinfo03.ibi.unicamp.br/soja/>) das bibliotecas: H03 de hipocótilo e plúmula de sementes germinantes; L05 de folhas não expandidas e ápices caulinares de plântulas com duas semanas; L06 de tecido foliar submetido a estresse hídrico; SH2 de brotos germinativos; L03 de folhas totalmente expandidas; UK1 de origem desconhecida; S10 de plântulas; L01 de tecido foliar senescente de plantas maduras e L08 de folha.

No genoma da soja, dois transcritos alinharam, utilizando o BLAST, com o *contig* 18151 com *e-value* igual a zero: um gene correspondente a este transcrito foi identificado no cromossomo Gm20 e outro no cromossomo Gm07 (Figura 13). A ancoragem desses dois transcritos em seus respectivos cromossomos é mostrada na Figura 13. Somente o Glyma20g01120.1 no cromossomo Gm20 apresentou o alinhamento com 100% de identidade ao *contig* 18151 e, por isso, sua sequência gênica foi selecionada para análises posteriores. A anotação funcional do *locus* Glyma20g01120 indica que o *contig* 18151 é membro da família PetM da subunidade 7 do complexo citocromo *b6f* (Figura 14-A). O *locus* Glyma07g21150 também apresentou a mesma anotação funcional (Figura 14-B). A partir dos resultados convencionou-se chamar o gene correspondente ao *contig* 18151 de *GmCit1* (*Glycine max* Citocromo1).

A



**Figura 13.** Análise comparativa da sequência do *contig* 18151, 802 pb, com o genoma da soja no Phytozome (<http://www.phytozome.net/>). (A) transcrito Glyma20g01120.1 no cromossomo Gm20 que alinhou com o *contig* 18151 com 100% de identidade; (B) transcrito Glyma07g21150.1 com 91,1% de identidade de sequência com o *contig* 18151 do cromossomo Gm07. O *contig* 18151 está representado em azul e os transcritos ao qual foi alinhado em amarelo.

A

Info: Locus name Glyma20g01120

Functional annotations for this locus:

Pfam:08041 PetM family of cytochrome b6f complex subunit 7

CO:000642 cytochrome b6f complex

The main content of the screenshot is obscured by a large black redaction box.

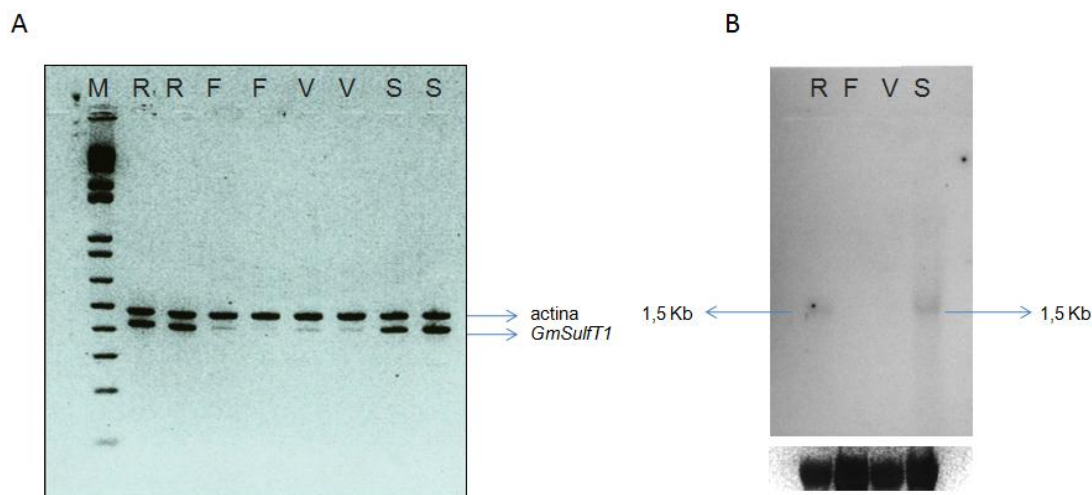
B

The content of this panel is completely obscured by a large black redaction box.

**Figura 14.** Anotação funcional dos *loci* que alinham com o *contig* 18151 no banco de dados do Phytozome (<http://www.phytozome.net/>). (A) Glyma20g01120 no cromossomo Gm20 e (B) Glyma07g21150 no cromossomo Gm07.

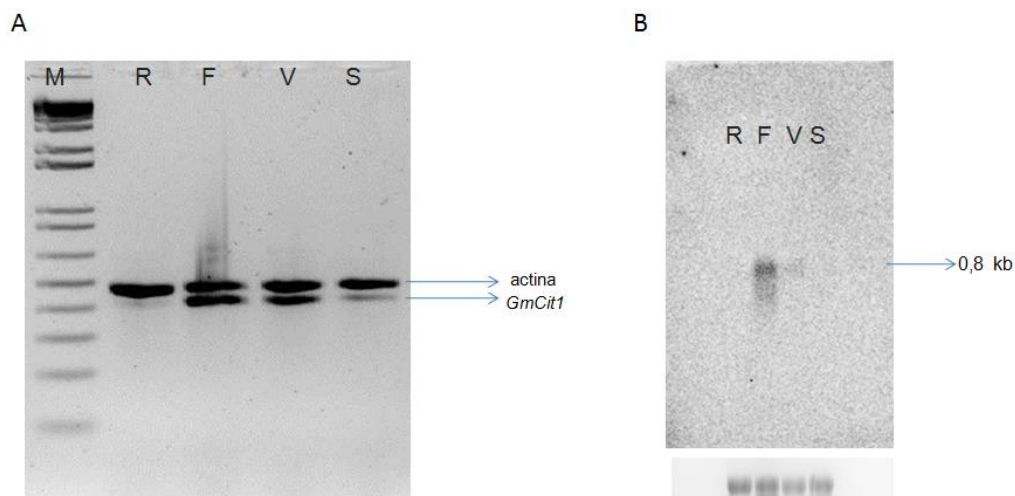
#### 4.2. Análise da expressão dos genes *GmSulft1* e *GmCit1*

Para avaliar a expressão de *GmSulft1* e *GmCit1* em órgãos de *G. max* cv. Conquista foram feitas reações de RT-PCR semiquantitativa e *Northern blot*. Foi amplificado um fragmento de 423 pb, correspondente à sequência predita codificadora de *GmSulft1*, usando pares de iniciadores desenhados dentro da sequência codificadora. Este fragmento amplificou preferencialmente em raiz e semente de soja por RT-PCR semiquantitativa (Figura 15-A) em níveis semelhantes ao do fragmento do gene de actina, gene constitutivo usado como controle. No ensaio de *Northern blot* a sonda marcada com  $^{32}\text{P}$  do fragmento de 423pb do gene *GmSulft1* hibridizou com um transcrito de aproximadamente 1,5 kb em raiz e semente (Figura 15-B).



**Figura 15.** Perfil de expressão do gene *GmSulfT1*. (A) Eletroforese em gel de agarose 1,5% dos produtos das reações de RT-PCR semiquantitativa feitas em duplicatas de reação, indicando os fragmentos amplificados do gene de actina (~ 500 pb) e *GmSulfT1*(423 pb); (B) Ensaio de *Northern blot* do *GmSulfT1* com RNA total de órgãos de soja, apresentando fragmento de 1,5 Kb correspondente ao tamanho estimado do transcrito *GmSulfT1*. No painel inferior, eletroforese em gel de agarose 1,5% mostrando concentrações de RNA ribossomal 25S equivalentes nas amostras correspondentes, após coloração com brometo de etídio. Em A e B: raiz (R); folha (F); vagem (V) ou semente (S). M: 1Kb *plus DNA LADDER*.

No caso de *GmCit1*, ocorreu maior amplificação do fragmento predito de 419 pb em folha e em vagem de soja (Figura 16- A) e a sonda do fragmento do *GmCit1* marcada com  $^{32}\text{P}$  hibridizou no *Northern blot* com um transcrito de aproximadamente 800 pb em folha e vagem (Figura 16-B).



**Figura 16.** Perfil de expressão do gene *GmCit1*. (A) Eletroforese em gel de agarose 1,5% dos produtos das reações de RT-PCR semiquantitativa, indicando os fragmentos amplificados do

gene de actina (~ 500 pb) e *GmCit1* (419 pb); (B): ensaio de *Northern blot* do *GmCit1* com RNA total de órgãos de soja, apresentando fragmento de 0,8 Kb correspondente ao tamanho estimado do transcrito *GmCit1*. No painel inferior, eletroforese em gel de agarose 1,5% mostrando concentrações equivalentes de RNA ribossomal 25S nas amostras correspondentes, após coloração com brometo de etídio. Em A e B: raiz (R); folha (F); vagem (V) ou semente (S). M: 1Kb *plus DNA LADDER*.

### **4.3. Isolamento dos promotores dos genes *GmSulft1* e *GmCit1***

Uma vez que o perfil de expressão preferencial em diferentes tecidos foi confirmado para os dois genes, foi montada uma estratégia para o isolamento das respectivas regiões promotoras. Para tanto, foram gerados fragmentos de diversos tamanhos a partir da região 5' UTR adjacente ao ATG (sítio de início da tradução) dos dois genes. Foram desenhados iniciadores, com base na sequência da região promotora de cada gene obtida a partir da sequência genômica no Phytozome. Nas Figuras 17 e 18 são mostradas as regiões onde os iniciadores anelam-se (sublinhados) para amplificar fragmentos promotores dos genes *GmSulft1* e *GmCit1*, respectivamente. Dessa maneira, para amplificar os fragmentos foi utilizado o mesmo iniciador anti-senso e diferentes iniciadores senso.

