

FLÁVIO TETSUO SASSAKI

**ISOLAMENTO E CARACTERIZAÇÃO DE PROMOTORES
ÓRGÃO-ESPECÍFICOS A PARTIR DE INFORMAÇÕES DO
BANCO FORESTS (*EUCALYPTUS* GENOME SEQUENCING
PROJECT CONSORTIUM)**

Botucatu - SP

2008

FLÁVIO TETSUO SASSAKI

**ISOLAMENTO E CARACTERIZAÇÃO DE PROMOTORES
ÓRGÃO-ESPECÍFICOS A PARTIR DE INFORMAÇÕES DO
BANCO FORESTS (*EUCALYPTUS* GENOME SEQUENCING
PROJECT CONSORTIUM)**

Tese apresentada ao Instituto de Biociências
da Universidade Estadual Paulista Júlio de
Mesquita Filho, Campus de Botucatu, para
obtenção do título de Doutor em Ciências
Biológicas - Genética

Orientador: Prof. Dr. Ivan de Godoy Maia

Botucatu - SP

2008

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO
DA INFORMAÇÃO
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: SELMA MARIA DE JESUS

Sasaki, Flávio Tetsuo.

Isolamento e caracterização de promotores órgão-específicos a partir de informações do Banco FORESTs (*Eucalyptus* Genome Sequencing Project Consortium) / Flávio Tetsuo Sasaki. – Botucatu : [s.n.], 2008.

Tese (doutorado) – Universidade Estadual Paulista, Instituto de Biociências de Botucatu 2008.

Orientador: Ivan de Godoy Maia

Assunto CAPES: 20203004

1. Eucalipto - Melhoramento genético 2. Genética vegetal
CDD 581.15

Palavras-chave: Biotecnologia; *Eucalyptus*; Expressão gênica; Promotores; Tecido-especificidade

AOS MEUS PAIS

AGRADECIMENTOS

- Ao professor Dr. Ivan de Godoy Maia, pela orientação, conselhos, amizade e paciência durante meu processo de aprendizado. Agradeço a confiança depositada em mim.
- Aos meus pais, pelo apoio moral, pela compreensão e apoio financeiro. Por acreditarem na minha carreira e depositarem confiança em mim.
- Aos meus irmãos Eduardo e Cristina, pela amizade e conversas.
- À Ana Elisa Madalena Rinaldi, pelo amor, carinho e companheirismo. Por ser uma batalhadora em que me inspiro.
- Ao Consórcio do FORESTs, pelo uso do banco de ESTs.
- À Votorantim Celulose e Papel S.A., pelo material vegetal coletado para extração de RNA total.
- À Suzano Bahia Sul Papel e Celulose S.A., especialmente ao Dr. Esteban Roberto González, pelo material vegetal coletado para PCR quantitativo em tempo real, e pela colaboração no experimento de ensaio transiente de GUS em *Eucalyptus*.
- Ao uso do banco de northern eletrônico do Prof. Dr. Marcelo Menossi e Renato Vicentini.
- Aos meus colegas de laboratório, Adriana, Antônio, Regiane, Akemi, Fábio, Roberto, Negin, Carla, Marcelo, Denise, Cíntia, Layra, Júlio, Vanusa, Tânia, Virgínia, Ana Teresa, Débora, Rodrigo, Edmárcia e Brandalise, pela amizade, ajuda e companheirismo.
- Ao pessoal do grupo do prof. dr. Paulo Ribolla, Diego, Letícia, Karina e Aline por me ajudarem nos seqüenciamentos de DNA. Agradeço igualmente a Virgínia também pelo auxílio no seqüenciamento.
- Aos meus colegas Regiane, Akemi e Fábio, por me ajudarem nas análises de qPCR.
- Ao Antônio, pela amizade e conversas interessantes sobre ciência. Pelas lições de bioinformática e proteínas.

- Aos meus companheiros de república, Eduardo e Heraldo, pela amizade e conversas.
- À colega Adriana, pela caracterização do candidato EGRT2 via RT-PCR e isolamento de promotores putativos.
- Ao Prof. Dr. Márcio de Castro Silva-Filho, pelo fornecimento das sementes de tabaco transgênico com o cassete 35S-GUS para o controle positivo no ensaio histoquímico de GUS.
- Ao apoio financeiro do CNPq.
- Ao apoio financeiro da FAPESP.

“Algo que aprendi em uma longa vida é que toda nossa ciência, comparada com a realidade, é primitiva e infantil, mas, ainda assim, é a coisa mais preciosa que temos”

Albert Einstein

Resumo

Os cassetes de expressão empregados atualmente para a produção de plantas geneticamente modificadas são baseados, na sua maioria, em promotores constitutivos que determinam a expressão generalizada do transgene na planta, o que muitas vezes não é necessário nem desejado. A alternativa mais viável para substituição de tais promotores é investir na identificação e caracterização de promotores órgão/tecido-específicos ou estímulo-dependentes nas espécies de interesse. O presente projeto teve como objetivos identificar e caracterizar genes com padrão de expressão órgão-específico em eucalipto e, a partir dessas informações, isolar e caracterizar as seqüências promotoras adjacentes. Predições *in silico* no banco de dados do projeto de seqüências expressas do eucalipto (FORESTs), e informações disponíveis na literatura a respeito de genes com padrão de expressão órgão-específico, foram utilizadas na seleção. Assim, 59 genes preditos como possuindo expressão exclusiva em dado órgão foram selecionados para a validação via RT-PCR. Dos candidatos validados, 2 eram de raiz, 1 de folha, 1 de botão floral e 1 de botão floral e fruto, concomitantemente. Três candidatos ubíquos também foram selecionados. Dois desses candidatos tiveram seus perfis de expressão em diferentes órgãos de eucalipto avaliados por PCR quantitativa, indicando que o primeiro é preferencialmente expresso em folhas, e o segundo expresso especificamente em raiz. A partir de estratégias de “genome walking” foram isolados diferentes promotores putativos, sendo que o promotor do candidato com expressão preferencial em folha (1.2 kb) foi selecionado para a caracterização funcional em plantas usando o gene *uidA*, que codifica a β -glucuronidase (GUS), como repórter. Em plântulas transgênicas de tabaco da geração T1, a expressão de GUS foi detectada majoritariamente em cotilédones. Em plantas adultas, porém, a expressão de GUS foi

detectada também no caule e na raiz, mas em menor intensidade se comparada com a expressão observada em folha. Esse promotor modulou a transcrição de GUS em ensaios transientes em plântulas de *Eucalyptus grandis*.

Palavras-chave: Expressão gênica; Promotores; Tecido-especificidade; *Eucalyptus*;

Biotecnologia

Abstract

Many plant genetic engineering applications require spatial expression of transgenes, which in turn depends upon the availability of tissue- and organ-specific promoters. In the present work, the identification of genes with organ-specific expression in *Eucalyptus grandis* was performed aiming the subsequent isolation and characterization of contiguous promoter sequences. Candidate genes were selected by *in silico* predictions in the *Eucalyptus* EST project (FORESTs) database or by searching the available literature for genes with specific expression patterns. Fifty-nine genes with predicted organ-specific expression were selected for further RT-PCR validation. Among the validated candidate genes, 2 were root-specific, 1 was leaf-specific, 1 was flower-bud-specific and 1 was flower-bud and fruit specific. Three genes ubiquitously expressed were also selected. The relative expression levels of two of them (one leaf-specific and one root-specific) over a broad range of eucalyptus organ/tissues were determined using real time PCR. The 5' putative regulatory sequences of the validated genes were isolated using genome walking strategies, and the activity of the leaf-specific promoter (1.2 kb) was further investigated in transgenic tobacco plants using a GUS reporter system. In transgenic seedlings from the T1 generation, GUS expression driven by this promoter was detected preferentially in cotyledons. In adult plants, however, GUS expression was detected in both stem and root, but in lower intensity if compared to the expression observed in leaves. This promoter was also able to modulate the transient expression of GUS in seedlings of *Eucalyptus grandis*.

Keywords: Gene expression; Promoters; Tissue-specificity; Eucalyptus;

Biotechnology

SUMÁRIO

RESUMO.....	1
ABSTRACT	3
SUMÁRIO	4
1) INTRODUÇÃO.....	5
1.1) <i>O eucalipto</i>	5
1.2) <i>Regulação transcricional</i>	8
2) OBJETIVOS	20
3) MATERIAL E MÉTODOS	20
3.1) <i>Identificação dos ESTs com expressão órgão-específica</i>	20
3.2) <i>Extração de RNA total de eucalipto e tabaco</i>	22
3.3) <i>Validação biológica por RT-PCR e qPCR</i>	24
3.4) <i>Isolamento das regiões promotoras</i>	27
3.5) <i>Clonagem e Sequenciamento dos produtos de amplificação</i>	30
3.6) <i>Confirmação da identidade das seqüências clonadas</i>	31
3.7) <i>Construção do cassete de expressão</i>	32
3.8) <i>Produção de células competentes de Agrobacterium tumefaciens</i>	33
3.9) <i>Transformação do vetor binário em células competentes de A. tumefaciens</i>	34
3.10) <i>Preparação da cultura de A. tumefaciens e infecção dos discos foliares de tabaco</i>	34
3.11) <i>Regeneração de plantas transformadas</i>	36
3.12) <i>Identificação e caracterização molecular das linhagens transgênicas</i>	37
3.13) <i>Ensaio histoquímico da atividade GUS em tabaco</i>	38
3.14) <i>Identificação do início de transcrição pela técnica 5' RACE</i>	39
3.15) <i>Expressão transiente do cassete EGLV1:GUS em E. grandis e ensaios histoquímicos da atividade GUS</i>	40
4) RESULTADOS	41
4.1) <i>Validação biológica dos candidatos com expressão órgão-específica</i>	41
4.2) <i>Isolamento das regiões promotoras</i>	48
4.3) <i>Sequenciamento das regiões amplificadas por GW</i>	51
4.4) <i>Correlação das seqüências amplificadas com os ESTs</i>	52
4.5) <i>Construção do cassete de expressão para análise funcional em tabaco</i>	54
4.6) <i>Transformação estável em tabaco</i>	54
4.7) <i>Caracterização molecular dos transformantes</i>	56
4.8) <i>Ensaio histoquímico de GUS nas plantas EGLV1:GUS</i>	58
4.9) <i>Análise da expressão relativa do gene repórter nas plantas EGLV1:GUS</i>	60
4.10) <i>Expressão transiente em E. grandis e ensaio histoquímico de GUS</i>	61
4.11) <i>Análise da expressão relativa empregando qPCR</i>	63
4.12) <i>Determinação da região 5' não traduzida dos candidatos EGLV1 e EGFB1 usando 5'- RACE</i>	66
4.13) <i>Elementos cis-regulatórios presentes no promotor EGLV1</i>	67
5) DISCUSSÃO.....	69
5.1) <i>Validação por RT-PCR e qPCR</i>	69
5.2) <i>O candidato EGLV1</i>	70
5.3) <i>Os elementos regulatórios presentes no promotor EGLV1</i>	76
5.4) <i>O candidato EGRT2</i>	78
6) CONCLUSÃO	80
7) PERSPECTIVAS.....	81
8) ANEXOS	82
9) REFERÊNCIAS	90

1) Introdução

1.1) O eucalipto

O eucalipto é uma das principais fontes de celulose no mundo. O Brasil, onde existem 220 empresas florestais com 225 unidades industriais distribuídas em 17 estados, destaca-se no ranking mundial alcançando o 6º lugar como produtor de celulose e o 11º como produtor de papel (Bracelpa, 2007). Com 1,7 milhão de hectares de área plantada para fins industriais, o setor produziu 11,9 milhões de toneladas de celulose e 9 milhões de toneladas de papel com exportações da ordem de 4,7 bilhões de dólares gerando 110 mil empregos diretos e 500 mil empregos indiretos (Bracelpa, 2007).

O gênero *Eucalyptus*, da família Myrtaceae, originário da Austrália, foi introduzido no Brasil por volta do final do século XIX e início do XX. O eucalipto, com destaque especial para a espécie *E. grandis* que possui crescimento acelerado no cultivo, se adaptou muito bem ao clima e solo brasileiros. O crescimento do eucalipto é considerado rápido (média 45 m³/ha/ano – plantio clonal), chegando à fase de corte entre 5 a 7 anos (Leite, 1998). Além disso, o eucalipto possui uma madeira com boa qualidade de fibras e de polpa.

Em contraste às espécies anuais cultivadas (como milho, soja), as espécies florestais, como o eucalipto, são consideradas semidomesticadas devido ao seu longo ciclo reprodutivo. Atualmente, os mais avançados programas de melhoramento de espécies coníferas encontram-se em sua quarta geração, o mesmo acontecendo com o programa do eucalipto (Séguin *et al.*, 1998). O melhoramento de árvores é, portanto, um processo normalmente longo e limitado pela taxa de crescimento das espécies envolvidas. Nesse cenário, a engenharia genética, auxiliada pelas metodologias de

cultura de tecidos que permitem a produção de um grande número de propágulos, pode acelerar esse processo pela transferência direta de características simples. Nesse contexto, quanto mais conhecimento sobre a biologia molecular das árvores for adquirido, maiores serão as oportunidades para seu melhoramento (Séguin, *et al.*, 1998).

A identificação e caracterização de genes têm viabilizado o melhoramento de várias espécies vegetais através de seleção assistida, identificação de QTLs e produção de plantas transgênicas. A produção de plantas geneticamente modificadas tem um grande potencial, uma vez que tem permitido o melhoramento de genótipos selecionados por métodos convencionais, pela introdução de um ou poucos genes que, em muitos casos, são encontrados em espécies distantes. Tais genes não poderiam ser transferidos via recombinação, o que obrigatoriamente acontece entre indivíduos sexualmente compatíveis. As características melhoradas são muitas e incluem o aumento da tolerância a estresses bióticos (Kazan *et al.*, 1998; Kanzaki *et al.*, 2002) e abióticos (Serrano *et al.*, 1999; Shen *et al.*, 2002).

A primeira árvore transgênica foi produzida em 1987, introduzindo-se o gene *aroA* em *Populus* para conferir resistência ao herbicida glifosato (Fillatti *et al.*, 1987). Outros genes de resistência a outros herbicidas, a patógenos, e genes relacionados ao controle de florescimento e qualidade de madeira foram introduzidos também em *Populus* (revisão em Merkle e Nairn, 2005). Recentemente, Li e colaboradores (2003) reportaram a produção de *Populus* transgênico co-transformado com as seqüências anti-senso do gene que codifica a 4-cumarato-CoA ligase e senso do gene que codifica a coniferaldeído 5-hidroxilase, obtendo uma linhagem com baixo teor de lignina (52% menor), com uma razão siringil/guaiacil considerada alta (64%), e com 30% a mais de celulose em relação às árvores não transformadas. A introdução de mais de uma característica em uma árvore transgênica demonstrou ser possível permitindo a

manipulação concomitante da qualidade e quantidade de lignina, assim como do conteúdo de celulose (Merkler e Nairn, 2005).

No caso do eucalipto, os avanços iniciais se deram na área de propagação *in vitro*, procurando o desenvolvimento de sistemas de regeneração via organogênese e embriogênese (revisão em Merkle e Nairn, 2005) visando fornecer material-alvo necessário para manipulação genética. Apesar dos avanços de técnicas de transformação estável como, por exemplo, via biobalística (Sartoretto *et al.*, 2002) e transformação mediada por *Agrobacterium* assistida por sonicação - SAAT (González *et al.*, 2002), poucos foram os relatos de transformação estável e regeneração nessa espécie. O primeiro relato de transformação estável demonstrou que *E. camaldulensis* poderia ser transformado via co-cultivação de explantes foliares de culturas *in vitro*, de plantas (Mullins *et al.*, 1997) ou de hipocótilo de plântulas (Ho *et al.*, 1998), com *Agrobacterium tumefaciens*. Mais recentemente, um híbrido de *E. grandis* X *E. urophylla* foi transformado com a seqüência antisense do gene que codifica a álcool cinâmico desidrogenase (CAD) de *E. gunni*, e algumas plantas regeneradas apresentaram uma atividade CAD 78% menor (Tournier *et al.*, 2003).

Os projetos genoma têm sido de extrema importância para a identificação de genes de interesse para o setor agrícola e florestal. No Brasil, duas iniciativas construíram bancos de seqüências expressas (ESTs) a partir de eucalipto: o projeto FORESTs, com 123.889 ESTs; e o projeto GENOLYPTUS com aproximadamente 120.000 ESTs. Ambas as iniciativas contaram com participação do governo, empresas de papel e celulose e de universidades/institutos de pesquisa.

Sabe-se, entretanto, que somente a caracterização de um gene não garante sua utilidade na obtenção de transgênicos. A produção eficiente de uma proteína heteróloga em determinado organismo depende da obtenção de altos níveis de transcrição do gene

introduzido e, para isso, promotores altamente ativos são necessários (Rance *et al.*, 2002).

Numa breve consulta em bancos de patentes de promotores, verificou-se que as patentes relacionadas ao gênero *Eucalyptus* se concentram em poucos promotores, os quais modulam a expressão preferencial em madeira, como o promotor do gene que codifica uma celulose sintetase (WO2007/141705), e o promotor do gene que codifica a cinamil álcool desidrogenase (US5633439); ou ainda expressão tecido-vascular, como o promotor do gene *COMT* (US20040146904). Na maioria dos casos, os promotores utilizados e patenteados são provenientes de outras espécies florestais, como por exemplo, o *Populus* cujo genoma está completamente seqüenciado. A caracterização de promotores órgão/tecido específicos representa, portanto, um tema relevante de pesquisa visando disponibilizar novas ferramentas para a manipulação genética em eucalipto.