CTTTGAACCTGATTATTACCTGGGTGCTGAAAAAATATTTTTCATATAAATTCTCTAATTTTCTACTTTTCTCATTCTGATCAGCAGATAGAGACATTTTCCCA  
 GCTATTTAAAAACATTATTTTTTTTATTCTCTGTGTTAAATTTTTTCTTAATTTTCTTCTCTTTTATTAATTTGTTTCTTTTCTGTTTTTCCCTCGATAT  
 TTCACGTTGAATCCAAATAAATATATACAGGATCGCTAAGGTATTATCCAAGGTAGGTTGTTCAAAAATAAATAAGGCAAGAAGCAGCCTAACAGAAAAAAAAG  
 TAAAACATAAGAAAATTTGTAATAAAAAATGTTATATGTTAGAGATCCGTAGTTCATAGTTGGTAAGGCATATCTTTCATATCGTGGGAGAGATCTTTTATTAAT  
 TTAATATGTTTTCAGTCTTACTAAGAATGTTGATATTTATATACATATATTTTTTAATAATAATTTCAATTATCATTTTACGTATAACGTGATATGCCTCTAAACATGC  
 CATATTATACATATCAGAGTAATTTATGATCATTTTATATATCTAAATGAATGTTATATTTACTAAGCTAATGGAAAAAAAACACTGTGCTGCATTTGACAA  
 AAGAGAAAGATGTACGTGAGAGATTATTAGTTATTTATATAGTTAAAAAGTTAATATACGTGGTAAAAAAAAGATAAATATGTTAGAGATGTGATTGACTAAA  
 AAAAAATAATTAATAATAGTGATGGGTACGTGCTGAATAACAGTTAAAGGATAAATCTTAAAAAACAAAAGTGGAAACCAATGTGTGCCCTCTATTATATA  
 TATGTACGGCAATGCAGGATTATGTTTTTTTAAATATATAAATTTATACACACAAAATATGTATAAACTATACAAGAAGTATATATATATATATATATATAT  
 ATATATATATATATATATATAATGTTTTATACTCTTTTATTACTTGTAGCTCTCCAATTTCTTTTTCTTTCCATCTCTTTGTTCTTTACTAACGACTATAATT  
 TTTATTACATTTCACTCTTTCTTCTCTCTCTTTCTAACTCTTTTATTTTTCTTATAAATCTTTCCAACATATTTTACTAAAGATAAATACATCTCCTCAAAATCTAT  
 ACAACCATCATTTTTTATATTTTGTAAAGATCTCTCCATTTTTTCTAAATATTTTATAATAAATAAAGAAAAATTTATTTTCAATTTGTGATTTTTTAAACAGGAAA  
 TGTATTGTTCTTATTACTCATATTAATAATTAATAATTTACTTAAACAAGATATTAGGATTGGAGTTGCTTACTTGAATTTTAAATTTGAATTTATTTCTCTATTACA  
 TTATAGTCTTGTATTTTTTTACTTAATTTATATAATGTTGTTTTTCATGATTTGTAGCTAATATATATCTTTTAGTTATTAATCAAATTTGAATATATAATTTTAAAT  
 TATAATTTAAATTTTAAATTTTATATTCGCCTCTTAAAAAGGTTTAGTTCTTCTTAGTCGATGATGATTTCTTTTAAATTTTCACAATGAATAAATTTGTGCGC  
 TCGAAACAGAAAGAAAGCTTTGAGTCCATCGCTTTTCATCCGGTGTGATCTGTGCGGTAGCATGATTCCTGCATGTTTATATATGTTGATCCAAATTAATTTATCA  
 TATTCTCATAAAAAATATCGCATTCTTTACTTCTACTTGTAAACAAAAATAGTGTATATAAATAACATATCTTTGTGATCAAATTTCTATGCCAATCAGTACTAGA  
 GATCGACCATTAGCACTCTAATGCTCCAAATAATGTCACATGCTTCAGAGAAGAAAATGAATCCGAGAAAGGGAGGAAAATAACAATAGAGAAGACAAGCTAAGT  
 RzGm1Pr R1  
 CAAGAATGTAAGGAGTTGATACTCTCTCTTCCCTAGGAGAGAGGTTGGAGAACACGTTATATATATCTATTTCAAGGATTTTGGTCCAGCCATTGGAAATCCAAGCA  
 ATAATCACTTTTCAGAAGCACTTCCAAGCTAAAGACAGTGATGTTATTGTGGCCCAAATTCAAAAATCAGGTACCACTTGGCTGAAAGCTCTCACCTTTGCCATTGTC  
 AATCGCCATACTCATAGTATCACTACATCAATGTCATCACATCCTTTGCTTACTTCTAATCCTCATGAACTTGTGCCCTTTCATAGAAATACACCCTTTATGGTAATGCC  
 CCTAGCCATGTTCCAAACCTATCCAACATGACTGAGCCAAAGACTTTTTTGGTACACATATTTCCATTTCCATGCAATGGCCAACTCAATCAAGGAGTCCAATAGTAGAATA  
 ATTTATATATGTAAGAACCACTTGACACTTTTTGTGTCCTACTTGGATTTTCTCAACAAAATTAAGCCAGAACATTTACCTGAATTTGAACTAGGGGAAAGCTTTTGAA  
 AAGTATTGCAAAAGAAATAAGGTTTGGTCCAACTTGGGACCAAATGTTGGGTTATTGGAAGGAGAGTATAGCTAGGCCTAGTAAGTTTGTCTTGAAGTACGAG  
 GATCTTAAAAAAGATGTCAAATTTTCATGTGAAAGAATAGCGGAGTTCTTAGGATGGCCTTTCACCTCGGAGGAAGAAGGTGATGGGACTATGAGAGCATAAACAAG  
 CTATGCACTTCCGAGAAGATGAAGGAATGGAGCCAAATAAATCTGGAACATTTGCTAGGAACCTTGGAGAAAGTACTTGTCCGAAAGGCTGAAATGGGAGATTGG  
 GTGAACTACCTTTCCCTGAAATGGGTGAAAAGTTATCGCAAATTTATGGAAGAAAGTAAAGTGGGTGAGGCTGTCATTTTAAAGTGTCTCCTTATTTTTGTGATCG  
 ATCACTAGCTAGTGATCATCATCTACATCAAGTGTCTAAATAAGAATCGTGATATATGTCAGGTCATGATGTGGCCTACAAGTGCATGTTGTTGTGATTTTCCCTTT  
 TCCCGTTTGGATGATTTTTCTACTAGTGCAATGAAGTCTCAGTAAAGTTATTTCTATCATCGTGTTTTGTGTTAATTAATTAAGTACTCATTGCTTGTGTTTG  
 TGGTATTATATATATATATGACAGATAAACTTTGGTCTTGTATTTTGGATATCCTTGGTGTCCGGGTGATGAAGATTAAGGTTTGGTGAAGATTAAGTATCGG  
 TCGCGGTAACTGTTGATG

**Figura 17.** Sequência genômica da região promotora e codificadora de *GmSulft1* de *G. max*. cv. Williams 82 obtida pelo Phytozome (<http://www.phytozome.net/>). O sítio de iniciação da tradução (ATG) está destacado em vermelho, verde corresponde à região 5' UTR, azul à sequência codificadora e rosa à região 3'UTR. As regiões onde os iniciadores se anelam para amplificação dos diversos fragmentos do promotor estão sublinhadas.

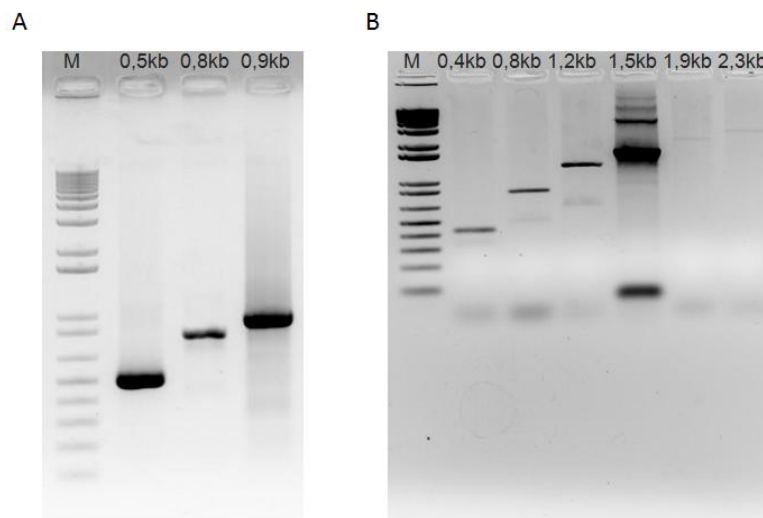
CTAAACTCTTGAACCTCACTTGCAGCATCGCGTCAGTACTACCAATTGATTAGAAATTAGCAGGAGATTTAATTTATGATAGTAAAATTGTCGCCATTTAGCCCT  
GGATTCTGAGTGATCTGTTGTTCACTAATTTCTTCACTTGGAACTTTGATGCTTTACATGTTTCATAAACATTGGAACCTTAGTTATTTGATGTTTTGATACAAATTTCA  
TGTTTTTGATTTGGGTTGTAATAGATTGTTAAATCTAGTTGATCATAAACTGAGTCGAAATGCTTGTGCTATTGCCGTGTTTTGGTTAATATTTTTTGTATGAAG  
TTATATTTGTTGTAGCTCCACCTGCATGAGTATGATGGTTTCTACTGCGTGTCTTACTGCTGATTTTTATTTAATCTGATCTGACATTTAATCTATATAG  
GATTAGGATTAGGATTTTGGGAGAATTGTAATATGATATGGATATTATGAGGTAACATGGATGATAGCTTTCTTCATATTTGTTGAGGAGCATGTTTTTTTTCTTTTC  
FoGm1P 1.9 F1  
TCAACTGAAATTATTTTAGTTGACAGCAAATACAGAGAATTGAAATTCGGCTGTTTGTATCTCTCTCTTTTGGCTCTAAAAATGTTTTCTAGATTTTGGG  
GTTATATTTATATATATATATAGTTATAGTTATTTTTAACAGAACTTCAGACAGCAATTAAGCACCTTCTCTTGTAGTTGTAGGCCAACGTGCAATTAACATGCACT  
TTAGACAGCAGCAATTCAGCACTTCTTTTACCATATTTCTACAAAAGTTTGTCTGGCTCTAATAGTTGATGAAATATTGAGCTTCTATTTTTTTTTATGGTGGAA  
AAGAGAAATGTGCTCAATAGAATAGAATTTGATTAATGGAAGGTAGGTTGGTGAATATGGTTTTTTTGGCTTATGCTGTGGTGTAGACATTTTTTGTCTAATGGGCC  
CTGTTGTTCCACTCGCATAATTTAGTGTCTTTGTTGGCAAATGATCCTCCTAATTTTCATTTTTCTTAATTTTGTGCAAGAAAAAGTGGTTCATCCTTATGTTT  
CTTGTTCCTGTTTTTAAATTAATTTACTATTATTTATAAAATTAATAATAGTCGTAACAGTGGCTAGTTGAGACAACAAATTTACATTTACACGTTATTTTTCTACTCAT  
GATCAAGATTAAGAGACGTTTCAATGGAGTTTTATTTTTATGATTTACTTTTTGTTTATCCTTTTTATTAATTAACCTTGTAAATTTGGGTAAATGATGATACCTCCATCATG  
TGGTCTTGTGCTGAATTCATCTCATTTTAATAAAAAATTTAGCTTATAAAAAATAGTTTAAAAATAATAACAAATTTAAATGTTTTATAAGCAAGTTGATTAAGA  
TTAAGTATTATAATTTTTTATAAGAAAATGATTTAACTTAACATAGTTTATCTTAAATGAATGGATCATGATTTAGTCAAACCTCGATCCAAATTAACCTTGTGAAC  
ACTTGTAGACAAGCTAATCAAGTAGTTTCACTATTATAGGACTTCTAGATTTTATGGCGGTTAATAACTTTTTGATTTGATTTCCATGAAATTTTTATTACGAAT  
FoGm1P 0.8 F1  
GAAAGAAATATAAGATGATTTCTCTTAAAAAAGTTTCAATTAAGATTAATAATATTTAGGCTAATCTTTTTGATTTATATATGAAATATGTTTGTAAATTTTTTATA  
AAAAAATAAAACCTAGCAATGAAATTAAGCGTGTCTTTGGATTAACCTAATAAAAAATTAATAATTTCTCAAAAATTAAGATAACCAACACCCCAATGCCAT  
ATTCCAATTTTTGGACTCTAAATTTAGTTGAACCTATTTACTTTCATTTGAGACGAGGTGGAACCTCTCAAAGCCAAGAGGAAGTAAATAGCGGAATAAGGTTCATAAA  
AAAGAGTTAACTCAACTAACAAAATGAATGATGGGTTTTCCAATGTTTTGTTGGTCAGCATGTATGTAATTTGTACACCACAAAATCAATAAGCTATCACATTTT  
FoGm1P 0.4 F1  
CTTATCAACCAACCACCTCTTCATACCTTATCCTTTTCGTTCTGTGGATCCAAATACAAAGTCCCTTCCCACACCCTGTAACATCTACTTTACAGAAGCACACCACAG  
GAACTAGAAAATCCATACAATCTTACCACCAAAAAGGTTAATGTAAGTACTATAACATAACTCTCTCTAGCCTATCATGATTACACATTTGTTTCACTATTTCTTGT  
ATATATTTGTTCAAGCAAACCTTCAACAGCATCAAGTGTGCTAAATCTAACTAACTGACCCTTTGTGTTGGTAATGTTATCTCATTTTACAGAAAATAATGCAATGGCC  
FoGm1Pr R1  
AGTGCAACAACTCTAGCCACAATCACTGTTGTGGCGGCTAGCCCCAGTGTGGCAGCAGAAGGAGTGGGAAGAGGAAATGTGAACCTTCATCAGAGGGTTGAATCTTTT  
GGAGGGTTGAAGGCTCAGAACAATGTGACATTACTAGGCCCTTCTGTGTGCACTGAGCAGTCTTTGCAAGGTTGGTGTGCTCATGAAATATCCATCATCATCA  
CGTAAGGGGAGAGGTGGAGGTGCTGCCTTTTCAACATGCAATGCTGCTGGTGTGAGATTTCCAGATTGCAGCCATCATGAATGGCCTTGTGCTTGTGGGGTGGCAATA  
GGGTTTGTCTTCTTCGAGTCGAAGCGTTTGTGGAAGAGTCTGAGTGAGAAATGTCATCAACTTTTAAATCAGTTTTTTCATATATGATTTTTTATAAAGAAAGAAAAAAG  
CGAAAAGTTGTATCTTCTTAATTTCTGTATGATGTAGAACTACATTTTTTTGGGGTGTCTGAATAGTACCATGTGAATTTTTTATTTGGGGAGAAAAATTCACACA  
CAAAGTGATCTGGCTTGTGACATGAAAATTTGGTGGATCTATGTTAATAATGAAGTGGTATTGGTATGAAAACGAACAT

**Figura 18.** Sequência genômica da região promotora e codificadora de *GmCit1* de *G. max*. cv. Williams 82 obtida pelo Phytozome (<http://www.phytozome.net/>). O sítio de iniciação da tradução (ATG) está destacado em vermelho, verde corresponde à região 5' UTR, azul à sequência codificadora, rosa à região 3'UTR e a região em amarelo corresponde a um íntron. As regiões onde se anelam os iniciadores para amplificação dos diversos fragmentos do promotor estão sublinhadas.