1.2) Regulação transcricional

1.2.1) Transcrição em eucariontes

Em eucariontes existem diversos mecanismos e fatores que regulam a expressão gênica, dentre eles é possível destacar: condensação da cromatina, metilação do DNA, iniciação transcricional, “splicing” alternativo de RNA, estabilidade do RNA mensageiro, controles traducionais, diversas formas de modificações pós-traducionais, tráfego intracelular, degradação de proteína (descritos em Lewin, 2000), e a ação de

microRNAs (Chen e Rajewsky, 2007). Para cada gene eucariótico, a iniciação da transcrição parece ser o determinante primário de controle da expressão gênica (Wray *et al.*, 2003). Nesse contexto, a formação de um complexo de pré-iniciação no tempo certo, e no promotor correto, constitui o pré-requisito básico para a execução correta da síntese de RNA (Hochheimer *et al.*, 2003).

O início da transcrição de um gene em eucariontes é orquestrado por uma série de etapas, que dependem do arranjo correto de complexos protéicos (Figura 01), podendo o mesmo ser dividido nas seguintes etapas: ação do complexo de remodelamento da cromatina expondo a região do gene a ser transcrita; ação de proteínas que mantêm a curvatura do DNA necessária (fatores de “looping”) para a aproximação de regiões *cis*-regulatórias mais distantes à região de formação do complexo de transcrição. No caso dos genes transcritos pela RNA polimerase II ocorre a ligação da proteína de ligação ao TATA-box (TBP) na sequência denominada TATA-box, o qual está presente no promotor-núcleo (ver item 1.2.2); a TBP é parte integrante de um complexo chamado TFIID (fator de transcrição geral da RNA polimerase II); em seguida ocorre a montagem do complexo de pré-iniciação da transcrição (PIC) através da associação de diferentes fatores gerais de transcrição (GTFs), de outros fatores associados a TBP (TAFs), e do complexo da holoenzima RNA polimerase II; após a fosforilação da cauda carboxílica (CTD) da subunidade β da RNA polimerase II ocorre a liberação dos GTFs do complexo e o início da transcrição se dá cerca de 30 pb à frente do TATA-box. À medida que o pré-RNA mensageiro é transcrito, ele sofre modificações co-transcricionais como a adição de um “cap” na terminação 5’, edição dos íntrons (splicing) e ligação dos éxons, através de proteínas que se ligam ao CTD fosforilado (Griffiths *et al.*, 2006).



Figura 01 – Esquema do início de transcrição de um gene idealizado. TBP – Proteína de ligação ao TATA-box; TAFs – Fatores de transcrição associados à proteína TBP (Wray *et al.*, 2003).

Na ausência de elementos *cis*-regulatórios e de proteínas capazes de direcionar a transcrição de um gene, este estará invisível à seleção natural. Todo gene com impacto fenotípico é flanqueado por elementos regulatórios, que em conjunto com a expressão e atividade de proteínas codificadas em outros loci, modula em que nível, tempo, condições ambientais, e em quais células ou tecidos a expressão ocorre (Wray *et al.*, 2003).

Num organismo eucariótico somente alguns genes são transcritos constitutivamente, e mesmo esses em condições de estresses podem ser modulados negativamente. A proporção de genes transcritos oscila consideravelmente durante o ciclo de vida, entre os tipos celulares, e em resposta às flutuações fisiológicas e condições ambientais diversas, como demonstrado nos trabalhos de White *et al.* (1999) em *Drosophila*; de Iyer *et al.* (2001) em levedura; de Kayo *et al.* (2001) em macaco rhesus; de Mody *et al.* (2001) em camundongo; e Ahn *et al.* (2004) em *Arabidopsis*.

1.2.2) Promotores em eucariontes

O promotor é o processador central da regulação de um gene uma vez que contém os sítios de ligação para os complexos protéicos das RNA polimerases responsáveis pela transcrição gênica (Griffiths *et al.*, 2000). Um promotor possui uma região proximal denominada promotor-núcleo que está situada próxima ao sítio de início da transcrição.

O promotor-núcleo é a mínima região de DNA contíguo suficiente para direcionar precisamente o início da transcrição através da maquinaria da RNA polimerase II. Em geral, o promotor-núcleo pode se localizar numa região que se estende por 35 nucleotídeos tanto na direção posterior e anterior ao sítio de início de transcrição. Há vários elementos conservados num promotor-núcleo, como o TATA-box descrito anteriormente, o iniciador (Inr), o elemento de reconhecimento ao TFIIB (BRE) e o elemento do promotor posterior (DPE), mas, em muitos outros casos não há conservação dessas regiões, de modo que aparentemente não existem elementos universais no promotor-núcleo (Butler e Kadonaga, 2002).

O TATA-box foi o primeiro elemento de promotor-núcleo identificado em eucariontes (Goldberg, 1979; Breathnach e Chambon, 1981) e se localiza tipicamente entre 25-30 nucleotídeos acima do ponto de início de transcrição. A sequência consenso se apresenta como TATAAA, muito embora, variações de um ou dois nucleotídeos em relação à sequência consenso, e de posição relativa ao sítio de início de transcrição, sejam observados. Esse elemento se assemelha à região -10 (Pribnow box) dos promotores de procariontes (Pribnow, 1975a,b), mas acredita-se que não sejam homólogos (Butler e Kadonaga, 2002).

O elemento Inr geralmente engloba o sítio de início de transcrição, sendo encontrado tanto em promotores que contém TATA-box como os que não o contém. A

seqüência consenso em células de mamíferos é Py-Py(C)-A₊₁-N-T/A-Py-Py (Corden *et al.*, 1980; Bucher, 1990; Javahery *et al.*, 1994; Lo e Smale, 1996; Smale *et al.*, 1998), onde **Py** é uma base pirimidínica, () indica preferência pela base nitrogenada, / uma base ou a outra, **N** qualquer base e **A₊₁** designa a posição +1 do início da transcrição, ou seja, o ponto onde a transcrição comumente se inicia (Butler e Kadonaga, 2002). Em *Drosophila*, o consenso de Inr é T-C-A₊₁-G/T-T-C/T (Hultmark *et al.*, 1986; Purnell *et al.*, 1994; Arkhipova, 1995; Kutach e Kadonaga, 2000).

O elemento DPE foi identificado como sendo um sítio posterior do promotor-núcleo para TFIID purificada de *Drosophila*. O complexo TFIID se liga cooperativamente aos elementos Inr e DPE, sendo que tanto mutações em Inr e DPE causam a perda de afinidade pelo TFIID (Burke e Kadonaga, 1996). O DPE está comumente presente em promotores sem TATA-box, e se encontra precisamente na posição +28 a +32 em relação à posição **A₊₁** do elemento Inr, sendo que a alteração de um único nucleotídeo de espaçamento entre Inr e DPE causa redução na afinidade de ligação do TFIID e da atividade transcricional basal (Burke e Kadonaga, 1997; Kutach e Kadonaga, 2000). A seqüência consenso para DPE é A/G₊₂₈-G-A/T-C/T-G/A/C, sendo que há uma menor preferência para G em +24 (Kutach e Kadonaga, 2000). Convém denotar que do ponto de vista funcional, o DPE é sempre dependente de Inr, e a seqüência consenso de DPE é sempre relativa à posição **A₊₁** de Inr.

O BRE se localiza imediatamente anterior a alguns TATA-box, e é um sítio de ligação ao TFIIB (Lagrange *et al.*, 1998). A seqüência consenso de BRE é G/C-G/C-G/A-C-G-C-C (onde o C em posição 3' do BRE é seguido por T na posição 5' do TATA-box). Análises de transcrição *in vitro* indicam que BRE facilita a incorporação de TFIIB em complexos de iniciação de transcrição funcionais (Lagrange *et al.*, 1998).

O promotor-núcleo não pode ser considerado um elemento passivo tendo como função única direcionar o acoplamento correto da maquinaria de transcrição da RNA polimerase II. Esse atua também como uma região de elementos regulatórios *cis*-atuantes que determina a especificidade para um complexo transcricional distinto e para outros elementos regulatórios mais distais, como os acentuadores (enhancers) (Butler e Kadonaga, 2002).

Em *Saccharomyces cerevisiae*, o promotor *his3* contém dois elementos TATA-box, denominados T_C (TATA constitutivo) e T_R (TATA regulatório), que direcionam o início de transcrição em diferentes pontos (Struhl 1986, 1987). O elemento T_C, constituído por cerca de 30 nucleotídeos, se localiza anteriormente ao T_R e não contém a seqüência canônica TATAAA de T_R, mas sim uma região rica em AT, sendo usado preferencialmente em transcrições de baixos níveis. Porém, após a indução do gene *his3*, há um aumento da transcrição relacionada a T_R, mas não a T_C. Adicionalmente foram identificados fatores de transcrição necessários para a transcrição dependente de T_C mas não de T_R (Moqtaderi *et al.*, 1996, 1998), e fatores que preferencialmente inibem a transcrição dependente de T_C (Collart e Struhl, 1994). Isso demonstra que cada elemento TATA recruta uma composição diferente de complexo protéico.

Em *Drosophila* foi verificado o fenômeno de especificidade promotor-acentuador em experimentos de competição de promotores, onde o acentuador é posicionado entre (ou na vizinhança) de dois promotores distintos. Os acentuadores AE1 e IAB5 em *Drosophila* ativam preferencialmente a transcrição no promotor-núcleo “even-skipped” que contém o elemento TATA-box em relação ao promotor-núcleo “white” sem TATA-box (com o DPE), demonstrando uma alta afinidade de ligação desses ao promotor-núcleo com TATA-box. Na ausência do promotor-núcleo “even-

skipped”, os mesmos acentuadores podem ativar o outro promotor (Ohtsuki *et al.*, 1998).

A especificidade do promotor-núcleo a fatores de transcrição e acentuadores distintos define um alto grau de controle transcricional que não pode ser relevado em estudos experimentais. Um acentuador deve ser estudado em conjunto com o seu promotor-núcleo cognato para averiguar sua especificidade; além disso, a presença de cada elemento num promotor-núcleo deve ser examinada separadamente, e em conjunto, para determinar sua função na transcrição (Butler e Kadonaga, 2002).

De maneira geral, uma transcrição baseada somente no promotor-núcleo produzirá baixas taxas. Funcionalmente, para se atingir taxas de transcrição significativas são necessárias associações específicas de fatores de transcrição às seqüências regulatórias presentes fora do promotor-núcleo (Weinzierl, 1999; Carey e Smale 2000; Lemon e Tjian 2000).

Há uma grande variabilidade de composição e organização de sítios de ligação a fatores de transcrição presentes nos genes eucariotos (Figura 02; Wray *et al.*, 2003). A arquitetura de sítios de ligação a fatores de transcrição demonstra que esses sítios podem estar localizados em uma região simples e discreta próxima ao sítio de início de transcrição de um gene (Figura 02 A-E), ou então dispersos em vários grupos distintos, altamente modulares (Figura 02 I, K, L e P). Em geral, a distância física dos elementos *cis*-regulatórios varia de algumas centenas de pares de bases a até mais que 100 kb (Figura 02 Q). Um exemplo extremo de dispersão física é o módulo regulatório do locus *Shh* em humanos e camundongos que se encontra a aproximadamente 800 kb de distância do início de transcrição (Lettice *et al.*, 2002). O posicionamento dos sítios *cis*-regulatórios em relação ao sítio de início de transcrição é bem variável, de poucas quilobases flanqueando anteriormente o promotor-núcleo (Figura 02 A-H, J, M, N),

dentro de regiões não transcritas 5' (Figura 02 P, módulo 1), em íntrons (Figura 02 K, módulos 3 e 4; L, módulos 7-11), nas regiões não traduzidas 3' (Figura 02 Q, módulos 1-5), e em raros casos, dentro de éxons, como demonstrado para o gene *keratin 18* em humanos (Neznanov *et al.*, 1997) e para o gene *nonA* em *Drosophila* (Sandrelli *et al.*, 2001). Em alguns casos determinadas regiões regulatórias influenciam a transcrição em mais de um gene (Figura 02 J, M, N, O), com origem filogenética comum (como parálogos; Figura 02 M, N: Yp1/Yp2 e Dlx6/Dlx4, respectivamente) ou não (Figura 02 J, M: GAL10/GAL1 e APOC3/APOA1, respectivamente). Essa diversidade de posições e interações só é possível graças à capacidade do DNA de se dobrar, auxiliando a interação entre proteínas associadas ao DNA em sítios de ligação distais. Sítios *cis*-regulatórios podem se situar até no lado mais distal de um locus adjacente (Figura 02 O). É importante ressaltar que genes que possuem padrões similares de expressão nem sempre apresentam similaridade de organização de seus promotores. Como exemplo disso é possível citar a expressão dos genes *even-skipped* e *ftz*, que produzem o padrão de 7 faixas no embrião de *Drosophila* (Figura 02 I: módulos 3-7, 10 e 11; P: módulo 2), mas não compartilham os mesmos elementos regulatórios.

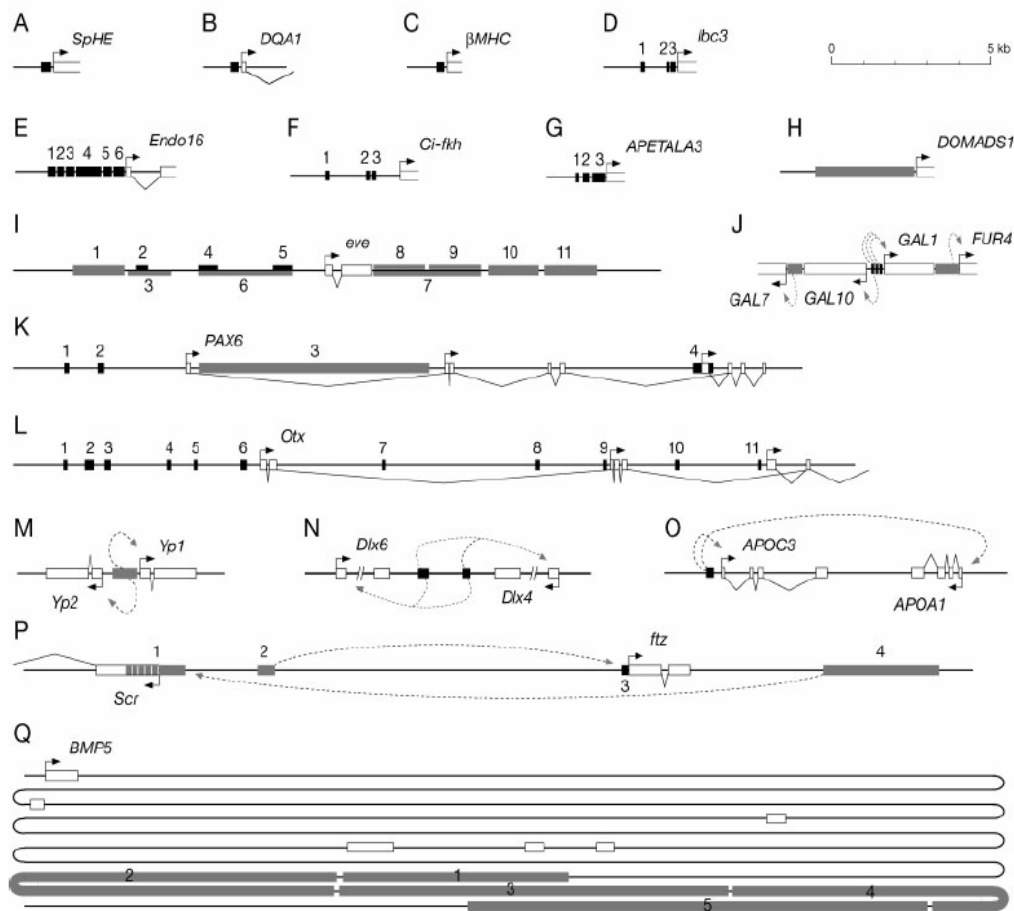


Figura 02 – Quadro representativo de tipos conhecidos de arquitetura de elementos cis-regulatórios em eucariotes (Wray *et al.*, 2003). A escala dos mapas dos loci está representada no canto superior direito. Quadros em preto – regiões regulatórias precisamente mapeadas; Quadros em cinza – regiões regulatórias que não foram precisamente mapeadas; Quadros em branco – éxons; Flecha dobrada – Início de transcrição; Números – Regiões regulatórias distintas em que as contribuições no perfil transcricional total foram determinadas experimentalmente; Linhas pontilhadas indicam interações entre um módulo e outros loci, sendo esses às vezes não adjacentes. Loci, produto gênico, táxon e referência ver anexo 8.1.

A informação transcricional de um promotor não pode somente ser analisada linearmente em função da presença ou ausência de sítios de ligação. A posição relativa, orientação, e as seqüências de nucleotídeos desses sítios, assim como os perfis de expressão dos fatores de transcrição e co-fatores cognatos, todos contribuem para

determinar o perfil de transcrição total de um gene, pois são interações complexas, não-lineares, e frequentemente contexto-dependentes. Sendo assim, inúmeros motivos levam os diversos sítios de ligação a fatores de transcrição a não serem funcionais ou não influenciarem na transcrição de um gene. Dentre os motivos podemos destacar: a ausência de um fator de transcrição; a condensação da cromatina local, tendo ou não o fator de transcrição; a ocupação do sítio de ligação adjacente, mascarando o sítio de interesse; o fator de transcrição na forma inativa; maior afinidade de uma proteína alternativa ao sítio; a diversidade de interação protéica entre co-fatores e fatores de transcrição permitindo ou não a ligação a determinado sítio (Wray *et al.*, 2003).

1.2.3) Promotores empregados em plantas

Atualmente, os promotores mais utilizados em plantas geneticamente modificadas, dicotiledôneas em particular, são o promotor 35S do vírus do mosaico da couve flor (CaMV) e os promotores dos genes da nopalina e octopina de *Agrobacterium tumefaciens*. Apesar dos grandes avanços que têm sido obtidos com o emprego de tais promotores, os padrões de expressão dos transgenes são variados e baixos em alguns casos (Zheng & Murai, 1997; Green *et al.*, 2002) não havendo por outro lado garantia de expressão no órgão/tecido desejado (Neuteboom *et al.*, 2002).

Além disso, o uso de promotores ubíquos e constitutivos como os citados acima é uma das causas da grande preocupação a respeito dos efeitos dos transgênicos no meio ambiente. Promotores como o CaMV 35S quando ligados a genes utilizados para seleção de transformantes (genes de resistência a antibióticos) determinam, em geral, a expressão do produto gênico em todos os órgãos/tecidos da planta. Tal característica,

desnecessária e indesejável em vegetais geneticamente modificados, tende a diminuir a aceitação dos produtos derivados pelos consumidores. Nesse contexto, promotores com padrão de expressão órgão-específico seriam de grande importância e alguns estudos com esse propósito já vem sendo conduzidos. Por exemplo, Huang e colaboradores (2001) desenvolveram, a partir da identificação de um promotor calo-específico, um método de seleção de plantas transformadas de arroz que certamente não expressarão o gene de resistência ao antibiótico em outros órgãos/tecidos.

O conhecimento de promotores também é importante já que se espera que, ao menos em parte, semelhanças nas regiões envolvidas na regulação da transcrição determinem padrões similares de expressão em diferentes espécies (Ohler e Niemann, 2001). Lauvergeat e colaboradores (2002) demonstraram que o padrão de expressão vascular determinado pelo promotor do gene *EgCAD2*, que codifica a cinamil álcool desidrogenase em *Eucalyptus gunnii*, se conserva em tabaco e *Populus*. Já o trabalho de Ayre *et al.* (2003), empregando o promotor do gene que codifica a galactinol sintetase de *Cucumis melo* (específico de canais das menores nervuras da folha), mostrou que o padrão de expressão tecido-específico é conservado em *Arabidopsis* e tabaco, apesar de nenhuma das duas espécies empregarem galactinol no processo de transporte via floema. Tais resultados representam exemplos de conservação de processos regulatórios em diferentes espécies.

A identificação de regiões promotoras com padrões definidos de expressão pode ser realizada por diferentes métodos. O ponto comum a esses métodos é a identificação de genes com expressão em dado órgão ou tecido, processo que poder ser realizado, por exemplo, empregando-se as informações disponíveis em um banco de seqüências expressas (EST), e posteriormente técnicas para verificação de expressão diferencial, como o “Northern Blot”, a Transcrição reversa (RT)-PCR, ou o PCR em tempo real. Os

cDNAs selecionados são então utilizados na identificação de seqüências genômicas, contendo as regiões regulatórias de interesse. Essa identificação pode ser realizada a partir do “screening” de uma biblioteca genômica ou utilizando-se estratégias de “genome walking”. Devic *et al.* (1997) desenvolveram uma metodologia bastante eficiente para este fim que consiste dos seguintes passos: isolamento de DNA genômico; digestão isolada com diferentes enzimas de restrição de corte abrupto; ligação de adaptadores aos fragmentos de restrição resultantes; uso desses fragmentos mais os adaptadores como moldes para ampliações via PCR utilizando-se um oligonucleotídeo de seqüência complementar ao adaptador e um complementar a um dado gene, cuja expressão específica a dado órgão e a seqüência de bases foram previamente verificadas; clonagem e sequenciamento dos fragmentos.

Do ponto de vista acadêmico, a identificação e análise de promotores órgão-específicos colaboram para um aprofundamento dos estudos sobre os mecanismos moleculares envolvidos no processo de expressão gênica e de diferenciação tecidual. No enfoque prático, a disponibilidade de tais promotores facilita a obtenção de cassetes de expressão que permitam a expressão dos genes de interesse nos tecidos alvos em que esses promotores estejam ativos, possibilitando assim um aprimoramento dos processos biotecnológicos. O uso de promotores com padrão de expressão órgão/tecido-específico tende a diminuir os problemas causados pela expressão constitutiva/ubíqua de transgenes que pode gerar, por exemplo, efeitos pleiotrópicos indesejados. Nesse contexto, o presente estudo visou explorar as informações do banco de ESTs do Eucalipto (FORESTs) visando identificar genes com padrão de expressão órgão/tecido-específico para posterior obtenção de suas seqüências promotoras.

2) Objetivos

Os objetivos gerais do presente trabalho foram:

- Identificar genes com perfil de expressão órgão-específico em eucalipto.
- Isolar e caracterizar funcionalmente as seqüências promotoras adjacentes à região codificadora dos genes caracterizados.

3) Material e Métodos

3.1) Identificação dos ESTs com expressão órgão-específica

Os genes candidatos com padrão de expressão órgão-específico foram selecionados de duas maneiras. Inicialmente, a partir de predições *in silico* utilizando-se de um banco de “Northern” eletrônico (<http://ipe.cbmeg.unicamp.br/bioinfo/forestsdb.html>) construído com base nos 33.080 contigs depositados no banco FORESTs (<https://forests.esalq.usp.br/>). Adicionalmente, foram utilizados dados coletados na literatura contendo a descrição de genes com padrão de expressão órgão-específico em diferentes espécies vegetais. Os ESTs selecionados desta maneira foram: EGUTST2032C06.g (madeira - Paux *et al.*, 2004), EGMCRT3145G04.g (raiz - Ahn *et al.*, 2004) e EGEPRT3362H10.g (raiz - Chaumont *et*

al., 2001), EGRFCL1206E01.g (madeira - Krauskopf *et al.*, 2005), EGQHSL1102C02.g (caule – Ludwikow *et al.* 2004), EGQHLV2242E10.g (folha - Ludwikow *et al.* 2004), EGBMRT6276A06.g (raiz - Ludwikow *et al.* 2004), EGEZRT6274D07.g (raiz - Ludwikow *et al.* 2004), EGBGBK1048G08.g (madeira - Ludwikow *et al.* 2004), EGCBRT6017A01.g (raiz - Ahn *et al.*,2004), EGEQRT3102B06.g (raiz - Okamoto *et al.*, 2006) e EGCCST6215A11.g (madeira - Moreau *et al.*, 2005).

O banco FORESTs é constituído por ESTs oriundos do sequenciamento de bibliotecas de cDNAs construídas utilizando RNA mensageiro proveniente de vários tecidos em condições normais e sob estresses bióticos e abióticos. Cabe ressaltar que para comparar os níveis de expressão nas diferentes bibliotecas, os dados foram normalizados pelo número de *reads* em cada biblioteca. As análises dos perfis de expressão foram realizadas considerando cada biblioteca individualmente ou pools onde bibliotecas de órgãos/tecidos similares foram agrupadas. Adicionalmente, a fim de aumentar a confiabilidade dos dados gerados pela análise *in silico*, seis testes estatísticos robustos foram aplicados na tentativa de minimizar falso-positivos (Fisher 2X2, 2X2 χ^2 e AC para comparação de pares, ou seja, comparação de 2 bibliotecas ou 2 pools e R, GT e χ^2 para testes de múltipla comparação, que permite comparar todas bibliotecas ou pools). A descrição detalhada dos métodos empregados para confecção do “northern” eletrônico pode ser encontrada em Vicentini *et al.* (2005).

Todos os contigs selecionados (via predição *in silico* ou pela literatura) foram analisados quanto a sua similaridade no banco de dados GenBank do *National Center for Biotechnology Information* (NCBI; <http://www.ncbi.nlm.nih.gov>), utilizando as seqüências dos *reads* que melhor os representavam, considerando a qualidade e extensão das mesmas. Essa análise foi realizada com auxílio da ferramenta BLAST (Altschul *et al.*, 1990) adotando como critério de seleção o valor de E-value menor ou igual a e-5.

Após a fase de seleção *in silico*, com intuito de validar os candidatos selecionados, pares de oligonucleotídeos foram desenhados com base nas seqüências dos ESTs considerados órgão-específicos com o auxílio dos programas computacionais Primer3 (Rozen & Skaletsky, 2000) e GeneRunner (Tabela 1). Para tal, os seguintes critérios foram adotados: tamanho do oligonucleotídeo (mínimo: 18 bases; ótimo: 20 bases; máximo: 25 bases); temperatura de anelamento (mínimo: 57°C; ótimo: 60°C; máximo 66°C); e conteúdo em GC (mínimo: 40%; ótimo: 50%; máximo: 60%).

3.2) Extração de RNA total de eucalipto e tabaco

O RNA total proveniente de diferentes órgãos de *Eucalyptus grandis* empregado na validação dos ESTs selecionados nas análises *in silico* foi extraído através da metodologia descrita por Korimbocus *et al.* (2002) (baseado em Chang *et al.*, 1993) com modificações. Foram utilizados aproximadamente 100 mg de tecido macerado em nitrogênio líquido com posterior solubilização em 500 µl do tampão de extração CTAB pré-aquecido (Anexo 8.2) e incubação por cerca de 10 a 15 min a 65°C. Adicionou-se em seguida 500 µl de CIA (clorofórmio:álcool isoamílico - 24:1) e homogeneizou-se a solução. Centrifugou-se a 12000g por 10 min à temperatura ambiente e coletou-se o sobrenadante, aplicando-se novamente o passo anterior de CIA seguido de uma nova centrifugação. Um volume de cloreto de lítio (LiCl 5M) foi adicionado ao volume de sobrenadante coletado e o RNA total foi precipitado a 4°C por no mínimo 1 hora com posterior centrifugação a 12000g por 30 min (4°C). O sedimento foi lavado em 200 µl de TE-SDS 1%, 100 µl de NaCl 5 M e 300 µl de isopropanol. A amostra foi precipitada durante 30 min a -20°C e centrifugada a 12000g por 20 min (4°C). Uma lavagem foi

realizada com 1 ml de etanol 70% com posterior centrifugação a 7500g por 5 min (4°C), sendo o precipitado ressuspenso em H₂O Milli-Q autoclavada (tratada com Dietilpirocarbonato - DEPC). Alíquotas das amostras foram misturadas ao tampão de carregamento e avaliadas quanto à sua integridade em gel de agarose desnaturante 1% corado com brometo de etídio (Anexo 8.3).

Os órgãos extraídos foram separados em dois grupos de estudo: o primeiro objetivando a validação do perfil de expressão via RT-PCR e o segundo para a avaliação da expressão relativa por PCR em tempo real (qPCR). O primeiro grupo consistiu de folha de eucalipto adulto, raiz de muda, caule de muda, feixe cambial de tronco, botão floral e fruto. O segundo grupo foi constituído, respectivamente, de folhas e caules de eucaliptos com 6 anos e 1 mês (72 meses), e com 2 anos e 7 meses (31 meses); folhas, caules e raízes de eucaliptos com 6 meses, plântulas de cultura *in vitro* (1 mês), e de propagação vegetativa *in vitro* (40 dias); botões florais e frutos.

Visando avaliar o padrão de expressão relativa do gene *uidA* em plantas transgênicas de tabaco foram utilizadas, para a extração de RNA total, amostras de folha, caule e raiz de plantas de 4 meses de idade, utilizando o protocolo de extração de CTAB citado acima.

3.3) Validação biológica por RT-PCR e qPCR

3.3.1) RT-PCR

Para as análises empregando RT-PCR, alíquotas de 2 µg de RNA total extraído do primeiro grupo de amostras foram tratadas com DNase I segundo as recomendações do fabricante (MBI Fermentas) visando degradar todo DNA genômico presente nas mesmas. Após o tratamento com DNase I, cada alíquota de RNA foi submetida à transcrição reversa empregando-se oligo-dTVN17 e a enzima “SuperscriptTM III Reverse Transcriptase” segundo as recomendações do fabricante (Invitrogen). Os cDNAs obtidos foram então submetidos à amplificação por PCR, empregando-se para tal oligonucleotídeos específicos aos ESTs escolhidos. As seguintes etapas foram adotadas: desnaturação inicial à 94°C por 3 minutos, seguida de 40 ciclos de 94°C por 30 segundos, 65°C por 45 segundos e 72°C por 1 minuto, com uma etapa posterior de extensão 72°C por 10 min. A especificidade a um dado órgão foi confirmada analisando-se os produtos gerados por esta amplificação em gel de agarose 1,5 % corado com brometo de etídio. Procurou-se nesta etapa avaliar a presença ou ausência de bandas correspondentes aos transcritos do gene candidato em cada órgão.

3.3.2) PCR quantitativa em tempo real (qPCR)

As análises de qPCR foram realizadas em duas fases do projeto: a primeira usando os genes validados como órgão-específicos na etapa de RT-PCR, com o objetivo

de mensurar a expressão relativa dos mesmos em diferentes órgãos de eucalipto. A segunda foi realizada visando analisar a expressão do gene *uidA* (GUS), sob controle do promotor putativo de eucalipto, em diferentes órgãos de tabaco transgênico. Por ser uma técnica mais sensível que a RT-PCR, a qPCR é capaz de detectar transcritos normalmente não detectáveis pela RT-PCR. Para tal, alíquotas de 1 µg de RNA total extraído dos órgãos do segundo grupo de extração também foram tratados com DNase I e submetido à transcrição reversa como descrito. Com auxílio do espectrofotômetro (Nanodrop), as amostras de cDNAs foram quantificadas e ajustadas a 60 ng por reação. O reagente utilizado na reação foi o *SYBR Green Master Mix* (Applied Biosystems) e os oligonucleotídeos gene-específicos foram desenhados no programa “Primer Express” (Applied Biosystems). O aparelho *ABI PRISM 7000 Sequence Detection System* (Applied Biosystems) foi utilizado como plataforma para a qPCR.

Os candidatos analisados foram EGEQLV2201B04.g (oligo reverso LV1RT-PCRR1 5' GAGGTCTGCAACGCAGGAGTC 3'; oligo senso LV1RT-PCR1 5' CAAACCTTAATTCTGTTCTCCTTCCT 3') e EGMCRT3145G04.g (oligo reverso EgKt1-RTPCRR 5' AAGAGTGCGCGTGGGACTT 3'; oligo senso 5' GCCGCGTTGACTCTCTTAACC 3') em eucalipto. Adicionalmente, a expressão realtiva do gene repórter *uidA* (GUS) em diferentes órgãos coletados de plantas de tabaco transgênico foi determinada (RT-GUS3F – 5' ATGGTAGATCTGAGGAACCGACG 3' e RT-GUS3R – 5' TTTCGCGATCCAGACTGAATG 3'; o oligonucleotídeo RT-GUS3F foi desenhado na junção éxon e intron).

O método de $2^{-\Delta Ct}$ (Livak e Schmittgen, 2001) foi utilizado para as análises, nas quais a expressão relativa do gene alvo foi calculada com base na expressão de um gene normalizador. No caso de eucalipto foi usado um gene que codifica uma NADP-

isocitrato desidrogenase de *Eucalyptus globulus* [EgIDHF – 5’ CTGCTGGAATCTGGTATGAACA 3’ e EgIDHR – 5’ TCACTCTGGACATCTCCATCA 3’ (Goicoechea *et al.*, 2005)], e no tabaco transgênico foi empregado o gene *18S* [18SF – 5’AGGAATTGACGGAAGGGCA 3’ e 18SR – 5’ GTGCGGCCCCAGAACATCTAAG 3’ (Levy *et al.*, 2004)]. A seguinte equação foi usada para o cálculo da expressão relativa:

$$ER = 2^{-\Delta Ct}$$

Onde:

ER = expressão relativa

$$\Delta Ct = Ct_{\text{gene-alvo}} - Ct_{\text{normalizador}}$$

Ct = “Threshold cycle”, ciclo em que a fluorescência do alvo se torna significativamente maior que o ruído (“background”)

Após a padronização das reações quanto a quantidade de cDNA/reação e quantidade de oligonucleotídeos, a eficiência de reação foi considerada perto de 1,0. Os componentes da reação de amplificação estão listados no anexo 8.4 sendo utilizado um ciclo prévio de 50°C por 2 min e 95°C por 10 min, seguido de um ciclo de desnaturação a 95°C por 15 s, anelamento e extensão a 62°C por 1 min, repetido por quarenta vezes. Após a reação foi adicionado um passo de dissociação através do aquecimento de 60°C

para 95°C, subindo 1°C por mim. Este procedimento serviu para verificar a especificidade dos oligonucleotídeos utilizados ou a formação de “primers-dimers”.