Foram isolados da região promotora do gene *GmSulfT1* três fragmentos, com aproximadamente 0,5 kb, 0,8 kb e 0,9 kb (Figura 19- A) nomeados, respectivamente, de PSulfT0,5; PSulfT0,8 e PSulfT0,9. Da região promotora do gene *GmCit1* foram isolados seis fragmentos com tamanhos aproximados de: 0,4 kb; 0,8 kb; 1,2 kb; 1,5 kb; 1,9 kb e 2,3 kb.



(Figura 19- B) que foram denominados PCit0,4; PCit0,8; PCit1,2; PCit1,5; PCit1,9 e PCit2,3, respectivamente.

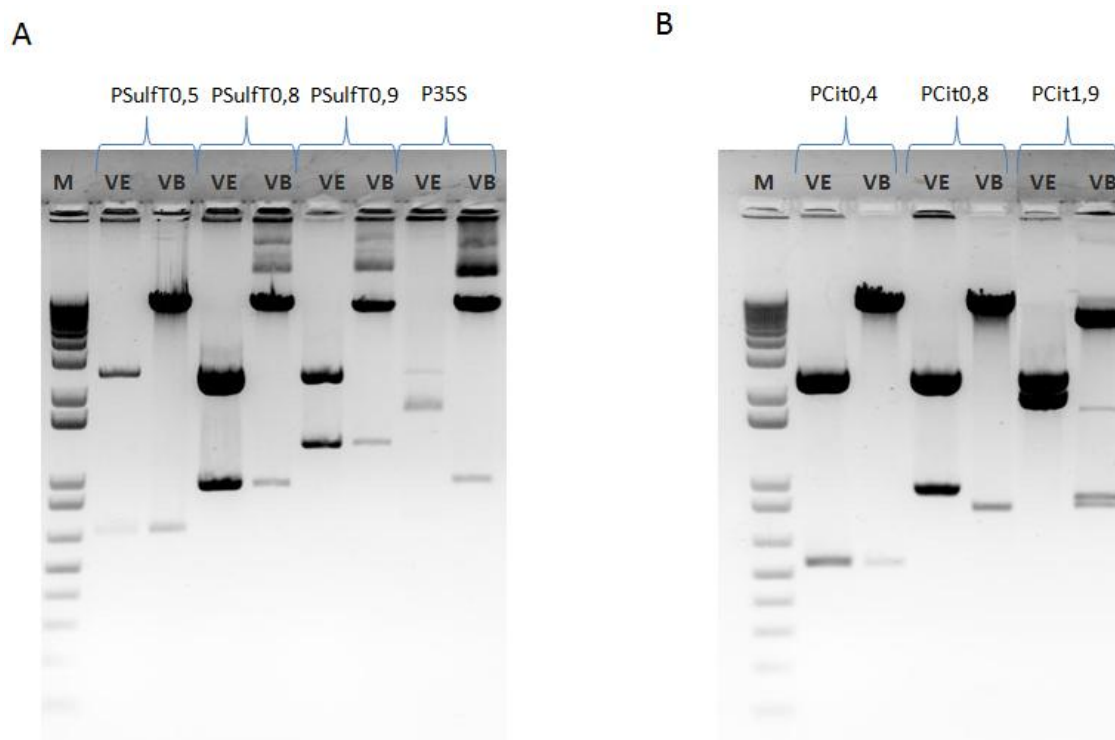


**Figura 19.** Eletroforese em gel de agarose 1,5% dos fragmentos amplificados da região promotora dos genes *GmSulfT1* e *GmCit1*. (A) fragmentos de 0,5 kb; 0,8 kb e 0,9 kb da região promotora de *GmSulfT1*; (B) fragmentos de 0,4 kb; 0,8 kb; 1,2 kb; 1,5 kb; 1,9 kb e 2,3 kb da região promotora de *GmCit1*. M: 1Kb plus DNA LADDER.

Os fragmentos da região promotora de cada gene, amplificados por PCR, foram inseridos no vetor de entrada pENTR™. Após a confirmação da clonagem pela digestão com enzima de restrição foi feita a reação de recombinação do tipo LR do sistema Gateway® entre esses vetores e o vetor binário pMDC162. O resultado da recombinação são os novos vetores binários contendo os promotores PSulfT0,5, PSulfT0,8, PSulfT0,9, PCit0,4, PCit0,8, PCit1,5, PCit1,9 e o P35S posicionados a montante do gene repórter *gus*, que codifica a enzima - glucuronidase.

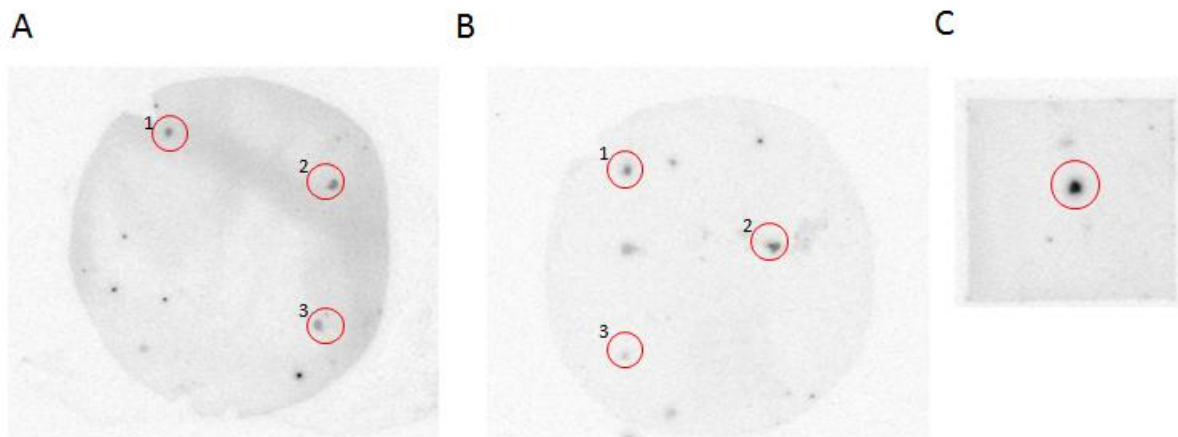
As etapas de clonagem do sistema Gateway® foram verificadas por eletroforese em gel de agarose 1% contendo os produtos das digestões com enzimas de restrição. Após digestão com *NotI* e *EcoRV*, dos vetores pENTR™ contendo os diferentes fragmentos da região promotora do gene *GmSulfT1* observa-se os fragmentos de aproximadamente 0,5 kb, 0,8 kb e 0,9 kb correspondendo aos respectivos promotores (Figura 20-A). Somente os vetores pENTR™ contendo os promotores PCit0,4, PCit0,8 e PCit1,9 liberaram os fragmentos de aproximadamente 0,4 kb; 0,8 kb; e 1,9 kb (Figura 20-B), como previsto para a digestão dos vetores com *NotI* e *EcoRV*. A restrição dos vetores binários, resultantes da reação de recombinação LR, com a enzima *XbaI* gerou fragmentos correspondentes aos promotores

PSulfT0,5; PSulfT0,8 e PSulfT0,9 (Figura 20-A), e aos promotores PCit0,4, PCit0,8, (Figura 20-B). O promotor PCit1,9 possui um sítio de reconhecimento da enzima utilizada na posição de 1134 pb. Dessa forma, além do fragmento com aproximados 1,9 kb foram liberados ainda dois fragmentos com aproximadamente 0,9 kb cada (Figura 20-B) da digestão do vetor binário com PCit1,9. O promotor P35S possui um sítio de reconhecimento da *EcoRV* na posição 739 pb. Por isso, a digestão do vetor pENTR™ contendo o P35S com as enzimas *NotI* e *EcoRV* liberou fragmentos de aproximadamente 2 e 0,7 kb (Figura 20-A). A liberação de um fragmento de aproximadamente 0,2 kb também era previsto, porém, sua visualização não foi possível após eletroforese em gel de agarose nas condições descritas na seção anterior. O fragmento liberado com 0,8 kb do vetor binário contendo o P35S após digestão com *XbaI* está de acordo com o previsto (Figura 20-A).



**Figura 20.** Eletroforese em gel de agarose 1% dos fragmentos de restrição gerados do vetor de entrada pENTR™ (VE) com *EcoRV* e *NotI* e dos vetores binários pMDC162 (VB) com *XbaI*. (A) digestão dos vetores contendo os promotores PSulfT0,5, PSulfT0,8, PSulfT0,9 e P35S; (B) digestão dos vetores contendo os promotores PCit0,4, PCit0,8 pMCit1,9.

Foram obtidas colônias de *Agrobacterium tumefaciens* contendo os vetores binários onde foram clonados os promotores PSulfT0,5; PSulfT0,8, PSulfT0,9, PCit0,4, PCit0,8, PCit1,9 e P35S as quais foram identificadas por hibridização de colônia (Figura 21).



**Figura 21.** Hibridização de colônias de *Agrobacterium tumefaciens* transformadas com os novos vetores binários: (A) colônias hibridizadas com o promotor PSulfT0,5. Na colônia 1 o vetor binário possui o promotor PSulfT0,5, na 2 o PSulfT0,8 e na 3 o PSulfT0,9, (B) colônias hibridizadas com o promotor PCit0,4. Na colônia 1 o vetor binário contém o promotor PCit0,4, na 2 o PCit0,8 e na 4 o PCit1,9 (C) colônia hibridizada com o promotor 35S de 0,8 kb. Todas as sondas foram marcadas com  $^{32}\text{P}$ .

#### 4.4. Análises *in silico* e funcional dos promotores isolados

Plantas transformadas com os vetores binários contendo os promotores isolados dos genes *GmSulfT1* e *GmCit1* clonados a montante do gene repórter *gus* foram regeneradas (Tabela 3). Essas plantas foram analisadas individualmente quanto à atividade da GUS em folha e raiz (Figuras 22-B e 23-B e Tabela 3). Devido à característica do ensaio histoquímico de ser uma análise qualitativa, foi feita uma comparação visual entre as intensidades da cor azul resultante da reação da GUS com o substrato X-Gluc. Foram obtidas cinco plantas transformadas com o PSulfT0,5 que apresentaram atividade da GUS tanto em raiz quanto em folha com variação de intensidade da cor azul (Tabela 3).

Quanto à região promotora do gene *GmCit1*, foram obtidas quatro plantas transformadas com o promotor PCit0,4. Destas, duas não apresentaram atividade da GUS em raiz e duas apresentaram pouca atividade, enquanto em folha as quatro plantas apresentaram atividade mediana. PCit0,8 apresentou alta atividade GUS em folha das cinco plantas transformadas e em raiz essa atividade variou (Tabela 3). No caso das duas plantas transformadas com PCit1,9, foi verificada expressão forte de *gus* tanto em folha quanto em

raiz (Tabela 3). Todas as plantas transformadas com o vetor binário sem promotor adjacente ao *gus* e as plantas não transformadas não apresentaram atividade da GUS (Tabela 3).