Adicionalmente, os produtos amplificados via qPCR para o candidato EGEQLV2201B04.g nas seguintes amostras de eucalipto: folhas com 6 anos e 1 mês, folhas com 6 meses; caule e raiz de 6 meses, foram clonados. Essa etapa visou confirmar se as seqüências amplificadas correspondiam ao EST validado.

3.4) Isolamento das regiões promotoras

3.4.1) A técnica de *genome walking* (GW)

Para a clonagem das regiões promotoras dos candidatos validados foi utilizado o kit *Universal Genome Walker*[™] (Clontech), e bibliotecas de DNA genômico de *Eucalyptus grandis*. Essas bibliotecas foram constituídas a partir de 4 pools de fragmentos oriundos da digestão de DNA genômico com diferentes enzimas de restrição de corte abrupto (*DraI*, *EcoRV*, *PvuII* e *StuI*). Esses fragmentos foram então ligados a adaptadores fornecidos pelo kit. A técnica “genome walking” permite amplificar, através de PCR, regiões localizadas à montante de uma seqüência conhecida (Devic *et al.*, 1997) empregando-se oligonucleotídeos complementares aos adaptadores e à seqüência conhecida (no nosso caso do EST validado).

A primeira etapa do procedimento de GW consiste de um PCR primário empregando oligonucleotídeos mais externos relacionados ao adaptador (AP1) e à seqüência do EST validado. Uma alíquota diluída (1:50) desse PCR primário é então

usada como molde para a realização do PCR secundário que utiliza oligonucleotídeos mais internos (AP2 e GSP2) à região amplificada (Figura 03).

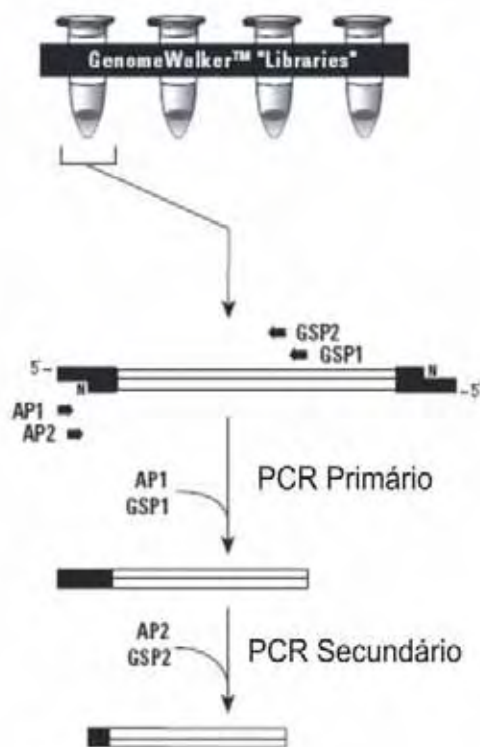


Figura 03 – Representação esquemática da metodologia de *genome walking* (adaptada de Clontech Inc., 1999). As extremidades em preto dos fragmentos de DNA representam os adaptadores; AP1 e AP2 – oligonucleotídeos adaptadores; GSP1 e GSP2 – oligonucleotídeos gene-específicos.

3.4.2) Desenho dos oligonucleotídeos gene-específicos (GSPs)

Oligonucleotídeos complementares à porção 5' terminal da seqüência do EST representativo de cada contig validado foram desenhados com o intuito de amplificar produtos a montante da referida região. Os critérios adotados foram os seguintes: tamanho do oligonucleotídeo (mínimo: 26 bases; ótimo: 28 bases; máximo: 30 bases); temperatura de anelamento (mínimo: 66°C; ótimo: 67°C; máximo 68°C) e conteúdo em GC (mínimo: 40%; máximo: 60%).

3.4.3) Extração de DNA genômico de eucalipto

A extração de DNA genômico foi baseada no protocolo de Ferreira & Grattapaglia (1998). Foram utilizados aproximadamente 150 mg de tecido foliar macerado em nitrogênio líquido com posterior solubilização em 700 µl do tampão de extração CTAB pré-aquecido (Anexo 8.5) e incubação por 30 min a 65°C. Adicionou-se em seguida 600 µl de CIA (clorofórmio-álcool isoamílico na proporção 24:1) seguido de uma centrifugação a 14000g durante 15 min. O sobrenadante foi coletado e ao mesmo foi adicionado 600 µl de CIA, centrifugando a amostra nas mesmas condições anteriores. O sobrenadante foi coletado, e nele foram adicionados 700 µl de isopropanol frio (-20°C). A amostra foi precipitada a -20°C por no mínimo 30 min com subsequente centrifugação a 7500g por 5 min. O precipitado foi lavado com 1 ml de etanol 70% e centrifugado a 7500g por 5 min. Descartou-se o sobrenadante e o precipitado foi ressuspendido com TE (10 mM Tris-HCl, 1 mM EDTA – pH 8.0) contendo 10 µg/ml de

RNAse, e incubado em banho-maria a 37°C por 30 a 120 min para eliminar contaminações com RNA. A quantificação do DNA foi feita em gel de agarose 1%, corado com brometo de etídio.

3.4.4) Construção das bibliotecas de GW e realização da PCR *nested*

Para a construção das bibliotecas foram utilizadas alíquotas de DNA genômico de eucalipto (2,5 µg) digeridas durante 16 horas a 37 °C com as enzimas de restrição *DraI*, *EcoRV*, *PvuII* e *StuI* separadamente. Uma alíquota da reação de digestão (5 µl) foi aplicada em gel de agarose 1% para verificar a qualidade da digestão, sendo o DNA restante purificado via extração clorofórmio/fenol e ligado aos adaptadores. Tanto a purificação quanto a ligação dos adaptadores foram processadas seguindo as instruções do fabricante do kit. Após a ligação dos adaptadores, as amplificações das seqüências promotoras foram realizadas conforme especificações do kit (Anexo 8.6 e Figura 3).

3.5) Clonagem e Sequenciamento dos produtos de amplificação

Os produtos de PCR derivados do GW foram isolados em gel de agarose 1% e purificados. Esses produtos foram clonados no vetor pGEM-Teasy (Promega) e transformados por choque térmico em *Escherichia coli* cepa DH5α (Sambrook *et al.*, 1989). Mini-preparação por lise alcalina (Anexo 8.7) foi utilizada para a extração dos plasmídeos e uma digestão com *EcoRI* foi realizada para verificar a presença dos

fragmentos clonados nos tamanhos corretos, descartando os clones com falsos positivos. Os transformantes validados foram crescidos em 1,5 ml de meio Luria Bertani (LB) líquido, volume ao qual foi adicionado glicerol líquido autoclavado (15%; v/v) para posterior estocagem em freezer -80°C.

A reação de sequenciamento (Anexo 8.8) foi realizada com um kit comercial de “Big Dye” versão 3.1 (Applied Biosystems) e os oligonucleotídeos M13 senso e antisenso. Como molde foram utilizados os DNAs plasmidiais contendo os produtos de PCR derivados do GW. As reações de sequenciamento foram realizadas em seqüenciador automático modelo ABI Prism 377 (Applied Biosystems).

3.6) Confirmação da identidade das seqüências clonadas

A estratégia adotada para a confirmação da identidade das seqüências clonadas baseou-se na realização de uma PCR utilizando um oligonucleotídeo senso desenhado internamente à região seqüenciada e o outro oligonucleotídeo reverso complementar a seqüência do EST validado. Os pares de oligonucleotídeos utilizados para o candidato EGACFB1015B10.g foram: EBFF1 (senso 5'CATGATCCTTGCGGAAATCT 3') e GSP-FB1 (reverso 5' CGTGACAGTAGCAAGACCAACAAAG 3') com amplicon de 255 bp; EBFF2 (senso 5' GGTGCGGCTAATTGCTAAAG 3') e GSP-FB1 com amplicon 287 bp. Já para o candidato EGEQLV2201B04.g foram utilizados o seguinte par de oligonucleotídeos: EGLV2B (5' CAATCTGTCTGCTTCCAAG 3') e GSP-LV1 (5' ATTCACGGTGACTGCTGAGACCTTCTT 3') com amplicon em torno de 300 bp.

Confirmada a correlação dos fragmentos sequenciados com seus respectivos ESTs, as análises em sistema de expressão estável no intuito de confirmar a atividade dos promotores clonados foram iniciadas.

3.7) Construção do cassete de expressão

Os candidatos EGEQLV2201B04.g (específico de folha e nomeado de EGLV1) e EGACFB1015B10.g (específico de botão floral e nomeado de EGFB1) foram selecionados para a análise funcional dos seus promotores em expressão estável em tabaco. Os vetores binários selecionados para tal foram o pCAMBIA-1381z (CAMBIA) e pBI101.1 (Clontech).

Para a construção do cassete de expressão contendo o promotor do candidato EGLV1 foi utilizado o vetor binário pCAMBIA-1381z, que tem um íntron inserido na região codificadora (ORF) do gene *uidA* (GUS). O fragmento correspondente ao promotor (1223 pb) foi amplificado por PCR para adição dos sítios de restrição *Bam*HI (na porção 5' terminal) e *Nco*I (na porção 3' terminal) em suas extremidades (Figura 4A). O sítio *Nco*I permite a inserção correta do promotor em fusão transcricional com o início da ORF do gene GUS no pCAMBIA.

Para a obtenção do cassete de expressão contendo o promotor do candidato EGFB1 foi utilizado o vetor binário pBI101.1 (Clontech). O fragmento promotor (1900 pb) foi amplificado por PCR com a adição dos sítios de restrição *Hind*III (na porção 5' terminal) e *Xba*I (na porção 3' terminal) em suas extremidades (Figura 4B). Nesse caso, a impossibilidade de se usar o vetor pCAMBIA-1381z ocorreu devido à presença de um sítio interno *Nco*I na região 3' terminal da seqüência a ser inserida.

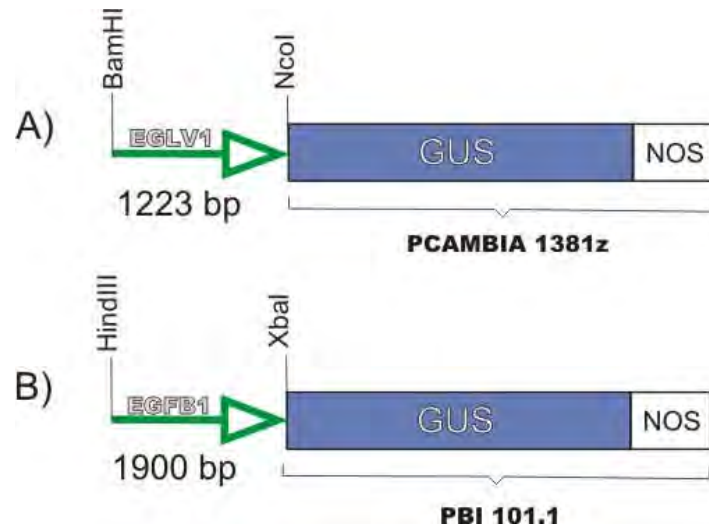


Figura 04 – Representação esquemática dos cassetes de expressão construídos e inseridos nos vetores binários. A) Promotor EGLV1 em fusão com o gene GUS inserido no vetor binário pCAMBIA-1381z; B) Promotor EGFB1 em fusão com o gene GUS inserido no vetor binário pBI 101.1.

3.8) Produção de células competentes de *Agrobacterium tumefaciens*

Para a transformação estável da construção EGLV1:GUS foram produzidas células competentes da linhagem LBA4404 de *Agrobacterium tumefaciens*. Essa linhagem tem genes de resistência à estreptomicina (origem plasmidial) e rifampicina (origem cromossômica) como agentes de seleção. O procedimento utilizado foi baseado no protocolo de Lacorte & Romano (1998; Anexo 8.9). Para manutenção das células

competentes ressuspendidas em CaCl_2 20 mM, adicionou-se glicerol estéril a uma concentração final de 15% (v/v). As células foram homogeneizadas, congeladas em nitrogênio líquido, e mantidas estocadas em freezer -80°C .

3.9) Transformação do vetor binário em células competentes de *A. tumefaciens*

O cassete de expressão EGLV1:GUS inserido no vetor binário pCAMBIA-1381z foi transformado nas células competentes de *A. tumefaciens* linhagem LBA4404. O vetor binário pCAMBIA-1381z possui gene de resistência a canamicina como agente de seleção bacteriano. A transformação seguiu o protocolo de Lacorte & Romano (1998; Anexo 8.10). Para confirmação da transformação, colônias isoladas foram cultivadas em meio LB líquido a 28°C por 20 h, e após a extração dos plasmídeos recombinantes por lise alcalina, os mesmos foram submetidos à reação de PCR utilizando oligonucleotídeos internos, um no sentido senso posicionado junto ao promotor e o outro no sentido reverso na região codificadora do gene GUS.

3.10) Preparação da cultura de *A. tumefaciens* e infecção dos discos foliares de tabaco

Cem microlitros de um pré-inóculo, obtido a partir de uma colônia isolada de *A. tumefaciens* da linhagem LBA4404 transformada com a construção EGLV1:GUS,

foram inoculados em 10 ml de meio líquido LB, sem a presença dos antibióticos de seleção no intuito de evitar interferência dos mesmos na co-cultura com os discos foliares. A cultura foi incubada sob agitação a 150 rpm por 16 horas a 28°C até atingir a absorbância a 600 nm igual a 1.5. Uma alíquota do inoculo foi então diluída até atingir uma absorbância a 600 nm de 0.1 para então ser utilizada na co-cultura com os discos foliares de tabaco (Aragão *et al.*, 2002).

Em um fluxo estéril, discos foliares de ~2 cm de diâmetro provenientes de tabaco (*Nicotiana tabacum* SR1) cultivado *in vitro* (Anexo 8.11) foram cortados de folhas jovens e colocados em uma placa de Petri umedecida com água deionizada estéril no intuito de manter os explantes hidratados. Um ml da cultura diluída de *A. tumefaciens* foi adicionado a 10 ml de LB líquido num tubo de centrifuga estéril de 50 ml. Com uma pinça estéril, os explantes foram colocados no tubo e a co-cultura foi mantida à temperatura ambiente por 5 min sob leve agitação. Transferiram-se os discos foliares para uma nova placa de Petri contendo papel de filtro estéril para retirar o excesso de bactérias.

Os discos foliares co-cultivados foram transferidos para placas de Petri contendo meio Murashige & Skoog (Murashige & Skoog, 1962) sólido, pH 5.7, suplementado com sacarose (30 g/l), 2 µg/ml de citocinina 6-benzilaminopurina (BAP) e 0.1 µg /ml de auxina ácido naftalenoacético (ANA), sendo as mesmas mantidas no escuro por 48 horas a 28°C, tempo necessário para a infecção da bactéria.

Os experimentos foram divididos em placas com discos infectados pelas bactérias transformadas com a construção EGLV1:GUS e sem infecção. Discos foliares sem infecção por *A. tumefaciens* foram usados como controle positivo de regeneração dos explantes em meio MS com BAP e ANA, e como controle negativo para testar posteriormente o efeito do antibiótico de seleção em tecidos não transformados.

3.11) Regeneração de plantas transformadas

Após 48 horas de incubação no escuro, os explantes transformados e os controles não transformados foram transferidos para novas placas com meio MS sólido contendo sacarose, BAP e ANA nas mesmas concentrações anteriores. Nessas placas foram adicionados o bacteriostático cefotaxima (300 µg/ml) e o antibiótico higromicina-B (50 µg/ml). Nesse meio, o tecido vegetal não transformado não irá apresentar regeneração já que não apresenta resistência a higromicina-B, enquanto que a cefotaxima irá inibir o crescimento da *A. tumefaciens*. Discos foliares correspondentes ao controle positivo foram também transferidos para placas sem antibióticos para testar o efeito do BAP e ANA na regeneração dos calos. Todas as placas foram mantidas em sala de cultivo sob fotoperíodo de 16 horas e temperatura de $22 \pm 2^{\circ}\text{C}$. Após o aparecimento de partes aéreas no material regenerante, essas foram transferidas para frascos de vidro estéreis contendo meio MS em presença dos agentes de seleção e de ANA para induzir o enraizamento.

Após o estágio de enraizamento, as plantas foram transferidas para copos de plástico de 300 ml contendo vermiculita e terra vegetal na proporção de 1:2, respectivamente, e aclimatadas em caixas cobertas fechadas parcialmente com filme PVC transparente. Após algumas semanas, discos foliares de 2 cm de diâmetro foram cortados de tais plantas e o DNA genômico extraído de acordo com o protocolo de Konieczny *et al.* (1993), com posterior tratamento com RNase (10 mg/ml). As plantas foram transferidas para a estufa em vasos maiores, sendo suas sementes coletadas posteriormente.

3.12) Identificação e caracterização molecular das linhagens transgênicas

3.12.1) Confirmação da inserção via PCR usando DNA genômico

Com o intuito de verificar a inserção do cassete de expressão EGLV1:GUS nas linhagens transgênicas regeneradas foi realizada uma PCR utilizando 15 ng/ μ l de DNA genômico extraído de cada linhagem (total de 14 plantas de eventos independentes). DNA genômico extraído de tabaco selvagem (WT) foi usado como controle. Para a amplificação foram utilizados, respectivamente, o oligonucleotídeo senso EGLV2B junto à região promotora, e o oligonucleotídeo reverso GUSR (5' GTCTGCCAGTTCAGTTCGTTGTTC 3') complementar a região codificadora do gene GUS . Uma temperatura de anelamento de 54°C foi utilizada, gerando um produto de amplificação de 467 pb.

3.12.2) Detecção de transcritos do gene repórter por RT-PCR

Após a confirmação da integração do cassete de expressão foi realizada uma reação de RT-PCR usando RNA total de algumas linhagens no intuito de verificar se o gene GUS estava sendo transcrito nas plantas transformadas. Com esse objetivo, extraiu-se RNA total de folhas coletadas das plantas transgênicas de tabaco com Trizol (Invitrogen), seguindo as recomendações do fabricante, para posterior transcrição reversa (RT) como descrito anteriormente. Uma amplificação foi então realizada

empregando os cDNAs produzidos e oligos internos à região codificadora do gene GUS, gerando um produto de amplificação de 400 pb.

3.13) Ensaios histoquímicos da atividade GUS em tabaco

Os ensaios histoquímicos para detecção da atividade de β -glucuronidase (GUS) nos tecidos transformados foram realizados empregando-se um substrato halogênico (5-bromo-4-chloro-3-indolyl β -D-glucuronide; X-Gluc), que em contato com a enzima se torna azulado.

Como controle positivo foram utilizados tabacos transgênicos transformados com um cassete de expressão contendo o promotor constitutivo CaMV 35S fusionado transcricionalmente ao gene repórter GUS (cedidos pelo Prof. Dr. Márcio de Castro Silva-Filho – Departamento de Genética - ESALQ-USP). Folhas retiradas de plantas adultas de tabaco transgênico (4 meses) e plântulas da geração T1 com 10 dias pós-germinação foram utilizadas no experimento. O ensaio histoquímico foi realizado baseado em Jefferson *et al.* (1987) e Lacorte & Romano (1998).

Inicialmente os discos foliares de 2 cm de diâmetro foram tratados com um tampão de pré-fixação (formaldeído 0,3 %, MES 10 mM - pH 5.6, manitol 0,3 M) por 45 min em temperatura ambiente, seguido de várias lavagens em tampão fosfato 50 mM pH 7.0; as folhas foram então submersas no tampão fosfato contendo 1 mM de X-Gluc, substrato que é oxidado pela enzima β -glucuronidase, sendo incubadas a 37°C por 16 horas. Para retirar-se a clorofila da amostra, após a reação histoquímica, os discos foram lavados em álcool 70 % por diversas vezes (Jefferson *et al.*, 1987).

Adicionalmente, plântulas inteiras das linhagens transgênicas EGLV1:GUS (geração T1) foram submetidas ao ensaio GUS. Essas foram inicialmente incubadas em tampão de reação contendo 100 mM $\text{Na}_2\text{PO}_4 \cdot \text{H}_2\text{O}$, 0,5 mM $\text{K}_4\text{Fe}(\text{CN})_6 \cdot 3\text{H}_2\text{O}$, 10 mM $\text{Na}_2\text{EDTA} \cdot 2\text{H}_2\text{O}$, 0,1 % Triton X-100, 1 mM X-Gluc (50 mg/ml) e infiltradas a vácuo por 7 min. O tempo de reação foi de 16 horas a 37°C. Após esse período as amostras foram lavadas em etanol 70 % para retirar a clorofila do material vegetal.

3.14) Identificação do início de transcrição pela técnica 5' RACE

A técnica RACE (Rapid Amplification of cDNAs Ends) foi utilizada para determinar as terminações 5' dos transcritos dos candidatos EGLV1 e EGFB1. Para tal foi empregado o kit 5'RACE (Invitrogen) segundo as especificações do fabricante. Na amplificação foram utilizados três oligonucleotídeos gene-específicos (GSPs) relacionados às terminações 5' dos ESTs validados e oligonucleotídeos complementares aos adaptadores. Na primeira etapa, a primeira fita de cDNA é sintetizada a partir do RNA total extraído do tecido-alvo de eucalipto empregando para tal o oligonucleotídeo mais externo (GSP1) e a enzima *SuperScript II* (Invitrogen). A molécula híbrida de RNA-cDNA é tratada em seguida com um mix de RNase H visando a eliminação do RNA molde. A fita de cDNA é purificada através de uma coluna de sílica com presença de um agente caotrópico, visando a ligação de produtos de cDNAs > 200 pb e eliminação do excesso de dNTPs, enzimas, tampões e oligonucleotídeos. O resíduo remanescente do agente é então eliminado em lavagem em álcool etílico 70% e o cDNA purificado é eluído da coluna em água deionizada autoclavada. Em seguida, a adição de uma cauda homopolimérica na terminação 5' do cDNA é realizada com auxílio da

enzima Terminal Transferase (TdT) e de dCTP. Uma PCR é realizada utilizando-se os oligonucleotídeos GSP2 (GSP2) e o “Abridged Anchor Primer” (AAP), seguida de outra rodada de PCR mais específica em “nested” empregando os oligonucleotídeos GSP adjacente e “Abridged Universal Amplification Primer” (AUAP). Os produtos gerados foram clonados em pGEM-Teasy (Clontech) para o sequenciamento.

3.15) Expressão transiente do cassete EGLV1:GUS em *E. grandis* e ensaios histoquímicos da atividade GUS

Os ensaios de expressão transiente em eucalipto foram realizados em colaboração com a empresa Suzano Papel e Celulose. Para tal foi utilizada a metodologia SAAT (*sonication-assisted Agrobacterium mediated transformation*) conforme protocolo adaptado para eucalipto por González (2002). Nesse método, microferimentos são provocados nos explantes de eucalipto por sonicação (60 s) a fim de aumentar a penetração da agrobactéria no tecido vegetal. Nesse caso foram usadas plântulas de eucalipto obtidas *in vitro*. Quarenta e oito horas após a infecção das plântulas de eucalipto pela agrobactéria contendo o cassete de expressão LGV1:GUS, as mesmas foram submetidas ao ensaio histoquímico empregando o substrato X-Gluc.

4) Resultados

4.1) Validação biológica dos candidatos com expressão órgão-específica

Dos 59 candidatos selecionados *in silico* para a validação biológica do perfil de expressão via RT-PCR (Tabela 01), 3 eram dados como apresentando padrão de expressão ubíquo, 18 eram preditos como sendo específicos de madeira e/ou caule de muda, 16 como apresentando expressão específica de raiz, 11 específicos de folha, 7 de botão floral e/ou fruto e 4 de madeira/caule e raiz.

Tabela 01 – Resultado das análises de similaridade (*Blast*), órgão-especificidade *in silico* e validação (RT-PCR) dos contigs selecionados *in silico* no banco FORESTs. A seqüência dos oligonucleotídeos e o tamanho (T) esperado dos produtos (pb) da validação também são apresentados.

Contigs	Blast	NCBI	E-Value	<i>In silico</i>	RT-PCR	Seqüência dos Oligos (5' → 3')	T (bp)
EGEQR3102A09.g	14-3-3 c-2 protein [<i>Nicotiana tabacum</i>]	GI:44917143	2e-113	Ubíquo	AI	R – GACCGCTGCTGTCTTGAAT F – ACAACGCTGCAACACAAGAG	258
EGACFB1015B10.g	Sem resultados significativos	-	-	Bofão floral	FB	R – CTGGCCCTGGAACAGAGTTA F – TCTCTCGCCTTGTGGTCTT	233
EGEQR3002E03.g	Boron transporter 1 [<i>Arabidopsis thaliana</i>]	GI:30690607	0.0	CL, RT	Todos	R – CCATCAGAAACAGCAAAGCA F – TTGGGGAACAACCTGGAGAG	259
EGAGLV2214H10.g	Hypothetical protein [<i>Vitis vinifera</i>]	GI:147826727	2e-89	L.V, SL	Todos	R – ACCACCCAGCACCTTTCCTTC F – GGCATGTTCTGTGCATCATC	250
EGABST2222G06.g	Beta-carotene hydroxylase [<i>Coffea arabica</i>]	GI:76560802	1e-119	ST, BK	LV, RT, ST, WD, FB	R – CCTCCAGAACGAAAGCATACC F – CGTCTCAATTCAACGCCACAC	250
EGEPRT3362H10.g*	unnamed protein product [<i>Vitis vinifera</i>]	GI:157334321	1e-84	RT, SL	RT	R – GCGTAGGACCCCGATGAAGAT F – AAGGGCTCGATAGGGATCAT	229
EGABSL1082B12.g	Hypothetical protein [<i>Vitis vinifera</i>] Glyceraldehyde-3-phosphate dehydrogenase [<i>Nicotiana tabacum</i>]	GI:147858936	8e-149	WD, SL, CL	Todos	R – GCCACACAGAGAGCCAAAAGT F – GAACCCCTCCCAGTAAATGC	247
EGEQFB1001C06.g	Glyceraldehyde-3-phosphate dehydrogenase [<i>Nicotiana tabacum</i>]	GI:4539543	3e-178	Ubíquo	Todos	R – CTTCTCGTTCACACCCACAA F – ACCACGAGCTCAAGGTCAAG	248
EGEQWD2247D05.g	Decreased apical dominance protein [<i>Petunia x hybrida</i>]	GI:57918684	0.0	WD	RT, ST, WD	R – CTCCATCCTCACCACCAAGT F – CTGACTGACAACGCCAACAC	243

EGCCLV2224H06.g	Early light inducible protein [<i>Lycopersicon esculentum</i>]	GI:46401440	4e-45	LV	Todos	R - GCCACAAAAGCCTATCATTCG F - CCTAGGACCCGCAATCAACAT	251
EGEQRT3301H05.g	Oxidoreductase, 2OG-Fe(II) oxygenase family protein [<i>Arabidopsis thaliana</i>]	GI:15219097	1e-118	RT	LV, RT, ST	R - GTGTTTCTCCACGCCGTAAT F - GGAGATCAAAGAGCCAGAACG	256
EGCART6008F08.g	Ubiquitin-like protein [<i>Phaseolus vulgaris</i>]	GI:1684855	8e-132	RT	AI	R - TCTTGCCAGTGAAGGGTTTTTC F - AAAGGCCAAGA TCCAAGACA	185
EGEQLV2201B04.g	Auxin-binding protein ABP19 [<i>Prunus persica</i>]	GI:18203684	1e-75	LV	LV	R - GCAGAAATCGACCCTTGAAT F - TCCTGCAAGAAAGTCTCAGC	245
EGBFFB1040G05.g	(1-4)-beta-mannan endohydrolase [<i>Coffea arabica</i>]	GI:10129675	1e-172	FB	SA	R - TTTCCTTATGATCGCCCTTCG F - ATAAGCACITTGCTGGCGATT	251
EGBFFB1042G10.g	Transcription factor MAD55 [<i>Prunus persica</i>]	GI:126842923	2e-24	FB	FB, FT	R - CGAGCTTCCACAGTCTGTCA F - GCTCATGGAAAGGTACCAAG	246
EGBFLV2238G01.g	Epsilon-adaptin, putative [<i>Arabidopsis thaliana</i>]	GI:30692596	1e-32	LV, SL	LV, ST, FB, FT	R - CCCAGTAAGCAAGTGCAATG F - CCAGGAGTGACAAATGGTAGC	240
EGBFLV2250F11.g	AT4g02920/T4I9_20 [<i>Arabidopsis thaliana</i>]	GI:15450579	4e-19	LV, SL	Todos	R - GATTGTTCCGGCAATCCAGT F - GAGCTCTGCTTGA TGGCTTC	242
EGEQFB1201B08.g	Hypothetical protein [<i>Vitis vinifera</i>]	GI:147777974	2e-155	FB	SA	R - CCCATGACCTCACTGGCTAT F - CATCTCGCAAAGCTCAAAAACA	259
EGEQWD2200E06.g	Hypothetical protein [<i>Vitis vinifera</i>]	GI:147795424	5e-112	WD	Todos	R - GAAACACCCAGAGGCAATTG F - TTCTGGGTGGTTGCTGATG	80
EGEQWD2200A06.g	Hypothetical protein [<i>Vitis vinifera</i>]	GI:147771889	8e-103	WD	LV, RT, ST, WD, FB	R - TCC TCAACCTCTCCAACAATG F - GTCCACATCAACGTCGACAA	233
EGEQWD2232H04.g	Hypothetical protein [<i>Vitis vinifera</i>]	GI:147841964	5e-172	WD	Todos	R - TCATATAGAAAAGCCGAGACCAA F - CAGCAATGGCGATGCTGAA	80
EGEZWD2266E05.g	Targeting for Xklp2 [<i>Medicago truncatula</i>]	GI:92874845	2e-49	WD	Todos	R - CGCACGTTCCGGTGTCTTAA F - TGCAC TATCGCATCCAAACAAA	80

EGEQWD2228F05.g	Histidine acid phosphatase family protein [<i>Tetrahymena thermophila</i> SB210]	GI:118370137	0.005	WD	Todos	R - CACAAAGCATGACGGCAACA F - CAAATTACGTATCGCCCAACGA	80
EGEZRT6201C03.g	AT4g32020/F10N7_170 [<i>Arabidopsis thaliana</i>]	GI:14326506	3e-05	RT	AI	R - GGCTCCAGTTTCGACTTTTG F - GACCCCTCTCTCCCCAGAAA	249
EGEZRT6201E05.g	C-repeat binding factor 1b [<i>Eucalyptus gunnii</i>]	GI:79013964	5e-100	RT	AI	R - ATGTCTTGGTGTGGGTGA F - AGAAGTCCGGGAGACAG	250
EGEZFB1006E01.g	Peroxidase 3 [<i>Scutellaria baicalensis</i>]	GI:4760704	4e-110	FB	AI	R - ATGTTGTCGACCCACGTCAAA F - GGTTCAAGCTATGCCAGACT	252
EGBFST2270C12.g	Hypothetical protein [<i>Vitis vinifera</i>]	GI:147770431	6e-27	ST	Todos	R - GGATCAAACACAGTGCCAAA F - TGCTCGGAGCCTTATACTGG	252
EGQHSL1102C02.g*	Phospholipid hydroperoxide glutathione peroxidase [<i>Momordica charantia</i>]	GI:17646156	1e-80	SL, ST	RT, ST, WD, BF	R - CTCTTCATTACTCCCTGGCTC F - CTTCAACGGGCAICGACC	286
EGQHLV2242E10.g*	Os08g0559400 [<i>Oryza sativa</i> (japonica cultivar-group)]	GI:115477813	6e-72	LV	LV, RT, ST, WD, FB	R - GATAAATCTCCATTCTCTGCG F - CCTCCCGAGGTCACCTC	387
EGBMR16276A06.g	Cinnamoyl-CoA reductase family [<i>Arabidopsis thaliana</i>]	GI:15226134	2e-71	RT, SL	LV, ST, WD, FB	R - GGGTTCATCATCATTGCAG F - GCTTCGTGGTTCGCAAG	353
EGRFCL1206E01.g*	Cellulose synthase 3 [<i>Eucalyptus grandis</i>]	GI:67003911	5e-126	CL, RT, WD	RT, ST, WD, BF	R - CTGCCATGTACTTGTGCTTGTTC F - CTGGACTTGTCCGCTGGTTCTC	360
EGMCR13145G04.g*	HAK5 [<i>Solanum lycopersicum</i>]	GI:94483077	1e-83	RT	RT	R - GCGACAGAACCCCAAGAATG F - GTGGCGCTGGAGGAAGAAG	316
EGEZRT6274D07.g*	oxidoreductase family protein [<i>Arabidopsis thaliana</i>]	GI:15221433	1e-87	RT, SL	AI	R - GCTTCCTTCCACCATCAG F - CTGCTTTTGCCTCTGTTCC	268

EGEQSL1057C07.g	Sem resultados significativos	-	ST, WD, RT, SL, CL, FB	Todos	R-CTACAAAACCGTTGATTCC F-GTTGTGAGGGTTCGGCG	150
EGEQRT3302D12.g	Nitrate transporter [<i>Populus tremula</i> x <i>Populus tremuloides</i>]	GI:57283317	CL, RT	LV, ST, WD, FT	R-GGTGATGAGGATGAGGGAAG F-GAAGCTCTCCGCCCATATG	423
EGEQRT3102B06.g*	Hypothetical protein [<i>Vitis vinifera</i>]	GI:147825149	RT, ST	LV, RT, ST, WD, FB	R-GGTTGTCCACGGTCTTGC F-CAAAAACCCCTCTGGCATC	384
EGEQRT3001H05.b	Cellulose synthase 1 [<i>E. grandis</i>]	GI:67003907	RT, ST, WD	LV, RT, ST, WD, FB	R-CAITGAGTTGAGCTGCCAATG F-GGGTCGGGAGAAAGATTGG	270
EGEQLV2001G11.g	Photosystem I subunit PsaD [<i>Glycine max</i>]	GI:148372347	LV	Todos	R-GCGACTCCCAAGGTGATCAC F-GGATTCACATAAGACCCGCTC	424
EGCEFB1016A09.g	Alpha-tubulin [<i>Eucalyptus grandis</i>]	GI:153799903	Ubíquo	AI	R-GTTGGCGGCATCTTCTTTG F-CTCTCTCGCCGGATCCTTC	363
EGCCST6215A11.g*	Fascilin-like arabinogalactan protein [<i>Eucalyptus grandis</i>]	GI:153799891	ST, WD	LV, RT, ST, WD, FB	R-GACGGAAAGGATGATGTG F-GATGAAAGCATCAACTGGTAC	319
EGCBRT6017A01.g*	KUP2 [<i>Vitis vinifera</i>]	GI:93115181	RT, SL	RT, ST, FB	R-GAACTAAAGTGAGCGTCC F-GGCTTCTGCATTTCAATCC	378
EGBGBK1048G08.g*	Dihydroflavonol-4-reductase [<i>Citrus sinensis</i>]	GI:41351796	RT, SL, WD	Todos	R-GAACTCCCTCCACCCGTCG F-GAATCAGCGGCAGCTAATGG	378
EGAGSL4212E01.g	Hypothetical protein [<i>Vitis vinifera</i>]	GI:147778418	SL, ST, WD	LV, RT, ST, WD, FB	R-GTGCCATCGCTAGTTTGTTC F-CAGTTCGTGCTCGCAATGG	275
EGBMLV2260A10.g	Pentatricopeptide (PPR) repeat-containing protein [<i>A. thaliana</i>]	GI:30694913	LV	Todos	R-CAITTCITTAGCTGTGTAAC F-GAGCTGCTTACCTGAAAAATC	443
EGCEFB1018G01.g	Multicopper oxidase, putative [<i>Medicago truncatula</i>]	GI:92891717	FB	AI	R-CTCAACCAGTTCGGCTCCT F-TCATCAGGGGATGGTTCGTA	254
EGJMST6256E09.g	Os01g0794400 [<i>Oryza sativa</i>]	GI:115440447	ST	LV, RT, ST, FT, FB	R-CCTTATTACCATCGGGCTGA F-GTACTCGCCATGTCCGAACT	251
EGEQWD2001B12.g	Sem resultados significativos	-	WD	Todos	R-CGGAGAGACGACAGGGGATTA F-GTCGCGGATCTTCGACCT	223