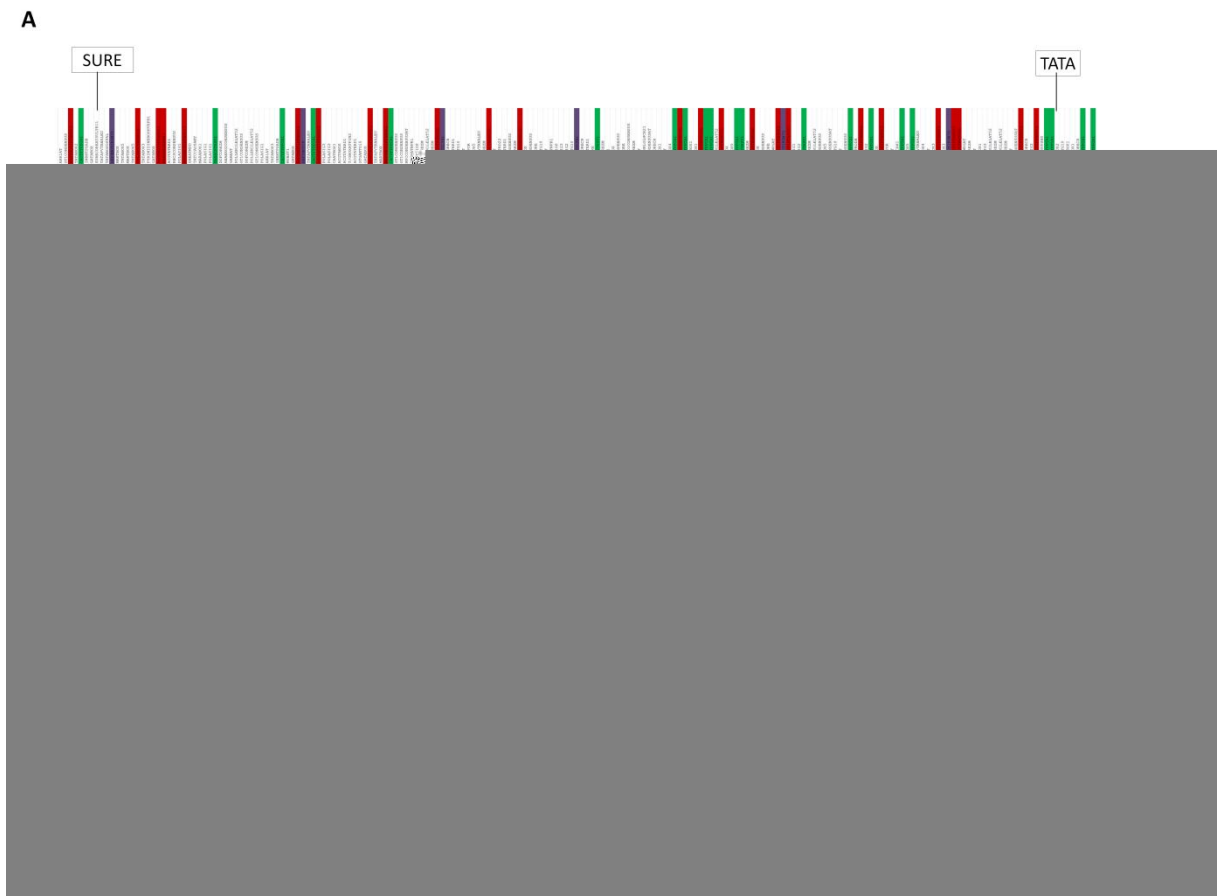
**Tabela 3.** Níveis de atividade da GUS encontrada nas plantas de *Nicotiana tabacum* transformadas com os promotores PSulfT0,5, PCit0,4, PCit0,8, PCit1,9, vetor binário sem promotor e em plantas não transformadas.

Promotor	Planta#	Folha	Raiz
PSulfT0,5	4	+++	++++
PSulfT0,5	5	+++++	+++
PSulfT0,5	14	+++	+++
PSulfT0,5	15	++++	++
PSulfT0,5	17	+++++	++++
PCit0,4	15	+++	++
PCit0,4	38	+++	+
PCit0,4	39	++	-
PCit0,4	40	+++	-
PCit0,8	7	+++++	++++
PCit0,8	8	+++++	+++++
PCit0,8	26	+++++	+++
PCit0,8	28	+++++	+++++
PCit0,8	31	+++++	++++
PCit1,9	17	++++	++++
PCit1,9	22	++++	++++
Controle negativo (sem promotor)	14; 16; 21; 31	-	-
Controle negativo (planta não transformada)	3; 4; 5; 10; 20; 26; 33; 34	-	-

Nota: O nível máximo de expressão é representado por +++++ e ausência de expressão é representada por -.

#### 4.4.1. Região promotora de *GmSulfT1*

De acordo com o programa de predição de promotores (TSSP - Softberry, Inc.), a região promotora do gene *GmSulfT1* com 1848 pb apresenta TATA Box putativo a 37 pb a montante do início da transcrição (TSS). O programa para reconhecimento de motivos regulatórios de plantas, o NSITE-PL (Softberry, Inc.) reconheceu na mesma região promotora com 1848 pb do *GmSulfT1* 43 motivos de dez elementos *cis* regulatórios diferentes. Segundo o PLACE, a sequência analisada possui 423 motivos regulatórios putativos nas fitas de DNA senso e anti-senso, sendo 148 relacionados à expressão órgão/tecido-específica de diversos genes. Desses, 31 motivos que regulam a expressão específica em semente, 46 em raiz e 21 em folha foram evidenciados na região promotora do *GmSulfT1*. Na figura 22-A estão representados os motivos putativos dos elementos *cis* encontrados na fita senso. Os elementos *cis* necessários para expressão em raiz foram OSE1ROOTNODULE, OSE2ROOTNODULE, ROOTMOTIFTAPOX1 e RHERPATEXPA7, os envolvidos na expressão gênica em semente são os ACGTOSGLUB1, SEF1MOTIF e SEF4MOTIFGM7S e os envolvidos na expressão gênica em folha foram o CACTFTPPCA1 e TAAAGSTKST1. Além desses elementos, foram detectadas sequências reconhecidas por famílias de fatores de transcrição como Myb, Myc e Dof e elementos responsivos a hormônios e metabólitos. O elemento *cis* SURECOREATSULTR11, responsivo a níveis de enxofre, também foi identificado na região distal dessa região promotora. Na figura 22-A estão destacados, além dos elementos necessários para expressão gênica órgão-específica em raiz, folha e semente, o sítio de início de transcrição (TSS) presentes na fita de DNA senso, o motivo putativo do elemento SURECOREATSULTR11 e os fragmentos dessa região que estão sendo estudados. A Figura 22-B mostra o ensaio histoquímico de plantas de fumo: não transformada, transformada com o vetor binário sem promotor e com o promotor PSulfT0,5 a montante do gene repórter *gus*. Observa-se que o PSulfT0,5 é capaz de ativar a expressão do gene repórter tanto em folha como em raiz.

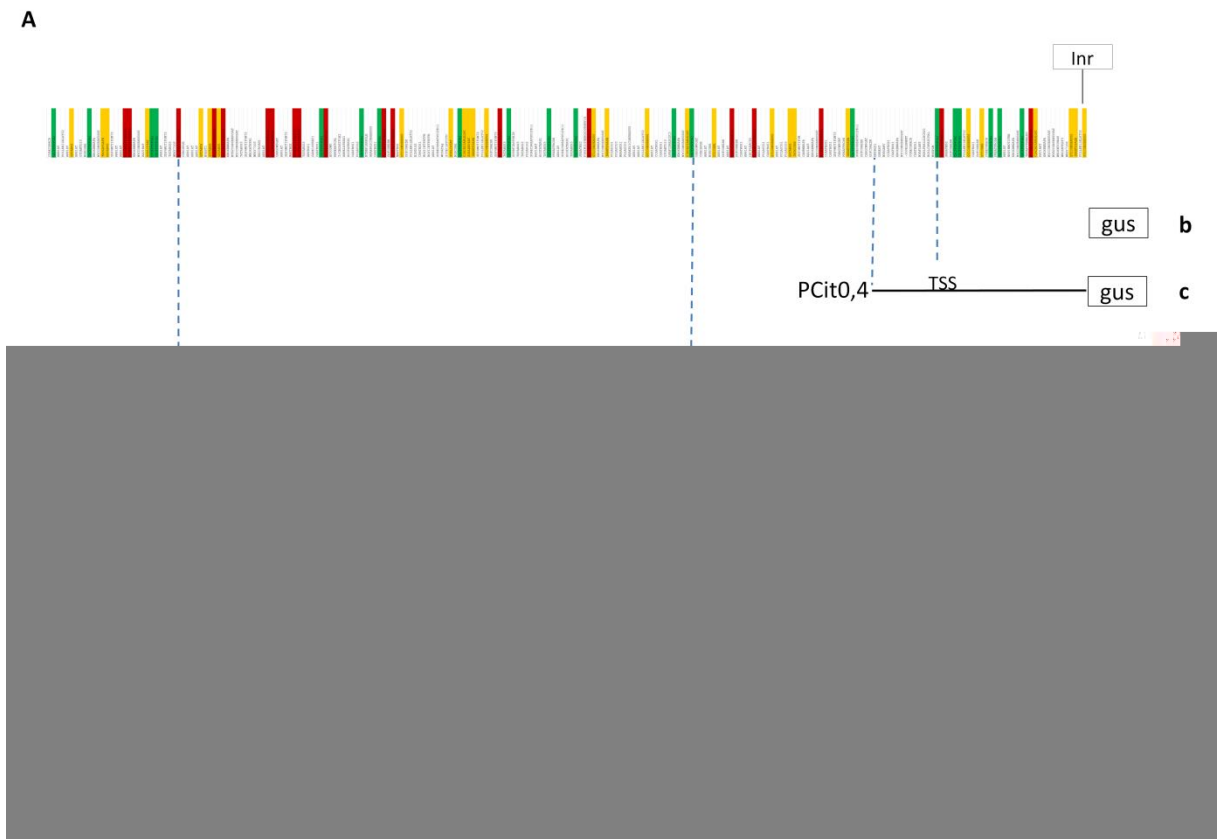


**Figura 22.** Análise *in silico* e funcional da região promotora do *GmSulfT1*. (A) representação dos motivos putativos dos elementos *cis* encontrados na região promotora (fita senso) com 1,8 kb. Na barra superior estão indicados o TATA Box e o elemento responsivo a enxofre Sure (SURECOREATSULTR11). Os elementos destacados necessários para expressão órgão-específica são: OSE1ROOTNODULE, OSE2ROOTNODULE, ROOTMOTIFTAPOX1 e RHERPATEXPA7 responsáveis pela expressão gênica em raiz, CACTFTPPCA1 e TAAAGSTKST1 necessários para expressão em folha e ACGTOSGLUB1, SEF1MOTIF e SEF4MOTIFGM7S relacionados com a expressão semente-específica. Na parte inferior **b** representa o gene *gus* sem promotor, **c**, **d**, e **e** representam o gene *gus* com os promotores PSulfT0,5, PSulfT0,8 e PSulfT0,9 respectivamente e (B) ensaio histoquímico em folha (a esquerda) e raiz (a direita) das plantas de fumo onde: **(a)** planta não transformada, **(b)** planta transformada com o vetor binário sem promotor e **(c)** planta com o vetor binário contendo o promotor PSulfT0,5. A barra na parte inferior das fotos corresponde a 1 mm.

#### 4.4.2. Região promotora de *GmCit1*

A predição do promotor do gene *GmCit1* não identificou nenhuma sequência putativa de TATA Box na sua região promotora com 2,3 kb. O NSITE-PL reconheceu na sequência 18 motivos de 18 elementos regulatórios diferentes. O banco PLACE, no entanto, detectou 477 motivos de elementos regulatórios localizados nas fitas de DNA senso e anti-senso, sendo 146 relacionados à regulação da expressão órgão/tecido-específica. Um motivo putativo do elemento iniciador, indicado na figura 23, foi detectado na região -14 pb do início da tradução. Os elementos *cis* putativos necessários para expressão em raiz encontrados foram OSE1ROOTNODULE, OSE2ROOTNODULE e ROOTMOTIFTAPOX1, o envolvido na expressão gênica em folha foi o CACTFTPPCA1 e também foram encontrados 63 motivos dos elementos responsivos à luz GT1CONSENSUS, GATABOX, INRNTPSADB, IBOXCORE, CIACADIANLELHC, -10PEHVPSBD, GT1CORE, IBOX, IBOXCORENT, SORLIP2AT, TBOXATGAPB, SORLIP1AT, SORLIP4AT e SORLREP4AT (Figura 23- A). Além desses elementos, foram detectadas sequências reconhecidas por famílias de fatores de transcrição como Myb, Myc e Dof e elementos responsivos a hormônios e metabólitos.

Os promotores gerados a partir de deleções 5' da região promotora do *GmCit1* estão representados com os respectivos motivos putativos de elementos *cis* na Figura 23-A. A Figura 23-B mostra o ensaio histoquímico de folha e raiz de plantas de fumo: não transformada, transformada com o vetor binário sem promotor, transformada com o promotor PCit0,4, transformada com o promotor PCit0,8 e transformada com o promotor PCit1,9. Nota-se que o PCit0,4 é capaz de ativar a expressão do gene repórter mais forte em folha do que em raiz e que os promotores PCit0,8 e PCit1,9 são ativos em folha e raiz.



**Figura 23.** Análise *in silico* e funcional da região promotora do *GmCit1*. (A) representação dos motivos putativos dos elementos *cis* encontrados na região promotora (fita senso) com 2,3 kb do *GmCit1*. Na barra superior estão indicados o TATA Box e um motivo putativo do elemento iniciador (Inr). Os elementos destacados necessários para expressão órgão-específica são: OSE1ROOTNODULE, OSE2ROOTNODULE e ROOTMOTIFTAPOX1 responsáveis pela expressão gênica em raiz, CACTFTPPCA1 necessário para expressão em folha e GT1CONSENSUS, GATABOX, INRNTPSADB, IBOXCORE, CIACADIANLELHC, -10PEHVPSBD, GT1CORE, IBOX, IBOXCORENT, SORLIP2AT, TBOXATGAPB, SORLIP1AT, SORLIP4AT e SORLREP4AT responsivos à luz. Na parte inferior **b** representa o gene *gus* sem promotor, **c**, **d**, e **e** representam o gene *gus* com os promotores PCit0,4, PCit0,8 e PCit1,9 respectivamente e (B) ensaio histoquímico em folha (a esquerda) e raiz (a direita) das plantas de fumo onde: (a) planta não transformada, (b) planta transformada com o vetor binário sem promotor e (c) planta com o vetor binário contendo o promotor PCit0,4, (d) planta com o vetor binário contendo o promotor PCit0,8, (e) planta com o vetor binário contendo o promotor PCit1,9. A barra na parte inferior das fotos corresponde a 1 mm.



## **5. Discussão**

A disponibilidade de bancos de ESTs de plantas em domínio público, o baixo custo da técnica e o avanço das ferramentas de bioinformática são vantagens do *Northern* eletrônico para análise do perfil de expressão gênica nos órgãos e tecidos de plantas (Ewing *et al.*, 1999; Murray *et al.*, 2007). No entanto, a falta de informação detalhada disponível sobre a origem dos tecidos e órgãos utilizados para montar os bancos de ESTs, a qualidade e quantidade do RNA utilizados, as taxas de falsos negativos e positivos são fatores que diminuem a confiabilidade dos resultados obtidos nesse tipo de análise (Murray *et al.*, 2007).

Usando essa técnica e visando o isolamento de promotores órgão-específicos, no presente trabalho, foram identificados os genes de soja *GmSulfT1*, preferencialmente expresso em raiz e semente, e *GmCit1*, preferencialmente expresso em folha. Para complementar os dados disponíveis nos bancos públicos, foi analisado o perfil de expressão gênica do gene correspondente aos *contigs* encontrados no banco da Embrapa Recursos Genéticos e Biotecnologia e em outros dois bancos de dados: o UniGene e o banco do GenoSoja. As sequências utilizadas nas análises dos três bancos possuem origens diferentes, bem como a forma de normalização dos dados.

O Teste Exato de Fisher garantiu que o perfil de expressão órgão-específico dos *contigs* resultantes não tenha sido um evento ocorrido ao acaso. Esse teste é mais conservador que outros utilizados, incluindo o de Audic e Claverie, para esse tipo de análise, isso significa que a diferença entre a frequência da EST nos grupos contrastados deve ser maior para que o *contig* seja considerado específico do órgão onde a EST é encontrada com maior frequência (Audic e Claverie, 1997). Partindo do princípio que a frequência da EST reflete a expressão do gene no órgão de origem, o uso desse teste levou à identificação de genes putativos com níveis de expressão altos. Os testes de significância estatística indicam uma tendência da ocorrência de expressão dos *contigs* em análise, no entanto, a validação dessa tendência precisa ser feita por análises biológicas, como realizado neste trabalho.

Os resultados das análises biológicas por RT-PCR semiquantitativa e *Northern* blot foram coerentes entre si e com as análises *in silico* de *GmSulfT1*, apresentando expressão preferencial em raiz e semente. A expressão localizada em semente pode estar relacionada à expressão apontada em cotilédone pelo *Northern* eletrônico. A proteína codificada por *GmSulfT1* possui um domínio de sulfotransferase, mais especificamente da família das flavonol-sulfotransferases. As sulfotransferases são proteínas localizadas no citosol das células vegetais e catalisam a sulfatação de moléculas orgânicas como glucosinolatos,

flavonóides, jasmonatos e brassinoesteróides (Takahashi *et al.*, 2011). Nessa reação a PAPS (3'-fosfoadenosina- 5'- fosfosulfato) é a molécula doadora de  $\text{SO}_4^{2-}$ , que por sua vez é adicionado a um grupo hidroxila de moléculas orgânicas ou a resíduos de tirosina em peptídeos (Leustek e Saito, 1999; Takahashi *et al.*, 2011). A assimilação de enxofre em plantas é feita por meio da sulfatação ou da via redutiva (Leustek e Saito, 1999). As sulfotransferases de plantas catalisam a sulfatação de flavonóides, desulfoglucosinolatos, colinas e ácido gálico glucosídeo (Varin *et al.*, 1997). A função biológica de conjugados sulfatados em plantas é desconhecida, suspeita-se que como em animais, a sulfatação possa estar relacionada à atividade hormonal e pode levar à produção de moléculas biologicamente ativas (Gidda e Varin, 2006).

Segundo Gidda e Varin, (2006) a distribuição da flavonol-sulfotransferase *AtST3a* de *Arabidopsis thaliana* na planta pode estar relacionada com a distribuição e acúmulo dos flavonóides e também é regulada temporalmente, de acordo com o estágio de desenvolvimento da planta, e responde a hormônios. O autor relata que em plantas com 15 dias e tratadas com citocinina, a expressão do gene *AtST3a*, de *Arabidopsis thaliana*, era confinada à raiz. A função e distribuição de flavonóides podem variar entre as espécies de plantas e seu estudo ainda está restrito a um pequeno número de espécies (Gidda e Varin, 2006).

Em plantas de soja a distribuição de flavonóides varia de acordo com o órgão, idade e com o tempo de exposição à luz (Graham, 1991). Graham (1991) detectou a presença de flavonóides somente nas folhas primárias de plântulas de soja. Sabe-se que flavonóides excretados pela raiz e tegumento de sementes germinantes de soja agem como indutores de genes da nodulação em rizóbios e desempenham um papel importante no transporte de auxina na raiz durante a formação de nódulos nesse órgão (Brechenmacher *et al.*, 2010). Sendo assim, os níveis de expressão elevados do *GmSulfT1*, detectados neste trabalho, em raiz e semente de plantas de soja podem estar relacionados a uma maior distribuição e acúmulo de flavonóides nesses dois órgãos.

Na análise de expressão do *GmCit1*, a RT-PCR semiquantitativa e o *Northern blot* também apresentaram resultados semelhantes e coerentes com os resultados dos dados obtidos *in silico*, que apontaram expressão em folha. No *Northern blot* o RNA de aproximadamente 800 pb está de acordo com o tamanho indicado pelo Phytozome de 930 e 942 pb, ambos na fita antisenso, nos *loci* dos cromossomos Gm20 e Gm07, respectivamente. Esses dois genes,

nos quais a sequência do *GmCit1* alinhou, codificam para uma proteína pertencente à família PetM da subunidade 7 do complexo citocromo *B<sub>6</sub>f*. O complexo citocromo *B<sub>6</sub>f* é uma proteína composta por múltiplas subunidades com diversos grupos prostéticos (Taiz e Zeiger, 2002). O complexo participa do transporte de elétrons através da membrana tilacóide dos cloroplastos, onde ocorre a fotossíntese e constitui a conexão eletrônica entre os fotossistemas II e I. O fluxo de elétrons por esse complexo gera o gradiente de prótons necessário para síntese de ATP, que constitui o passo chave e limitante da fixação de CO<sub>2</sub> (Yamori *et al.*, 2011). Assim, o complexo citocromo *B<sub>6</sub>f* é o responsável pelo controle fotossintético (Yamori *et al.*, 2011). Nesse contexto é esperado que o complexo *B<sub>6</sub>f* esteja presente nos órgãos fotossintetizantes, como folha e vagem de plantas de soja. Geralmente, a taxa de fotossíntese líquida em vagens de leguminosas é menor do que em folha (Furbank *et al.*, 2004). Adicionalmente, os níveis menores de *GmCit1* em vagem podem ser relacionados ao decréscimo da capacidade fotossintética e conseqüentemente das enzimas participantes desse processo após a maturação desse órgão (Quebedeaux e Chollet, 1975). Diferenças apontadas entre os níveis de *GmCit1* na RT-PCR semiquantitativa e no *Northern blot* em folha e vagem de soja podem ter ocorrido devido à diferença de sensibilidade das técnicas ou às nossas condições experimentais.

As regiões promotoras dos genes *GmSulfT1* e *GmCit1* foram isoladas visando posterior utilização em transgenes para aplicações biotecnológicas. Sabendo-se que esses promotores apresentam módulos de direcionamento de expressão, a manipulação por deleções ou mutagênese dirigida desses módulos poderá levar ao direcionamento da expressão exclusivamente para um órgão (Cazzonelli e Velten, 2008).

Três fragmentos da região promotora do *GmSulfT1* com 0,5 kb, 0,8 kb e 0,9 kb foram amplificados com sucesso. A dificuldade de amplificar fragmentos maiores de modo específico se deu pela baixa complexidade da região promotora desse gene, que apresentou 35% de adenosina e 41% de timina, e pela alta diferença entre os TMs dos iniciadores senso e antisense. O desenho de iniciadores senso com TM aproximado do iniciador antisense de 56 °C (com ajuste de sal) não foi possível devido ao auto-anelamento previsto do iniciador.

No caso do *GmCit1*, a região promotora foi isolada do gene no cromossomo Gm20, por apresentar 100% de identidade de sequência com o *contig* selecionado. Pela comparação entre as regiões promotoras do gene do cromossomo Gm20 com o gene do cromossomo Gm07 (dados não mostrados), verificou-se que elas são semelhantes em 200 pb na extremidade 3', na região onde o iniciador antisense foi desenhado e no meio da sequência

(na posição de aproximadamente 1 150 pb nas sequências com 2 000 pb). As sequências em comum nas regiões promotoras desses dois genes podem ser devido à presença de motivos de elementos *cis* conservados entre os membros da família PetM da subunidade 7 do citocromo *B<sub>6</sub>f*. Nesse contexto, os iniciadores sensois desenhados são específicos para amplificação da sequência a montante do gene localizado no cromossomo Gm20, pois não foram desenhados nas sequências semelhantes. Todos os seis pares de iniciadores desenhados foram capazes de amplificar o respectivo fragmento da região promotora do *GmCit1* com 0,4 kb, 0,8 kb, 1,2 kb, 1,5 kb, 1,9 kb e 2,3 kb.

Apesar do uso da sequência genômica de *G. max* cv. Williams 82 como molde para desenho dos iniciadores, análises preliminares de sequenciamento dos promotores PSulfT0,5, PSulfT0,8, PSulfT0,9, PCit0,4, PCit0,8 e PCit1,9 amplificados de *G.max* cv. Conquista mostram que as sequências possuem alta similaridade. No entanto, são necessários dados de sequenciamento com maior qualidade e em maior volume para detectar se tais diferenças de fato ocorrem ou são artefatos da técnica.

A caracterização da região promotora dos genes *GmSulfT1* e *GmCit1* por análise qualitativa da expressão de *gus* fusionado aos fragmentos truncados foi realizada in vivo em plantas transgênicas de fumo. Uma vez que a regulação da expressão gênica é resultado da interação de diversos fatores, o primeiro passo para interpretar os dados obtidos foi identificar motivos putativos de elementos *cis* presentes na região promotora de *GmSulfT1* e *GmCit1*.

Neste trabalho foram considerados para interpretação dos resultados somente aqueles elementos *cis* essenciais para expressão órgão/tecido- específica e aqueles encontrados em promotores de genes essenciais ao desempenho das funções semelhantes ao *GmSulfT1*, ou seja, assimilação de enxofre e ao *GmCit1*, transporte de elétrons e participação na fotossíntese. Os outros elementos são sequências de DNA reconhecidas por famílias de fatores de transcrição cujos membros participam da regulação gênica de diversos modos, tornando difícil inferir a função que esse elemento possa ter na região promotora em questão.

O nível de expressão alto em folhas de plantas transformadas com PSulfT0,5 pode ser explicado, em parte, pela presença de motivos do elemento CACTFTPPCA1 nesse promotor. O mesmo elemento *cis* foi identificado no módulo de expressão em mesófilo (*Mem1*) do promotor do gene fosfoenolpiruvato carboxilase A (*ppcA*) de *Flaveria* sp. e é necessário para gerar o padrão de expressão mesófilo-específica desse gene (Gowik *et al.*, 2004). Os

elementos relacionados com a expressão raiz-específica ROOTMOTIFTAPOX1, com quatro motivos putativos e OSE2ROOTNODULE, com um motivo, também foram identificados no PSulfT0,5. O elemento ROOTMOTIFTAPOX1 foi encontrado em dois módulos do promotor do gene *rolD* de *A. rhizogenes* que regulam a expressão gênica em raiz em altos níveis na zona de alongação e no tecido vascular e em menor grau no ápice (Elmayan e Tepfer, 1995). O padrão de expressão em raiz do gene *gus* sob regulação de PSulfT0,5, mais elevado no tecido vascular do que no ápice, sugere que esse elemento seja responsável por esse padrão.