EGEQLV2001D10.g	Hypothetical protein OsI_035938 [<i>Oryza sativa</i> (indica cultivar-group)]	GI:125535491	4e-109	LV	LV, ST, FT, FB	R - GTTCTCATCCAGCCAGTTG F - TCTTCCACAGCAAATGTGCA	199
EGEQLV2001A11.g	Putative chloroplast terpene synthase [<i>Quercus</i> <i>ilex</i>]	GI:14331015	8e-45	LV	LV, FB	R - TCCTTCTCTCAAGCGCTAA F - CGAGTCTTTGAAC TGGCTCA	250
EGEQFB1004C12.g	Peptidase C1A, papain [<i>Medicago truncatula</i>]	GI:140053138	1e-81	FB	Todos	R - ATCGCCGTATCCGGTAATC F - GAAGGCCTCAACCAGATCAA	243
EGQHRT6254F07.g	Laccase [<i>Populus</i> <i>trichocarpa</i>]	GI:3805956	3e-120	RT	LV, RT, ST, WD, FB	R-CTCATCTCCACC GACCAGAG F-CGACGGTAAGAGGGACTTTG	247
EGBMR T3131F09.g	ATSEB1 [<i>Arabidopsis</i> <i>thaliana</i>]	GI:18414592	2e-124	RT	Todos	R-CAGACTACCAGTGGGTCAA F-AAGAAGAA TGGCTGGAGCTG	260
EGEQRT3101C03.b	Laccase [<i>Populus</i> <i>trichocarpa</i>]	GI:3805964	0.0	RT, WD	LV, RT, ST, WD	R-CGAA TAAAGACCCGACCAA F-AACTTCTGAT TGGGCTG TGG	205
EGEZRT3007A11.g	Hypothetical protein [<i>Vitis vinifera</i>]	GI:147861513	0.0	SL, RT	RT, FT, FB	R-ATCTCCAGGTCAAACCATCG F-GGTGCTCCTCGTTTGCTAAG	254
EGEZRT6001A02.g	Peroxidase [<i>Glycine max</i>]	GI:4204765	7e-127	RT	LV, RT, ST, WD, FB	R-AAGATCTCGTCGCTTTTCG F-ACGTGTTGTCGAAAGGTGGTT	216
EGBMR T6276A06.g*	Cinnamoyl-CoA reductase family [<i>Arabidopsis thaliana</i>]	GI:15226134	1e-102	RT, SL	Todos	R-CTTCGTGGGTCCGTC AAGATT F-GGTGCAAGCAGCTCTACCTC	253
EGBMLV2260A10.g	Pentatricopeptide (PPR) repeat-containing protein [<i>A. thaliana</i>]	GI:30694913	1e-116	LV	Todos	R-TCAAATGACCCAAACCCTGT F-TTCATCAGCCTCCTCCAGTT	237
EGEZWD2231B04.g	Methyladenine glycosylase family protein [<i>A. thaliana</i>]	GI:42571455	2e-98	WD	LV, RT, ST, WD, FB	R-TCCAACGACTCGTCCCTTAC F-CCAGCACTCATCATGAAACG	252
EGUTST2032C06.g*	Putative endo-1,4-beta- glucanase [<i>E. globulus</i>]	GI:70779689	0.0	ST	LV, RT, ST, WD, FB	R-ACGTCAAAGTATGCCGGTCTC F-GGCTGAGGATGTCCGAGTAG	255

AI - Amplificação inespecífica (mais de um tamanho de banda), SA - Sem amplificação, SL - Plântula, CL - Callus, LV - Folha, RT - Raiz, ST - caule de muda, WD - tronco, FB - botão floral BK - casca e madeira.

* baseado em literatura

Ao final de todas as análises de RT-PCR, os seguintes candidatos tiveram seus padrões de expressão confirmados: EGEQFB1001C06.g com padrão ubíquo, EGEQLV2201B04.g com expressão específica em folha, EGEPR3362H10.g e EGMCRT3145G04.g com expressão em raiz, EGACFB1015B10.g com expressão em botão floral e EGBFFB1042G10.g com expressão em botão floral e fruto concomitantemente (sendo os mesmos denominados EGUB1, EGLV1, EGRT1, EGRT2, EGFB1 e EGFT1, respectivamente; Figura 05).

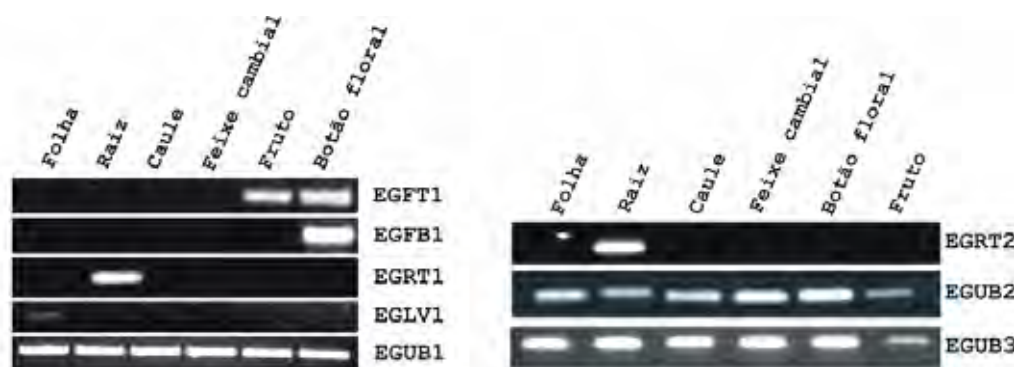


Figura 05 - Validação dos candidatos por RT-PCR. Horizontalmente estão dispostos os tecidos/órgãos testados e verticalmente estão os ESTs candidatos. Os produtos de amplificação foram analisados em gel de agarose 1,5 % corado com brometo de etídio. Somente alguns perfis selecionados estão apresentados.

Os demais candidatos analisados não apresentaram o padrão de expressão apontado pelas análises *in silico* (padrões obtidos estão disponibilizados na Tabela 1 – coluna RT-PCR), fato que pode ser confirmado na Figura 5 pelo perfil de expressão obtido, por exemplo, para os candidatos EGUB2 e EGUB3, respectivamente EGEQSL1057C07.g e EGBGBK1048G08.g, que deveriam ter expressão específica em caule/madeira. Tal predição foi baseada na análise *in silico* para o candidato EGUB2 e na literatura para o candidato EGUB3 (Ludwikow *et al.* 2004). Cabe ressaltar que nas diferentes análises realizadas o candidato EGUB1 foi sempre utilizado como controle

positivo de amplificação, pois seu padrão é ubíquo, e que o candidato EGRT1, que tem expressão somente em raiz, foi usado como indicador de ausência de contaminação das amostras de cDNA com DNA genômico. Com esses controles procurou-se eliminar a presença de possíveis falso-positivos e/ou negativos.

4.2) Isolamento das regiões promotoras

Os candidatos EGLV1, EGFB1, EGFT1, EGUB1, EGRT1 e EGRT2 (Figura 05) foram selecionados para o isolamento de suas regiões promotoras. Nas tentativas iniciais de amplificação via PCR, observou-se a presença de fragmentos com adaptadores nas duas extremidades, sendo que os mesmos não possuíam correlação com os ESTs validados. Em outros casos observou-se ausência total de amplificação, fato possivelmente associado à amplificação de fragmentos muito grandes (maiores que 6 kb).

Para se detectar as presenças de falsos positivos amplificados com os adaptadores foram realizadas amplificações controles utilizando-se o oligonucleotídeo do adaptador (AP) nas 4 “bibliotecas” genômicas geradas. Se algum fragmento amplificado nas reações específicas apresentasse tamanho semelhante ao fragmento amplificado nas reações AP, o mesmo era descartado. As Figuras 06 e 07 representam a amplificação secundária do GW dos candidatos EGLV1 e EGRT2, respectivamente, onde os controles negativos foram utilizados e as bandas não coincidentes foram cortadas, purificadas e clonadas para sequenciamento.



Figura 06 – Produtos de amplificação obtidos para o candidato específico de folha (EGLV1) após o PCR secundário do *genome walking*. DL1-4 são bibliotecas genômicas produzidas pela clivagem das enzimas *DraI*, *EcoRV*, *PvuII* e *StuI* respectivamente. Gel de agarose 1% corado com brometo de etídio. M-MassRuler DNA ladder Mix (Fermentas); (-) Controle negativo com AP2; O círculo representa o fragmento purificado e clonado.



Figura 07 – Produtos de amplificação obtidos para o candidato específico de raiz (EGRT2) após o PCR secundário do *genome walking*. As 4 bibliotecas utilizadas para amplificação com o oligo GSP2 estão anotadas com os nomes das enzimas de restrição que as geraram. Ladder 1 kb; 1A - 1D (controles negativos de amplificação). Os produtos foram observados em gel de agarose 1% corado com brometo de etídeo. Os círculos representam os fragmentos purificados e clonados.

Foram amplificadas e clonadas regiões promotoras de tamanho adequado para os candidatos EGLV1, EGFB1, EGUB1 e EGRT2 (Tabela 02). Os candidatos EGRT1 e EGRT2 ainda estão em processo de caracterização nas bibliotecas genômicas. Já para o candidato EGFT1, o fragmento de 750 pb gerado pelo sistema não foi suficiente para sair da região 5' do EST. Neste caso, o fragmento de EST disponível no banco para o desenho dos oligonucleotídeos não atingia a porção 5' do gene, fato verificado quando esse foi comparado com homólogos de outras espécies ou usando programas de predição de ORFs.

Tabela 02 – Nome dos candidatos, os respectivos padrões de expressão validados via RT-PCR e o tamanho dos fragmentos amplificados por GW. O atual status do processo de sequenciamento também está indicado.

Candidatos	Padrão de expressão órgão-específico	Tamanho da região amplificada (kb)	Seqüenciamento
1) EGBF1	Botão floral	1,9	1,1 kb
2) EGLV1	Folha	1,5	OK
3) EGUB1	Ubíquo	3,0	0,8 kb
4) EGRT2a	Raiz	1,4	0.8 kb
5) EGRT2b	Raiz	3,0	2,0 kb
6) EGRT2c	Raiz	3,0	1,8 kb

4.3) Sequenciamento das regiões amplificadas por GW

Os fragmentos amplificados para os candidatos EGLV1, EGFB1, EGFT1 e EGUB1 foram purificados e clonados em vetor pGEMT-easy para posterior sequenciamento. Em função do tamanho de alguns dos fragmentos seqüenciados, houve a necessidade de se realizar reações de sequenciamento em série, empregando para tal oligonucleotídeos desenhados com base nas seqüências recém-obtidas. A situação atual do processo de sequenciamento dos fragmentos obtidos no GW encontra-se descrita na Tabela 02.

É importante frisar que para o candidato EGUB1 foram observados problemas de correlação entre o fragmento amplificado e seqüenciado e o EST originalmente usado para desenho dos oligos. Detectou-se que a seqüência de nucleotídeos do fragmento amplificado apresentava alta identidade com um EST não relacionado com

aquele usado para o desenho dos oligos. Esse novo EST codifica uma isoforma da proteína codificada pelo EST originalmente empregado. Genes cujos produtos gênicos apresentam regiões muito conservadas normalmente dificultam as amplificações por GW.

4.4) Correlação das seqüências amplificadas com os ESTs

Conforme descrito no item 3.6, amplificações foram realizadas no intuito de confirmar a identidade e posicionamento do fragmento amplificado, e cuja seqüência foi determinada, em relação ao seu respectivo EST. Nas Figuras 08 e 09 estão apresentados dois exemplos de tais amplificações, as quais correspondem aos candidatos EGFB1 e EGLV1. Nesse caso foi empregado tanto DNA genômico de eucalipto quanto plasmidial (contendo o fragmento clonado por GW), e oligonucleotídeos sintetizados no sentido senso do fragmento seqüenciado e no sentido reverso à seqüência dos ESTs depositados no banco. Os resultados obtidos confirmam o posicionamento dos produtos de amplificação à montante do EST validado.

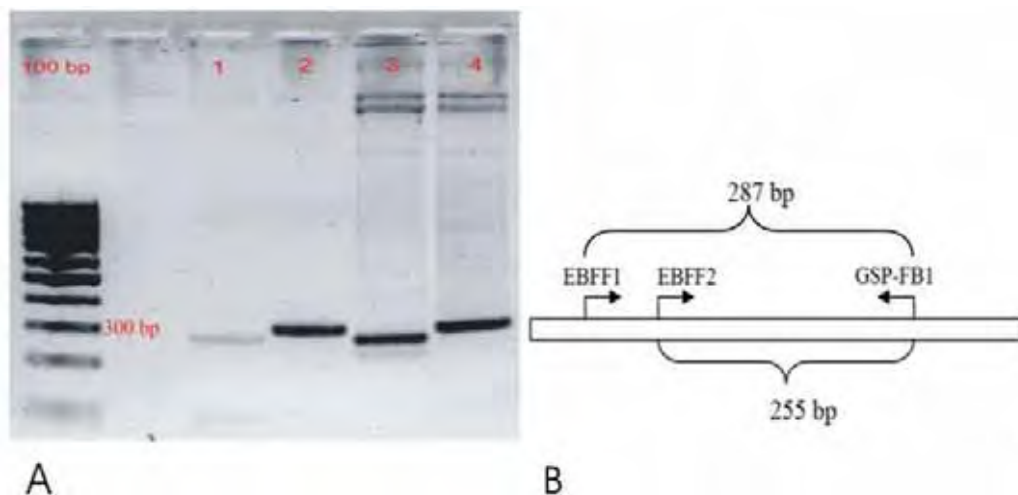


Figura 08 – Confirmação do posicionamento da região promotora amplificada para o candidato EGFB1 em relação ao EST selecionado no banco FORESTs (A). Amplificações foram realizadas com o par de oligos EBFF1 e GSP-FB1 (colunas 2 e 4 – amplicon de 287 pb) e com o par de oligos EBFF2 e GSP-FB1 (colunas 1 e 3 – amplicon de 255 pb), usando DNA genômico (colunas 1 e 2) ou o plasmídeo pGEMT-Easy contendo o fragmento EGFB1 (colunas 3 e 4). 100bp –. Ladder 100 pb da MBI Fermentas. Gel de agarose 1% corado com brometo de etídio. (B) A diferença de tamanho das bandas resulta do posicionamento do oligo EBFF1 à montante do oligo EBFF2, levando a amplificação de um fragmento ligeiramente maior.



Figura 09 – Confirmação do posicionamento da região promotora amplificada para o candidato EGLV1 em relação ao EST selecionado no banco FORESTs. O fragmento (300 pb) foi amplificado usando oligo específico à região promotora (EGLV2B) e outro ao EST validado (GSP). M – Marcador MassRuler DNA Ladder Mix (Fermentas); 1–3 Amplificações usando DNA genômico de eucalipto; (-) Controle negativo (sem DNA); 4 – Amplificação usando o plasmídeo pGEMT-Easy contendo o fragmento EGLV1. Gel de agarose 1% corado com brometo de etídio.

4.5) Construção do cassete de expressão para análise funcional em tabaco

Dentre os diferentes promotores putativos amplificados (Tabela 2), o promotor putativo do candidato EGLV1 (específico de folha) foi inicialmente selecionado para a análise funcional em expressão estável em tabaco. Para tal, a região promotora putativa foi amplificada por PCR para posterior inserção em vetor de transformação de plantas. O promotor putativo amplificado relativo ao candidato EGLV1 tem 1.2 kb e sua região 3' terminal se limita ao último nucleotídeo imediatamente adjacente a região codificadora do EST validado. Esse fragmento foi inserido no vetor pCAMBIA-1381z em fusão transcricional com o gene *uidA* (GUS).