Os altos níveis de expressão detectados em raiz e folhas das plantas transformadas com PSulfT0,5 podem ser resultado do efeito sinérgico da interação entre os fatores responsáveis pela expressão específica nesses dois órgãos, a exemplo do funcionamento dos domínios A e B do promotor CaMV 35S em plantas de fumo (Benfey e Chua, 1990). Outra possibilidade é a de que o módulo da região promotora responsável pela órgão-especificidade do *GmSulfT1* esteja na região distal e exerça controle negativo sobre os outros elementos *cis*, ou seja, reprimindo atividade dos outros módulos, como observado em outros estudos com promotores de soja (Strömvik *et al.*, 1999; Waclawovsky *et al.*, 2006). Waclawovsky *et al.* (2006) identificaram um domínio na região distal do promotor do gene de soja *GmSBP2* necessário para expressão específica no tecido vascular dos órgãos vegetativos e nos tecidos da semente e cuja deleção resultou em um promotor constitutivo altamente ativo. Tendo em vista os resultados obtidos pelos autores do estudo citado e a concentração de elementos *cis* relacionados à expressão específica em folha na região proximal e à concentração de motivos de elementos relacionados à expressão em raiz na região distal do PSulfT0,5, a hipótese proposta anteriormente é, então, reforçada.

O elemento OSE (*organ specific element*) está envolvido na expressão órgão-específica do gene leghemoglobina *lbc3* de soja em nódulos de raiz, com duas sequências consenso: AAAGAT (OSE1ROOTNODULE) e CTCTT (OSE2ROOTNODULE) (Ramlov *et al.*, 1993). A presença desses dois motivos putativos na região promotora com 1,8 kb sugere que o *GmSulfT1* possa ter algum papel na formação de nódulos em raízes de soja. A evidência de que a distribuição de algumas flavonol-sulfotransferases pode ser relacionada ao acúmulo de flavonóides nos órgãos da planta (Gidda e Varin, 2006) reforça a suposição feita, uma vez que a presença de flavonóides na raiz é necessária para a formação de nódulos em leguminosas (Subramanian *et al.*, 2007; Wasson *et al.*, 2006).

A análise *in silico* do promotor de *GmSulfT1* acusou o elemento *cis* putativo SURE (*sulfur-responsiv element*) na posição -1876pb a partir do sítio de início da tradução. Esse elemento é necessário para assimilação e regulação de enxofre nas plantas e também foi encontrado na região distal do promotor do gene SULTR1.1, um transportador de sulfato de *Arabidopsis* localizado nos pelos, epiderme e córtex de raiz que é altamente afetado pela deficiência desse composto (Maruyama-Nakashita *et al.*, 2005). Ainda segundo o autor, a sequência consenso GAGAC do SURE foi encontrada em outros quinze promotores de genes induzíveis por enxofre, indicando que esse elemento é conservado e pode ter um papel chave na indução. A presença desse elemento sugere que a região promotora do *GmSulfT1* seja induzível pelos níveis de enxofre disponíveis à planta.

Na região promotora de *GmSulfT1* com 1,8 kb foram detectadas também duas sequências ACGT que são essenciais para expressão endosperma-específica do gene da proteína de armazenamento glutelina *Glub-1* de arroz (Washida *et al.*, 1999). Adicionalmente, foram encontrados sítios reconhecidos pelas proteínas nucleares SEF1 (*soybean embryo factor 1*) e SEF4 (*soybean embryo factor 4*) de soja. Esses sítios foram encontrados em um *enhancer* relacionado ao aumento da expressão semente específica do gene que codifica a subunidade ' da -conglucina de soja (Allen *et al.*, 1989; Lessard *et al.*, 1991). Assim, a expressão de *GmSulfT1* em semente pode estar relacionada a esses motivos presentes em sua região promotora.

O perfil de expressão de *GmCit1* preferencial em folhas de plantas de soja pode ser conferido pela região proximal de sua região promotora. Isso pode ser proposto por meio da observação da expressão de *gus* a jusante de PCit0,4 em plantas de fumo transgênicas. A expressão em raiz, fraca quando comparada com folha, demonstra que esse fragmento pode ser capaz de regular a expressão órgão-específica ou, pelo menos, órgão-preferencial em folha.

Do TSS à extremidade 5' da construção PCit0,4 nota-se a presença de somente um motivo do elemento CACTFTPPCA1, necessário para a expressão específica em folha e nenhum motivo relacionado à expressão específica em raiz. Se o início da transcrição ocorreu no TSS putativo, o motivo encontrado do CACTFTPPCA1 pode ser por si só responsável pela regulação específica. No entanto, a capacidade da RNA polimerase II pode reconhecer outros sítios para início da transcrição em promotores sem TATA (Achard *et al.*, 2003; Nakamura *et al.*, 2002; Smale e Kadonaga, 2003). Adicionalmente, a existência de um motivo do elemento

INRNTPSADB que possui a sequência consenso de iniciador (Inr) (Nakamura *et al.*, 2002), sugere a possibilidade que os cinco motivos do CACTFTPPCA1 entre o ATG e o TSS também participem da atividade de PCit0,4. Essa proposição é apoiada pela observação da atividade da GUS, embora fraca, em raiz de duas plantas, que pode ser resultado da presença de um motivo do elemento ROOTMOTIFAPOX1 entre o ATG e o TSS putativo.

O outro motivo relacionado à expressão específica em raiz presente nessa região é o OSE2ROOTNODULE, porém sabe-se que ele dirige essa expressão apenas na presença do OSE1ROOTNODULE (Ramlov *et al.*, 1993). O resultado obtido nas análises de expressão da construção PCit0,4 é comparável ao obtido por Achard *et al.* (2003) nas análises do promotor sem TATA do gene RPL21 de expressão em folha. No trabalho citado, o autor relata que o promotor mínimo de RPL21 é capaz de conferir o padrão folha-específico de expressão do gene. Além da expressão folha-específica, sugere-se que o promotor de PCit0,4 também seja responsivo à luz devido à presença de seis motivos putativos de quatro elementos *cis* diferentes relacionados à resposta a luz.

Os dois motivos de ROOTMOTIFTAPOX1 presentes na região entre -0,4 kb e -0,8 kb de PCit0,8 podem ter sido responsáveis pelo aparecimento do alto nível de expressão em raiz. Além disso, propõe-se que o módulo entre -0,4kb e -0,8kb da região promotora do *GmCit1* atue de forma sinérgica com o módulo compreendido entre o ATG e -0,4kb, pois, embora as análises *in silico* não terem indicado elementos *cis* putativos relacionados à expressão específica em folha, a atividade de GUS nesse órgão aumentou significativamente ao ser comparada com a conferida pela construção de PCit0,4.

A atividade da GUS em plantas transformadas com PCit1,9 manteve-se em níveis similares àquela encontrada nas plantas com PCit0,8, apesar do aumento do número de motivos de elementos *cis* relacionados à expressão específica em folha e raiz.

A ausência de TATA Box canônico na região proximal do promotor do *GmCit1* reforça a hipótese proposta por Nakamura *et al.* (2002) de que promotores sem TATA Box são característicos de grupos gênicos envolvidos na fotossíntese. Outra característica marcante desses grupos é a resposta à luz (Nakamura *et al.*, 2002). A detecção dos elementos responsivos à luz no promotor de 2,3 kb do *GmCit1* sugere que esse promotor também possui tal característica.



Além dos motivos *cis* regulatórios identificados *in silico* outros fatores podem influenciar atividade do promotor. Como em todos os trabalhos envolvendo obtenção de plantas transgênicas, a diferença observada de atividade da GUS entre folha e raiz das plantas transformadas com a mesma construção pode ter sido influenciada pelo número de cópias e pelas regiões que flanqueiam o DNA inserido no genoma da planta hospedeira, uma vez que o local de inserção no genoma não pode ser previsto em transformações via agrobactéria.

Devido à natureza combinatória da maioria dos promotores eucarióticos, não foi possível determinar exatamente quais elementos *cis* contribuem para a função dos promotores. Outro aspecto a ser considerado são os mecanismos regulatórios epigenéticos que impactam diretamente os níveis de mRNA, dificultando a associação direta entre uma estrutura específica do promotor com a produção de mRNA (Cazzonelli e Velten, 2008). No entanto, as análises *in silico* auxiliam o estudo dessas contribuições e podem ser utilizadas para direcionar outros experimentos.

## **6. Conclusão**

- O *contig* 8703 possui o perfil de expressão baseado em ESTs preferencial em semente de plantas de soja.
- O *contig* 18151 possui o perfil de expressão baseado em ESTs preferencial em folha de plantas de soja.
- Os perfis de expressão baseados em ESTs dos genes putativos nos bancos de sequências utilizados neste trabalho possuem significância biológica.
- O gene *GmSulfT1* é expresso preferencialmente em raiz e em semente de plantas de soja.
- *GmSulfT1* está inserido em um *locus* no cromossomo Gm13 do genoma da soja e codifica o transcrito Glyma13g26070.1.
- O gene *GmCit1* é expresso preferencialmente em folha de plantas de soja.
- *GmCit1* está inserido no cromossomo Gm20 do genoma da soja e codifica o transcrito Glyma20g01120.1.
- A partir da sequência genômica de *G.max* cv. Williams 82 foi possível o isolamento de fragmentos das regiões promotoras dos genes *GmSulfT1* e *GmCit1* do genoma de *G.max* cv. Conquista.
- A região promotora de *GmCit1* não possui TATA Box canônico, semelhantemente a outros genes de atuação na fotossíntese.
- Os promotores PSulfT0,5, PCit0,4, PCit0,8, PCit1,9 foram capazes de ativar a expressão de *gus* em plantas transgênicas de fumo.
- O promotor PSulfT0,5 confere forte expressão em folha e raiz de fumo.
- O promotor PCit0,4 confere expressão preferencial em folha de fumo.
- Os promotores PCit0,8 e PCit1,9 conferem forte expressão em folha e raiz de fumo.

## **7. Perspectivas**

Os promotores isolados nesse trabalho são potenciais ferramentas biotecnológicas em programas de melhoramento de plantas. Geração de deleções mais específicas ou mutagênese dentro dessas regiões poderão resultar no isolamento de promotores com atividade exclusiva em um desses órgãos. O caráter de alta expressão em mais de um órgão, observado em alguns fragmentos desses promotores, pode, da mesma forma, ser explorado para expressão de transgenes em plantas de soja e outras leguminosas. Existe a possibilidade dos promotores isolados neste trabalho regular diferencialmente a expressão em plantas de soja e fumo e isso deve ser testado, já que o repertório de fatores de transcrição nessas duas espécies pode ser diferente.

Para melhor caracterizar as regiões promotoras obtidas dos genes *GmSulfT1* e *GmCit1*, novas plantas de fumo deverão ser obtidas com os vetores binários construídos nesse trabalho contendo os promotores PSulfT0,5, PSulfT0,8, PSulfT0,9, PCit0,4, PCit0,8, PCit1,5, PCit1,9 e PCaMV 35S. Isso será feito com intuito de observar a contribuição para o funcionamento global do promotor das regiões deletadas que ainda não foram analisadas e de obter mais plantas transformadas com as construções já avaliadas. A atividade de GUS nos órgãos das plantas transformadas e de sua progênie será avaliada quantitativamente bem como os níveis de expressão do gene.

## **8. Referências Bibliográficas**

ACHARD, P.; LAGRANGE, T.; EL-ZANATY, A. F.; MACHE, R. Architecture and transcriptional activity of the initiator element of the TATA-less RPL21 gene. **Plant Journal**, v. 35, n. 6, p. 743-752, 2003.

AINSWORTH, E. A.; YENDREK, C. R.; SKONECZKA, J. A.; LONG, S. P. Accelerating yield potential in soybean: potential targets for biotechnological improvement. **Plant, Cell & Environment**, v. 35, n. 1, p. 38-52, 2012.

ALBERTS, B.; JOHNSON, A.; JULIAN LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. Control of Gene Expression. In: (Ed.). **Molecular Biology of the Cell**. 4. New York: Garland Science, p.375-466, 2002.