4.6) Transformação estável em tabaco

4.6.1) Confirmação de transformação da agrobactéria

O vetor binário pCAMBIA-1381z contendo o cassete de expressão EGLV1:GUS foi transformado em *Agrobacterium tumefaciens* de acordo com protocolo já descrito (item 3.9). As agrobactérias recombinantes foram identificadas através de PCR empregando um oligonucleotídeo posicionado no sentido senso da sequência promotora e um outro complementar à região codificadora do gene repórter próximo ao ATG, originando um fragmento em torno de 500 pb (Figura 10).

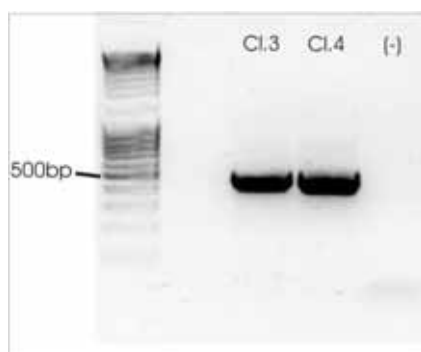


Figura 10 – PCR confirmatório da transformação do vetor binário contendo o promotor EGLV1 em *A. tumefaciens*. Da esquerda para direita: MassRuler DNA Ladder (Fermentas); Cl.3 - clone 3; Cl.4 – clone 4 e (-) controle negativo (sem DNA).

4.6.2) Transformação dos discos foliares e regeneração dos transformantes

Após os eventos de transformação via *A. tumefaciens* (item 3.10), os discos foliares foram mantidos em placas contendo meio de cultura MS com os agentes de seleção próprios descritos anteriormente. O experimento de regeneração foi separado em controle positivo de regeneração, sem antibióticos, e controle negativo de seleção, sem evento de transformação, visando testar a eficiência da higromicina B. O controle negativo após algumas semanas apresentava perda progressiva de clorofila e ausência de calos, enquanto o controle positivo e os explantes infectados com agrobactérias apresentavam calos nas partes transformadas.

Após várias semanas houve a indução da parte aérea nos explantes. Nessa etapa transferiu-se a parte aérea para potes com meio MS com a presença de ANA para induzir o enraizamento, mantendo sempre os agentes de seleção. Após algumas semanas

desenvolveu-se o sistema radicular nos tabacos transformados. Nesse estágio, as plantas foram transferidas para copos plásticos e aclimatadas para o novo ambiente.

4.7) Caracterização molecular dos transformantes

No total, quatorze plantas transformadas (representando eventos independentes) foram recuperadas do processo de regeneração. Para averiguar se o cassete de expressão EGLV1:GUS estava integrado nessas linhagens foi realizada a extração de DNA genômico, o qual foi então usado como molde em reação de PCR empregando os oligonucleotídeos utilizados no item 3.12.1. Como pode ser observado na Figura 11, a amplificação de um fragmento correspondente ao cassete de expressão inserido foi observada. Os fragmentos inespecíficos presentes na Figura 11A puderam ser eliminados aumentando-se a temperatura de anelamento de 53°C para 54°C (Figura 11B).

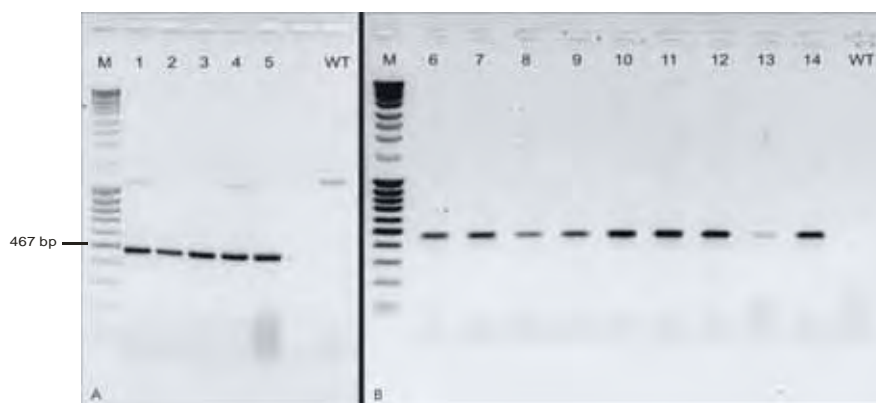


Figura 11 – PCR confirmatório da inserção do cassete de expressão EGLV1:GUS nas linhagens transgênicas de tabaco (1-14). Amplicon esperado de 467 pb. WT – Tabaco não transgênico usado como controle negativo; M - Marcador MassRuler DNA Ladder Mix (Fermentas).

Mais uma etapa confirmatória foi realizada no intuito de verificar se a transcrição do gene repórter nas linhagens transformadas estava ocorrendo. Para tal, RNA total foi extraído de folhas de algumas linhagens escolhidas aleatoriamente e uma RT-PCR foi realizada empregando oligonucleotídeos internos ao gene repórter, gerando um fragmento de 400 pb. Nesse caso, a presença de transcrito correspondente ao gene repórter foi observada em todas as linhagens analisadas (Figura 12).

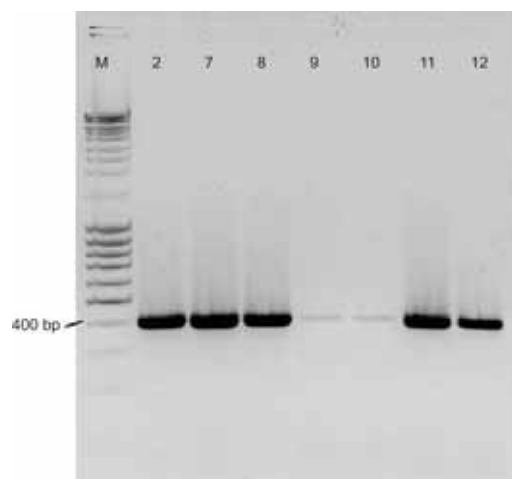


Figura 12 – RT-PCR confirmatório da expressão do gene repórter (GUS) nas plantas de tabaco transformadas com o cassete EGLV1:GUS. Gel de agarose 1% corado com brometo de etídio. M – Marcador MassRuler DNA Ladder Mix (Fermentas); Os números 2, 7, 8, 9, 10, 11 e 12 correspondem às linhagens transgênicas testadas.

As plantas transgênicas caracterizadas molecularmente foram transplantadas para vasos maiores com terra vegetal e areia (2:1) para recuperar as sementes da geração T1 e usá-las no ensaio histoquímico de GUS.

4.8) Ensaio histoquímico de GUS nas plantas EGLV1:GUS

Inicialmente, folhas e discos foliares dos transformantes primários (geração T0) para o cassete EGLV1:GUS foram submetidas ao ensaio histoquímico de GUS. A intenção era averiguar se a expressão de GUS estava efetivamente ocorrendo confirmando assim os eventos de transformação. Na Figura 13 é possível observar, nas folhas e discos foliares de duas linhagens analisadas (EGLV1.2 e EGLV1.7), a presença de coloração azulada relativa a reação da enzima β -glucoronidase com o substrato X-Gluc. O padrão parcial e variegado observado é provavelmente resultante da absorção parcial do substrato nos tecidos vegetais analisados.

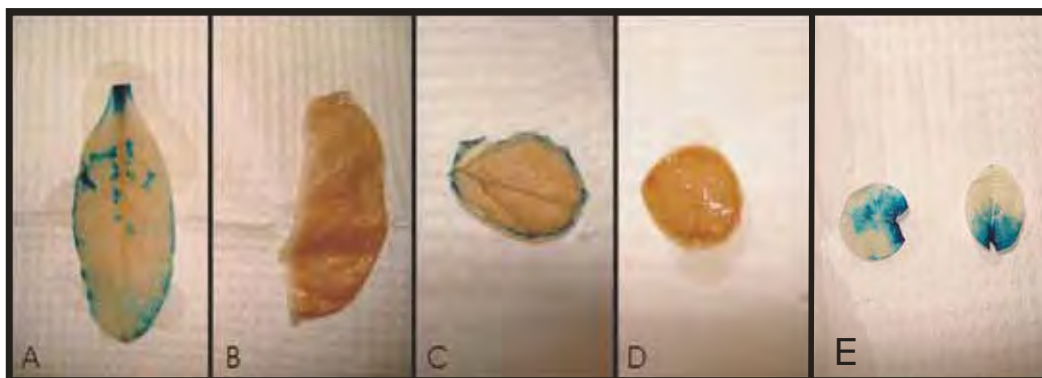


Figura 13 – Ensaio histoquímico de GUS em tabaco. A) Folha da linhagem transgênica 7 com marcação azulada correspondente a atividade GUS; B) Controle negativo – planta selvagem; C) Disco foliar da Linhagem 2; D) Controle negativo – planta selvagem; E) Controle positivo – planta transformada com um cassete contendo o gene repórter GUS sob controle do promotor constitutivo 35S.

Com a disponibilidade das sementes da geração T1 foi possível realizar o ensaio histoquímico de GUS em plântulas recém germinadas (Figura 14). Nesse experimento foi usado como controle negativo, plântulas de tabaco selvagem SR1, e como controle positivo, plântulas contendo o cassete de expressão 35S CaMV-GUS. As linhagens transgênicas contendo o cassete EGLV1:GUS apresentaram padrão de expressão de GUS específico de folhas, incluindo em alguns casos, os primórdios foliares (Figura 14F). Um padrão variegado foi novamente observado nas folhas coradas, padrão esse que pode estar relacionado à inadequada absorção do substrato.

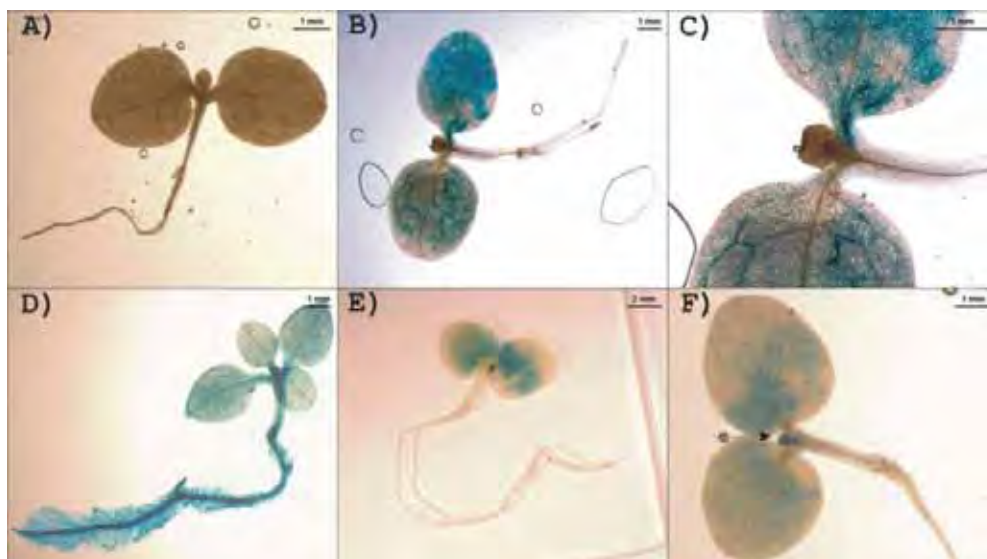


Figura 14 – Ensaio histoquímico de GUS em plântulas de tabaco da geração T1 contendo o cassete EGLV1:GUS. A) Controle não transgênico (SR1); B e C) Plântulas da linhagem 7; D) Controle positivo 35S-CaMV-GUS; E e F) Plântulas da linhagem 2. Seta – Primórdio foliar.

4.9) Análise da expressão relativa do gene repórter nas plantas EGLV1:GUS

Em uma análise mais detalhada, a expressão relativa do gene *uidA* foi avaliada em tabacos transgênicos da geração T1 transformados com a construção EGLV1:GUS. Para tal, foram utilizadas amostras de RNA total extraído de folha, caule e raiz de plantas da linhagem EGLV1.2 com 4 meses de idade. Nesse caso, o gene 18S de *Nicotiana tabacum* (Levy *et al.*, 2004) foi utilizado como normalizador. Pelos resultados obtidos é possível notar que a expressão do gene repórter foi majoritária em

folhas, sendo detectada também em caule e raiz, porém em menor intensidade (Figura 15).

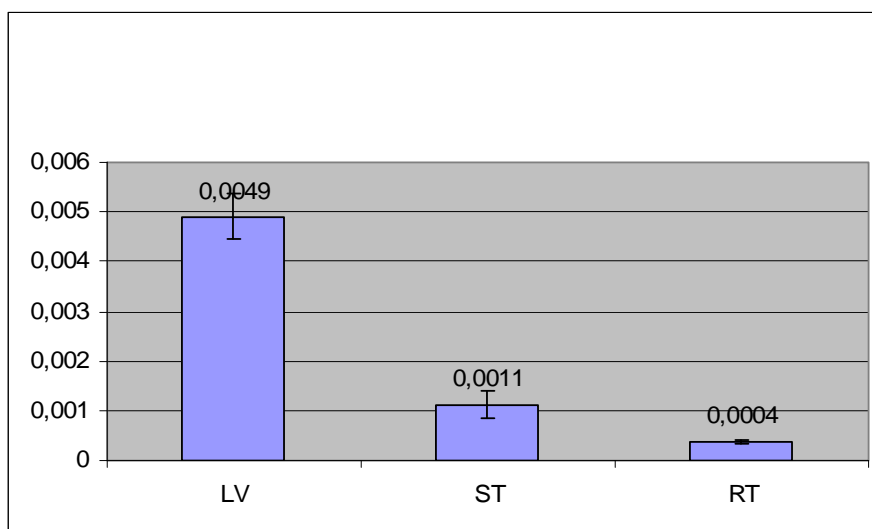


Figura 15 – Análise da expressão relativa do gene *uidA* (GUS) em diferentes órgãos de tabaco transgênico EGLV1:GUS (linhagem 2). LV – folha, ST – caule, RT – raiz. As barras correspondem à média mais desvios para três reações independentes.

4.10) Expressão transiente em *E. grandis* e ensaio histoquímico de GUS

A fim de confirmar a funcionalidade da região promotora do candidato EGLV1 em eucalipto, ensaios de expressão transiente nessa espécie foram realizados. Os resultados obtidos (Figura 16) mostram que a atividade GUS pôde ser detectada nas folhas cotiledonares das plântulas de *E. grandis* infectadas pela agrobactéria transformada com o cassete de expressão EGLV1:GUS (Figura 16; Painel B), demonstrando ser este, um promotor ativo em eucalipto. Um padrão similar de

expressão foi observado no controle positivo usando um cassete 35S:GUS (Painel A). Por outro lado, não foi observada a expressão de GUS no controle negativo (Painel C). Adicionalmente, ensaios histoquímicos em tufos de propágulos vegetativos de *E. grandis* infectados com agrobactéria contendo o cassete EGLV1:GUS foram realizados, contudo, não houve detecção de atividade GUS (Painel E) como também observado no controle negativo (Painel F). Por outro lado, no controle positivo 35S:GUS (Painel D) foi detectada atividade GUS.

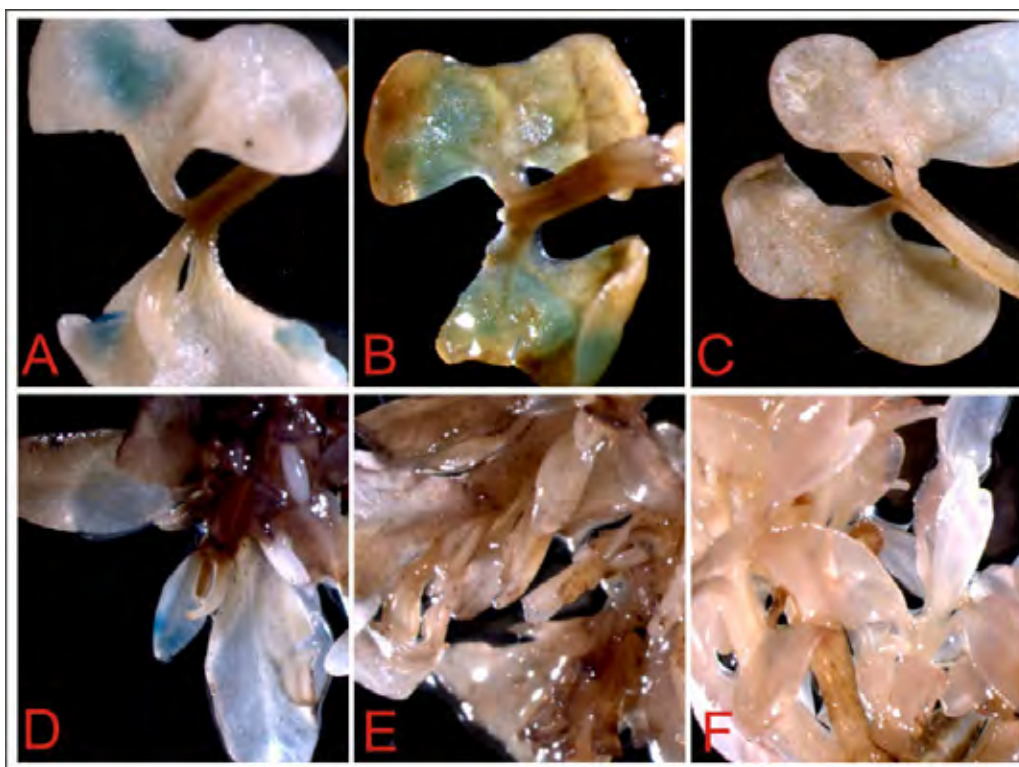


Figura 16 – Expressão transitória em eucalipto usando o cassete de expressão EGLV1:GUS. A, B e C – folhas cotiledonares. D, E, F – Tufo de propágulos vegetativos. A e D - Controle 35S:GUS; B e E - Cassete EGLV1:GUS; C e F - Controle negativo sem inoculação.