ALLEN, R. D.; BERNIER, F.; LESSARD, P. A.; BEACHY, R. N. Nuclear factors interact with a soybean beta-conglycinin enhancer. **Plant Cell**, v. 1, n. 6, p. 623-31, Jun 1989.

ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, v. 25, n. 17, p. 3389-402, Sep 1997.

AUDIC, S.; CLAVERIE, J. M. The significance of digital gene expression profiles. **Genome Res**, v. 7, n. 10, p. 986-95, Oct 1997.

BARROS, L. M. G. Transformação genética de *Nicotiana tabacum* cv Xanthi utilizando *Agrobacterium tumefaciens* e eletroporação. **Dissertação de mestrado**. Universidade de Brasília, DF, Brasil, 117p, 1989

BENFEY, P. N.; CHUA, N. H. The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants. **Science**, v. 250, n. 4983, p. 959-66, Nov 1990.

BENFEY, P. N.; REN, L.; CHUA, N. H. The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. **EMBO Journal**, v. 8, n. 8, p. 2195-202, Aug 1989.

BENFEY, P. N.; REN, L.; CHUA, N. H. Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. **EMBO Journal**, v. 9, n. 6, p. 1685-96, Jun 1990.

BLUMENTHAL, T. Operons in eukaryotes. **Briefing in Functional Genomics and Proteomics**, v.3, n.3, p.199-211, Nov 2004.

BUHEL, A. S.; BREDERODE, F. T.; BOL, J. F.; LINTHORST, H. J. Mutation of GT-1 binding sites in the Pr-1A promoter influences the level of inducible gene expression in vivo. **Plant Molecular Biology**, v. 40, n. 3, p. 387-96, Jun 1999.

BRECHENMACHER, L.; LEI, Z.; LIBAULT, M.; FINDLEY, S.; SUGAWARA, M.; SADOWSKY, M. J.; SUMNER, L. W.; STACEY, G. Soybean Metabolites Regulated in Root Hairs in Response to the Symbiotic Bacterium *Bradyrhizobium japonicum*. **Plant Physiology**, v. 153, n. 4, p. 1808-1822, Aug 2010.

CAMMUE, B. P. A.; BUTAYE, K. M. J.; DELAURE, S. L.; DE BOLLE, M. F. C. Approaches to minimize variation of transgene expression in plants. **Molecular Breeding**, v. 16, n. 1, p. 79-91, Aug 2005.

CARNEIRO, M.; VILAINE, F. Differential expression of the rolA plant oncogene and its effect on tobacco development. **Plant Journal**, v. 3, n. 6, p. 785-792, 1993.

CAZZONELLI, C. I.; VELTEN, J. In vivo characterization of plant promoter element interaction using synthetic promoters. **Transgenic Research**, v. 17, n. 3, p. 437-457, Jun 2008.

CHEN, M.; XU, Z.; XIA, L.; LI, L.; CHENG, X.; DONG, J.; WANG, Q.; MA, Y. Cold-induced modulation and functional analyses of the DRE-binding transcription factor gene, GmDREB3, in soybean (*Glycine max* L.). **Journal of Experimental Botany**, v. 60, n. 1, p. 121-35, 2009.

COMINELLI, E.; TONELLI, C. Transgenic crops coping with water scarcity. **New Biotechnology**, v. 27, n. 5, p. 473-477, 2010.

CUNHA, N. B.; MURAD, A. M.; CIPRIANO, T. M.; ARAÚJO, A. C. G.; ARAGÃO, F. J. L.; LEITE, A.; VIANNA, G. R.; MCPHEE, T. R.; SOUZA, G. H. M. F.; WATERS, M. J.; RECH, E. L. Expression of functional recombinant human growth hormone in transgenic soybean seeds. **Transgenic Research**, v. 20, n. 4, p. 811-826, 2011.

CURTIS, M. D. A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. **Plant Physiology**, v. 133, n. 2, p. 462-469, 2003.

DANIELL, H. Molecular strategies for gene containment in transgenic crops. **Nature Biotechnology**, v. 20, n. 6, p. 581-586, 2002.



DHANKHER, O. P.; LI, Y.; ROSEN, B. P.; SHI, J.; SALT, D.; SENECOFF, J. F.; SASHTI, N. A.; MEAGHER, R. B. Engineering tolerance and hyperaccumulation of arsenic in plants by combining arsenate reductase and  $\gamma$ -glutamylcysteine synthetase expression. **Nature Biotechnology**, v. 20, n. 11, p. 1140-1145, 2002.

DOYLE, J. J.; DOYLE, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin**, v. 19, p. 11-15, 1987.

ELMAYAN, T.; TEPFER, M. Evaluation in tobacco of the organ specificity and strength of the rolD promoter, domain A of the 35S promoter and the 35S2 promoter. **Transgenic Research**, v. 4, n. 6, p. 388-96, Nov 1995.

Embrapa Soja. Disponível em:  
<[http://www.cnpso.embrapa.br/index.php?op\\_page=294&cod\\_pai=17](http://www.cnpso.embrapa.br/index.php?op_page=294&cod_pai=17)>. Acesso em cinco de janeiro de 2012.

EWING, R. M.; BEN KAHLA, A.; POIROT, O.; LOPEZ, F.; AUDIC, S.; CLAVERIE, J. M. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. **Genome Research**, v. 9, n. 10, p. 950-9, Oct 1999.

FEHR, W. R.; CAVINESS, C. E. Stages of Soybean Development. **Ames: Iowa State University of Science and Technology**, v. 80, p. 1-12, 1977.

FIELD, B.; FISTON-LAVIER, A.-S.; KEMEN, A.; GEISLER, K.; QUESNEVILLE, H.; OSBOURN, A. E. Formation of plant metabolic gene clusters within dynamic chromosomal regions. **Proceedings of the National Academy of Sciences**, v. 108, n. 38, p. 16116-16121, September 20, 2011.

FURBANK, R. T.; WHITE, R.; PALTA, J. A.; TURNER, N. C. Internal recycling of respiratory CO<sub>2</sub> in pods of chickpea (*Cicer arietinum* L.): the role of pod wall, seed coat, and embryo. **Journal of Experimental Botany**, v. 55, n. 403, p. 1687-1696, Aug 2004.

GHANEM, M. E.; HICHRI, I.; SMIGOCKI, A. C.; ALBACETE, A.; FAUCONNIER, M. L.; DIATLOFF, E.; MARTINEZ-ANDUJAR, C.; LUTTS, S.; DODD, I. C.; PÉREZ-ALFOCEA, F. Root-targeted biotechnology to mediate hormonal signalling and improve crop stress tolerance. **Plant Cell Reports**, v. 30, n. 5, p. 807-823, 2011.

GIDDA, S. K.; VARIN, L. Biochemical and molecular characterization of flavonoid 7-sulfotransferase from *Arabidopsis thaliana*. **Plant Physiology and Biochemistry**, v. 44, n. 11–12, p. 628-636, 2006.

GITTINS, J. R.; PELLNY, T. K.; HILES, E. R.; ROSA, C.; BIRICOLTI, S.; JAMES, D. J. Transgene expression driven by heterologous ribulose-1,5-bisphosphate carboxylase/oxygenase small-subunit gene promoters in the vegetative tissues of apple. **Planta**, v. 210, n. 2, p. 232-240, 2000.

GOWIK, U.; BURSCHEIDT, J.; AKYILDIZ, M.; SCHLUE, U.; KOCZOR, M.; STREUBEL, M.; WESTHOFF, P. cis-Regulatory elements for mesophyll-specific gene expression in the C4 plant *Flaveria trinervia*, the promoter of the C4 phosphoenolpyruvate carboxylase gene. **Plant Cell**, v. 16, n. 5, p. 1077-90, May 2004.

GRAHAM, T. L. Flavonoid and Isoflavonoid Distribution in Developing Soybean Seedling Tissues and in Seed and Root Exudates. **Plant Physiology**, v. 95, n. 2, p. 594-603, Feb, 1991.

HAAG, J. R.; PIKAARD, C. S. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. **Nature Reviews Molecular Cell Biology**, v. 12, n. 8, p. 483-492, 2011.

HARTMAN, G. L.; WEST, E. D.; HERMAN, T. K. Crops that feed the World 2. Soybean-worldwide production, use, and constraints caused by pathogens and pests. **Food Security**, v. 3, n. 1, p. 5-17, 2011.

HERNANDEZ-GARCIA, C. M.; BOUCHARD, R. A.; RUSHTON, P. J.; JONES, M. L.; CHEN, X.; TIMKO, M. P.; FINER, J. J. High level transgenic expression of soybean (*Glycine max*) GmERF and Gmubi gene promoters isolated by a novel promoter analysis pipeline. **BMC Plant Biology**, v. 10, n. 1, p. 237, 2010.

HIGO, K.; UGAWA, Y.; IWAMOTO, M.; HIGO, H. PLACE: a database of plant cis-acting regulatory DNA elements. **Nucleic Acids Research**, v. 26, n. 1, p. 358-9, Jan 1998.

IRANI, R. J.; ÖZKUMUR, E.; AHN, S.; YALÇIN, A.; ÜNLÜ, M. S.; DELISI, C. P. TATA Binding Protein binds tightly to single stranded DNA: A new model for Eukaryotic Transcription. **Nature Precedings**, 2010.

- JAMES, C. Global Status of Commercialized Biotech/GM Crops: 2009. **ISAAA Brief**, v. 41, 2009.
- JUVEN-GERSHON, T.; HSU, J.-Y.; THEISEN, J. W. M.; KADONAGA, J. T. The RNA polymerase II core promoter — the gateway to transcription. **Current Opinion in Cell Biology**, v. 20, n. 3, p. 253-259, 2008.
- JUVEN-GERSHON, T.; KADONAGA, J. T. Regulation of gene expression via the core promoter and the basal transcriptional machinery. **Developmental Biology**, v. 339, n. 2, p. 225-229, 2010.
- KAY, R.; CHAN, A.; DALY, M.; MCPHERSON, J. Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes. **Science**, v. 236, n. 4806, p. 1299-302, Jun 1987.
- KEYSER, H.H.; LI, F.D. Potential for increasing biological nitrogen-fixation in soybean. **Plant and Soil**, v. 141, n. 1-2, p. 119-135, 1992.
- KIM, D. W.; LEE, S. H.; CHOI, S. B.; WON, S. K.; HEO, Y. K.; CHO, M.; PARK, Y. I.; CHO, H. T. Functional conservation of a root hair cell-specific cis-element in angiosperms with different root hair distribution patterns. **Plant Cell**, v. 18, n. 11, p. 2958-70, Nov 2006.
- KOMATSU, S.; THIBAUT, D.; HIRAGA, S.; KATO, M.; CHIBA, M.; HASHIGUCHI, A.; TOUGOU, M.; SHIMAMURA, S.; YASUE, H. Characterization of a novel flooding stress-responsive alcohol dehydrogenase expressed in soybean roots. **Plant Molecular Biology**, v. 77, n. 3, p. 309-322, 2011.
- KORNBERG, R. D. Eukaryotic transcriptional control. **Trends in Cell Biology**, v. 9, n. 12, p. 46-49, 1999.
- KORNBERG, R. D. The molecular basis of eukaryotic transcription. **Proceedings of the National Academy of Sciences**, v. 104, n. 32, p. 12955-12961, 2007.
- LAURIDSEN, P.; FRANSSEN, H.; STOUGAARD, J.; BISSELING, T.; MARCKER, K. A. Conserved regulation of the soybean early nodulin ENOD2 gene promoter in determinate and indeterminate transgenic root nodules. **Plant Journal**, v. 3, n. 3, p. 483-492, 1993.

LESSARD, P. A.; ALLEN, R. D.; BERNIER, F.; CRISPINO, J. D.; FUJIWARA, T.; BEACHY, R. N. Multiple nuclear factors interact with upstream sequences of differentially regulated beta-conglycinin genes. **Plant Molecular Biology**, v. 16, n. 3, p. 397-413, Mar 1991.

LEUSTEK, T.; SAITO, K. Sulfate transport and assimilation in plants. **Plant Physiology**, v. 120, n. 3, p. 637-44, Jul 1999.

LODISH, H.; BERK, A.; ZIPURSKY, S. L.; MATSUDAIRA, P.; BALTIMORE, D.; DARNELL, J. Regulation of Transcription Initiation. In: (Ed.). **Molecular Cell Biology**. 4. New York: W. H. Freeman and Company, 2000.

MARUYAMA-NAKASHITA, A.; NAKAMURA, Y.; WATANABE-TAKAHASHI, A.; INOUE, E.; YAMAYA, T.; TAKAHASHI, H. Identification of a novel cis-acting element conferring sulfur deficiency response in Arabidopsis roots. **Plant Journal**, v. 42, n. 3, p. 305-14, May 2005.

MASTON, G. A.; EVANS, S. K.; GREEN, M. R. Transcriptional regulatory elements in the human genome. **Annual Review of Genomics and Human Genetics**, v. 7, p. 29-59, 2006.

MATSON, P. A.; PARTON, W. J.; POWER, A. G.; SWIFT, M. J. Agricultural Intensification and Ecosystem Properties. **Science**, v. 277, n. 5325, p. 504-509, July 25, 1997.

MCCABE, D. E.; SWAIN, W. F.; MARTINELL, B. J.; CHRISTOU, P. Stable transformation of soybean (*Glycine max*) by particle acceleration. **Biotechnology**, v. 6, p. 923-926, 1988.

Ministério da Agricultura Pecuária e Abastecimento. Disponível em: <<http://www.agricultura.gov.br/vegetal/culturas/soja>>. Acesso em: cinco de janeiro de 2012.

Ministério da Agricultura Pecuária e Abastecimento. Listagem de OGM autorizados no Brasil, 2011. Disponível em: <<http://www.agricultura.gov.br/portal/page/portal/Internet-MAPA/pagina-inicial/vegetal/organismos-geneticamente-modificados/plantas-autorizadas>>. Acesso em cinco de janeiro de 2012.

MURASHIGE, T.; SKOOG, F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. **Physologia Plantarum**, v. 15, p. 473-497, 1962.

- MURRAY, D.; DORAN, P.; MACMATHUNA, P.; MOSS, A. C. *In silico* gene expression analysis – an overview. **Molecular Cancer**, v. 6, n. 1, p. 50, 2007.
- NAKAMURA, M.; TSUNODA, T.; OBOKATA, J. Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. **Plant Journal**, v. 29, n. 1, p. 1-10, Jan 2002.
- ODELL, J. T.; NAGY, F.; CHUA, N. H. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. **Nature**, v. 313, n. 6005, p. 810-2, Feb-Mar 1985.
- POTENZA, C.; ALEMAN, L.; SENGUPTA-GOPALAN, C. Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. **In Vitro Cellular & Developmental Biology - Plant**, v. 40, n. 1, p. 1-22, 2004.
- PREISZNER, J.; VANTOAI, T.; HUYNH, L.; BOLLA, R.; YEN, H. Structure and activity of a soybean Adh promoter in transgenic hairy roots. **Plant Cell Reports**, v. 20, n. 8, p. 763-769, 2001.
- PRIEST, H. D.; FILICHKIN, S. A.; MOCKLER, T. C. Cis-regulatory elements in plant cell signaling. **Current Opinion in Plant Biology**, v. 12, n. 5, p. 643-9, Oct 2009.
- PUGH, B. F.; TJIAN, R. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. **Genes & Development**, v. 5, n. 11, p. 1935-1945, November 1, 1991.
- QUANDT, H. J.; BROER, I.; PÜHLER, A. Tissue-specific activity and light-dependent regulation of a soybean rbcS promoter in transgenic tobacco plants monitored with the firefly luciferase gene. **Plant Science**, v. 82, n. 1, p. 59-70, 1992.
- QUEBEDEAUX, B.; CHOLLET, R. Growth and Development of Soybean (*Glycine max* [L.] Merr.) Pods: CO<sub>2</sub> Exchange and Enzyme Studies. **Plant Physiology**, v. 55, n. 4, p. 745-748, Apr, 1975.
- RAMLOV, K. B.; LAURSEN, N. B.; STOUGAARD, J.; MARCKER, K. A. Site-directed mutagenesis of the organ-specific element in the soybean leghemoglobin lbc3 gene promoter. **Plant Journal**, v. 4, n. 3, p. 577-80, Sep 1993.

ROZEN, S.; SKALETSKY, H. Primer3 on the WWW for general users and for biologist programmers. **Methods in Molecular Biology**, v. 132, p. 365-86, 2000.

SAEED, H. A.; VODKIN, L. O.; STRÖMVIK, M. V. Promoters of the soybean seed lectin homologues Le2 and Le3 regulate gene expression in vegetative tissues in Arabidopsis. **Plant Science**, v. 175, n. 6, p. 868-876, 2008.

SALVAGIOTTI, F.; CASSMAN, K.G.; SPECHT, J.E.; WALTERS, D.T.; WEISS, A.; DOBERMANN, A. Nitrogen uptake, fixation and response to fertilizer N in soybeans: A review. **Field Crops Research**, v.108, n. 1, p. 1-13, 2008.

SAMBROOK, J.; RUSSELL, D. W. **Molecular cloning - A laboratory manual**. 3. New York: Cold Spring Harbor Laboratory Press, 2001.

SCHMUTZ, J.; CANNON, S. B.; SCHLUETER, J.; MA, J.; MITROS, T.; NELSON, W.; HYTEN, D. L.; SONG, Q.; THELEN, J. J.; CHENG, J.; XU, D.; HELLSTEN, U.; MAY, G. D.; YU, Y.; SAKURAI, T.; UMEZAWA, T.; BHATTACHARYYA, M. K.; SANDHU, D.; VALLIYODAN, B.; LINDQUIST, E.; PETO, M.; GRANT, D.; SHU, S.; GOODSTEIN, D.; BARRY, K.; FUTRELL-GRIGGS, M.; ABERNATHY, B.; DU, J.; TIAN, Z.; ZHU, L.; GILL, N.; JOSHI, T.; LIBAULT, M.; SETHURAMAN, A.; ZHANG, X. C.; SHINOZAKI, K.; NGUYEN, H. T.; WING, R. A.; CREGAN, P.; SPECHT, J.; GRIMWOOD, J.; ROKHSAR, D.; STACEY, G.; SHOEMAKER, R. C.; JACKSON, S. A. Genome sequence of the palaeopolyploid soybean. **Nature**, v. 463, n. 7278, p. 178-183, 2010.

SHOEMAKER, R.; KEIM, P.; VODKIN, L.; RETZEL, E.; CLIFTON, S. W.; WATERSTON, R.; SMOLLER, D.; CORYELL, V.; KHANNA, A.; ERPELDING, J.; GAI, X.; BRENDEL, V.; RAPH-SCHMIDT, C.; SHOOP, E. G.; VIELWEBER, C. J.; SCHMATZ, M.; PAPE, D.; BOWERS, Y.; THEISING, B.; MARTIN, J.; DANTE, M.; WYLIE, T.; GRANGER, C. A compilation of soybean ESTs: generation and analysis. **Genome**, v. 45, n. 2, p. 329-338, 2002.

SHOEMAKER, R. C.; SCHLUETER, J.; DOYLE, J. J. Paleopolyploidy and gene duplication in soybean and other legumes. **Current Opinion in Plant Biology**, v. 9, n. 2, p. 104-109, 2006.

SINGH, O. V.; GHAI, S.; PAUL, D.; JAIN, R. K. Genetically modified crops: Success, safety assessment, and public concern. **Applied Microbiology and Biotechnology**, v. 71, n. 5, p. 598-607, 2006.

SINGH, R. J.; HYMOWITZ, T. Soybean genetic resources and crop improvement. **Genome**, v. 42, n. 4, p. 605-616, Aug 1999.

SMALE, S. T. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. **Biochimica et Biophysica Acta**, v. 1351, n. 1-2, p. 73-88, Mar 1997.

SMALE, S. T.; KADONAGA, J. T. Therna Polymeraseii Corepromoter. **Annual Review of Biochemistry**, v. 72, n. 1, p. 449-479, 2003.

STRÖMVIK, M. V.; SUNDARARAMAN, V. P.; VODKIN, L. O. A novel promoter from soybean that is active in a complex developmental pattern with and without its proximal 650 base pairs. **Plant Molecular Biology**, v. 41, n. 2, p. 217-231, 1999.

SUBRAMANIAN, S.; HU, X.; LU, G.; ODELLAND, J. T.; YU, O. The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. **Plant Molecular Biology**, v. 54, n. 5, p. 623-639, 2004.

SUBRAMANIAN, S.; STACEY, G.; YU, O. Distinct, crucial roles of flavonoids during legume nodulation. **Trends in Plant Science**, v. 12, n. 7, p. 282-285, 2007.

TAIZ, L.; ZEIGER, E. Photosynthesis: The Light Reactions. In: (Ed.). **Plant Physiology**. Sunderland: Sinauer Associates, Inc., 2002.

TAKAHASHI, H.; KOPRIVA, S.; GIORDANO, M.; SAITO, K.; HELL, R. Sulfur Assimilation in Photosynthetic Organisms: Molecular Functions and Regulations of Transporters and Assimilatory Enzymes. **Annual Review of Plant Biology**, v. 62, n. 1, p. 157-184, 2011.

TAKAI, D.; JONES, P. A. Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. **Molecular Biology and Evolution**, v.21, n.3, p. 463-467, 2004.

TRAN, L. S. P.; MOCHIDA, K. Functional genomics of soybean for improvement of productivity in adverse conditions. **Functional and Integrative Genomics**, v. 10, n. 4, p. 447-462, 2010.

TWYMAN, R. M., Ed. Growth and development: Control of gene expression, regulation of transcription. **Encyclopedia of Applied Plant Sciences**. London UK: Elsevier Science, p.558-567, 2003.

TWYMAN, R. M.; STOGER, E.; SCHILLBERG, S.; CHRISTOU, P.; FISCHER, R. Molecular farming in plants: host systems and expression technology. **Trends in Biotechnology**, v. 21, n. 12, p. 570-578, Dec 2003.

TYAGI, A. K. Plant genes and their expression. **Current Science**, v. 80, n. 2, p. 161-169, Jan 2001.

VARIN, L.; MARSOLAIS, F.; RICHARD, M.; ROULEAU, M. Sulfation and sulfotransferases 6: Biochemistry and molecular biology of plant sulfotransferases. **The FASEB Journal**, v. 11, n. 7, p. 517-525, Jun 1997.

VISARADA, K. B. R. S.; MEENA, K.; ARUNA, C.; SRUJANA, S.; SAIKISHORE, N.; SEETHARAMA, N. Transgenic breeding: Perspectives and prospects. **Crop Science**, v. 49, n. 5, p. 1555-1563, 2009.

WACLAWOVSKY, A. J.; FREITAS, R. L.; ROCHA, C. S.; CONTIM, L. A. S.; FONTES, E. P. B. Combinatorial regulation modules on GmSBP2 promoter: A distal cis-regulatory domain confines the SBP2 promoter activity to the vascular tissue in vegetative organs. **Biochimica Et Biophysica Acta-Genes Structure and Expression**, v. 1759, n. 1-2, p. 89-98, 2006.

WASHIDA, H.; WU, C. Y.; SUZUKI, A.; YAMANOUCHI, U.; AKIHAMA, T.; HARADA, K.; TAKAIWA, F. Identification of cis-regulatory elements required for endosperm expression of the rice storage protein glutelin gene GluB-1. **Plant Molecular Biology**, v. 40, n. 1, p. 1-12, May 1999.

WASSON, A. P.; PELLERONE, F. I.; MATHESIUS, U. Silencing the Flavonoid Pathway in *Medicago truncatula* Inhibits Root Nodule Formation and Prevents Auxin Transport Regulation by Rhizobia. **The Plant Cell Online**, v. 18, n. 7, p. 1617-1629, Jul 2006.



WEI, W.; PELECHANO, V.; JARVELIN, A. I.; STEINMETZ, L. M. Functional consequences of bidirectional promoters. **Trends in Genetics**, v.27, n.7, Jul, p.267-76, 2011.

YAMAMOTO, Y. Y.; YOSHIOKA, Y.; HYAKUMACHI, M.; OBOKATA, J. Characteristics of core promoter types with respect to gene structure and expression in *arabidopsis thaliana*. **DNA Research**, v. 18, n. 5, p. 333-342, 2011.

YAMORI, W.; TAKAHASHI, S.; MAKINO, A.; PRICE, G. D.; BADGER, M. R.; VON CAEMMERER, S. The Roles of ATP Synthase and the Cytochrome b6/f Complexes in Limiting Chloroplast Electron Transport and Determining Photosynthetic Capacity. **Plant Physiology**, v. 155, n. 2, p. 956-962, Feb 2011.

ZENZIE-GREGORY, B.; KHACHI, A.; GARRAWAY, I. P.; SMALE, S. T. Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. **Molecular Cell Biology**, v. 13, n. 7, p. 3841-9, Jul 1993.