4.11) Análise da expressão relativa empregando qPCR

4.11.1. Candidato EGLV1 (EGEQLV2201B04.g)

O candidato EGLV1 foi inicialmente validado nas análises de RT-PCR como apresentando um perfil de expressão específico em folhas (Figura 5). Porém, nas análises histoquímicas em tabaco transgênico, uma expressão residual em outros órgãos (caule e raiz) também foi observada (Figura 15). Com o intuito de melhor caracterizar o padrão de expressão desse candidato, uma análise de expressão mais aprofundada foi realizada empregando-se material vegetal coletado de plantas de eucalipto em diferentes idades. Nesse caso, optou-se por quantificar a expressão relativa do mesmo em relação ao gene normalizador *EgIDH* (Goicoechea *et al.*, 2005) usando-se qPCR.

As análises de expressão relativa revelaram que esse candidato é preferencialmente expresso em folhas, sendo a maior quantidade de transcritos detectada em folhas de eucalipto com 6 meses e em folhas de eucalipto proveniente de propagação vegetativa *in vitro* (Figura 17). Por outro lado, a sua expressão em folhas provenientes de plantas com 6 anos e 1 mês, 2 anos e 7 meses e em estágio de plântula (cerca de 1 mês) foi muito baixa, sendo ainda praticamente nula nos demais órgãos analisados.

O seqüenciamento dos produtos de amplificação gerados via PCR quantitativo em tempo real revelou que a região analisada é 97,3% idêntica à região correspondente do EST EGEQLV2201B04.g, sendo observadas duas mudanças de nucleotídeos. Em duas posições ocupadas por citosinas no EST foi observada uma timina em uma delas, mas nunca nas duas.

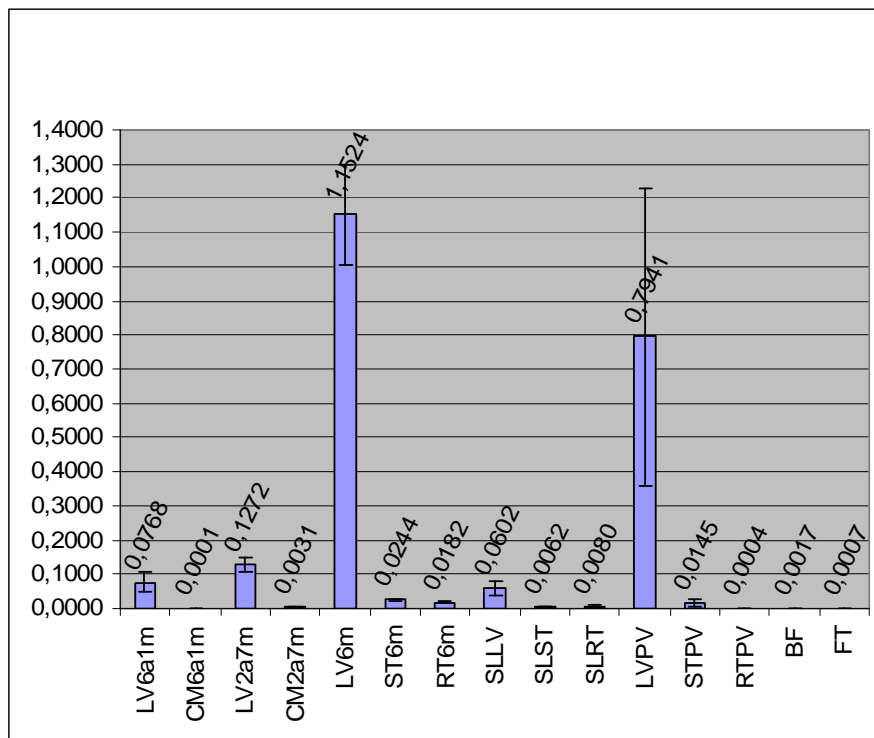


Figura 17 – Expressão relativa do candidato EGLV1 em diferentes órgãos, de diferentes idades, de *E. grandis*. LV6a1m e CM6a1m – folha e feixe cambial de eucalipto com 6 anos e 1 mês; LV2a7m e CM2a7m – folha e feixe cambial de eucalipto com 2 anos e 7 meses; LV6m, ST6m e RT6m – folha, caule e raiz de eucalipto com 6 meses; SLLV, SLST e SLRT – folha, caule e raiz de plântulas; LVPV, STPV e RTPV – folha, caule e raiz de eucalipto em propagação vegetativa in vitro; BF – botão floral; FT – fruto. As barras correspondem à média mais desvios para três reações independentes.

4.11.2. EGRT2 (EGMCRT3145G04.g)

Para efeito de comparação, a expressão relativa do candidato EGRT2, inicialmente validado nas análises de RT-PCR como possuindo um perfil de expressão específico em raiz (Figura 5), foi também analisada. Sua expressão relativa também foi normalizada usando o gene *EgIDH* (Goicoechea *et al.*, 2005).

Ao contrário do observado para o candidato EGLV1, o candidato EGRT2 manteve seu padrão de expressão nas análises de qPCR. Nesse caso, verificou-se que a sua expressão ocorria exclusivamente em raízes de *E. grandis*, sendo a mesma majoritária em raízes obtidas de plantas de propagação vegetativa, seguida de plantas com 6 meses, e de uma menor expressão em raízes provenientes de plântulas (Figura 18). A detecção nos demais órgãos foi nula, ou seja, não se detectou transcritos desse gene em folhas e feixes cambiais de plantas com 6 anos e 1 mês e de plantas com 2 anos e 7 meses, e em folhas e caule de plantas com 6 meses (dados não apresentados).

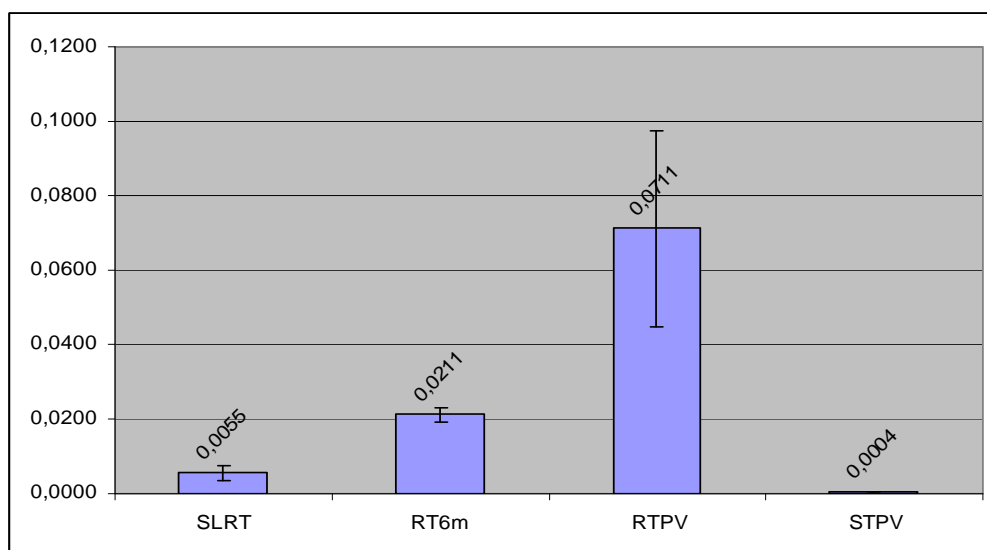


Figura 18 – Expressão relativa do candidato EGRT2 em diferentes órgãos, de diferentes idades, de *E. grandis*. SLRT – raiz de plântulas; RT6m – raiz de plantas com 6 meses; STPV e RTPV – caule e raiz de eucalipto em propagação vegetativa in vitro. As barras correspondem à média mais desvios para três reações independentes.

4.12) Determinação da região 5' não traduzida dos candidatos EGLV1 e EGFB1 usando 5'- RACE

Um experimento adicional foi empreendido a fim de determinar o tamanho aproximado da região não traduzida dos candidatos EGLV1 e EGFB1. Usando a técnica 5'-RACE foram amplificados e isolados fragmentos de aproximadamente 250 pb e 200 pb para os candidatos EGLV1 e EGFB1, respectivamente. As amplificações foram realizadas com cDNA molde usando os devidos controles para verificar a especificidade dos produtos amplificados (Figura 19).

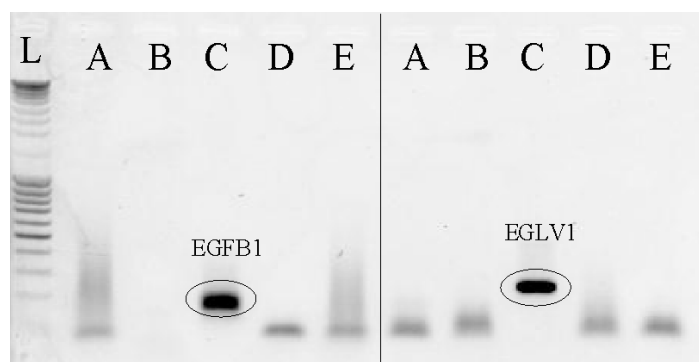


Figura 19 – Produtos de amplificação obtidos por 5'-RACE para os candidatos EGFB1 e EGLV1. A) Controle da primeira reação de síntese. B) Controle do eluato da coluna de sílica. C) cDNA tratado com TdT correspondente ao produto final do RACE. D) Controle sem tratamento de TdT. E) Controle sem cDNA. L) Ladder MassRuler DNA Mix (Fermentas). Círculos – fragmentos clonados e seqüenciados.

Os produtos foram clonados em vetor pGEMT-easy e seqüenciados. As seqüências obtidas foram alinhadas com as seqüências dos EST depositados no banco e com as regiões promotoras para determinar os putativos pontos de início de transcrição.

Até o presente momento, somente a seqüência do candidato EGLV1 teve boa qualidade indicando que o gene possui uma região 5' não traduzida de 48 nucleotídeos.

4.13) Elementos *cis*-regulatórios presentes no promotor EGLV1

Foram realizadas buscas por elementos *cis*-regulatórios putativos presentes na região amplificada do candidato EGLV1 utilizando-se o banco de dados PLACE (<http://www.dna.affrc.go.jp/htdocs/PLACE/>) (Higo *et al.*, 1999). O promotor EGLV1 tem um tamanho total de 1267 nucleotídeos, incluindo a região 5' não traduzida (48 pb) a qual foi estipulada por 5'-RACE. O ponto de início de transcrição determina o posicionamento relativo dos elementos *cis*-atuantes. Na busca por elementos regulatórios na seqüência promotora putativa em análise, procurou-se evidenciar e investigar os elementos mais repetitivos (disponibilizados no Anexo 8.12). De maneira geral, sabe-se que os programas de predição de sítios regulatórios ainda não são acurados gerando um alto índice de falso-positivos. Prestridge e Burks (1993) demonstraram que num conjunto de seqüências não promotoras de mamíferos submetidas a uma análise com uma matriz de peso (Bucher, 1990), gera-se em média um TATA-box a cada 120 pb.

Dentre os diferentes elementos de regulação encontrados no promotor ELGV1, os que mais chamam a atenção são os elementos do tipo DOFCOREZM (AAAG), que são sítios de ligação para os fatores de transcrição do tipo DOF, e que se encontram repetidos inúmeras vezes ao longo da seqüência promotora (24 vezes) (Figura 20).

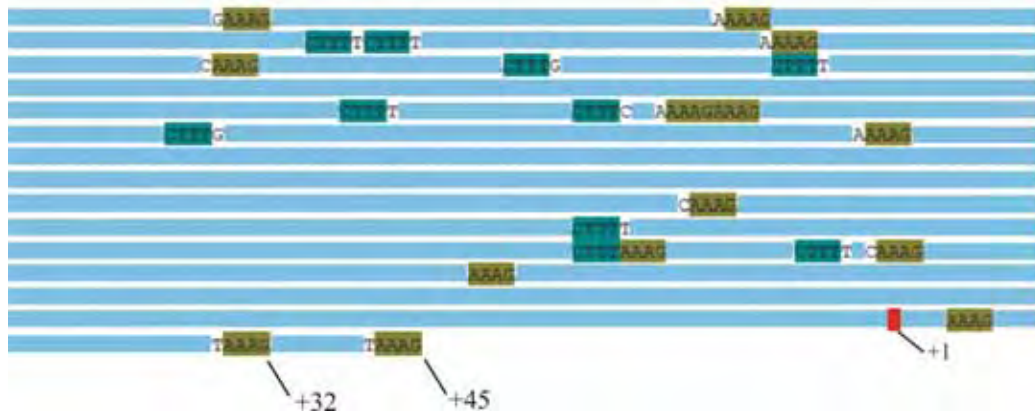


Figura 20 – Distribuição dos elementos *cis*-regulatórios do tipo DOFCOREZM no promotor EGLV1 (1267 pb). O motivo AAAG está representado pelos retângulos bege (senso) e verde-azulado (reverso). As barras em azul representam a seqüência de nucleotídeos do promotor. O retângulo em vermelho representa o ponto de início de transcrição do gene.

5) Discussão

5.1) Validação por RT-PCR e qPCR

Dos 59 candidatos selecionados por predições *in silico* e dados da literatura, 6 candidatos foram confirmados como apresentando expressão órgão-específica num primeiro momento (cerca de 10%). Em trabalhos semelhantes de validação de perfil de expressão de ESTs em café e cana-de-açúcar, a taxa de validação também é baixa (Hoshino, 2007; Brandalise, 2007). Isso pode ter ocorrido devido ao fato de alguns genes apresentarem níveis reduzidos de expressão em alguns órgãos/tecidos, o que faz com que os mesmos não sejam amostrados nas bibliotecas de ESTs. Nesse caso, tais candidatos apresentam-se como preferencialmente expressos em dado órgão/tecido, porém sem serem específicos.

Como a etapa de RT-PCR visou identificar candidatos órgão-específicos pela detecção ou ausência de produtos de amplificação nos diferentes órgãos analisados, sem a exigência de uma quantificação, procurou-se refinar as análises dos perfis de expressão de alguns candidatos usando PCR quantitativo em tempo real. Com essa ferramenta foi possível verificar que o candidato EGLV1 era na verdade, preferencialmente expresso em folhas. No caso de EGRT2 o perfil se manteve específico de raiz. Com base nesses resultados fica evidente que a análise quantitativa é um passo essencial para a validação final dos candidatos antes da realização dos ensaios funcionais.

O candidato EGFB1 não apresentou similaridade com nenhuma seqüência conhecida em bancos públicos. A sua seqüência deduzida de aminoácidos indicou a presença de uma região rica em prolinas, ocupando cerca de 38% da seqüência primária da proteína, resíduo normalmente presente em grande número em proteínas estruturais.

Muito embora um cassete de expressão tenha sido obtido para o candidato EGFB1, nos ensaios funcionais optou-se pela caracterização do candidato EGLV1, já que sua região promotora estava completamente seqüenciada e sua expressão em folha era mais interessante do ponto de vista aplicado. O candidato EGRT2 foi, do ponto de vista cronológico, caracterizado mais tardiamente. Por esse motivo, as análises funcionais do promotor correspondente não foram realizadas. Entretanto, esse candidato foi usado nas análises de expressão relativa como um controle com padrão de expressão específico em raiz.

5.2) O candidato EGLV1

Numa busca por seqüências similares à seqüência deduzida de aminoácidos do candidato EGLV1 (usando para tal a seqüência do EST depositada no banco FOREST) em bancos de seqüências públicas, verificou-se que essa apresentava similaridade com diversas seqüências relacionadas às proteínas do tipo germininas. As proteínas germininas fazem parte da superfamília Cupina, que tem como característica principal um domínio com dois motivos, que apresentam pelo menos um resíduo de histidina cada, separados por uma região de tamanho variável (Dunwell *et al.*, 2000, 2001). O termo Cupina vem do latim *cupa*, que significa pequeno barril, devido à estrutura de barril beta desse domínio (Dunwell, 1998). A seqüência deduzida de aminoácidos do

EST apresenta o mesmo tipo de arquitetura do domínio das cupinas, onde a seqüência consenso do primeiro motivo é $G(X)_5HXH(X)_{3,4}E(X)_6G$ e a do segundo motivo é $G(X)_5PXG(X)_2H(X)_3N$ [X representa qualquer aminoácido (Gane *et al.*, 1998)]. Os dois resíduos de histidina e o ácido glutâmico presentes no motivo 1 bem como o resíduo de histidina presente no motivo 2 agem em conjunto como ligantes, criando um possível sítio de ligação com afinidade a metais.

A superfamília Cupina foi concebida com base no fato de que a proteína germinina de trigo, uma proteína com termo-estabilidade produzida durante a fase inicial de germinação de embriões de trigo, compartilhava nove aminoácidos (HI/THPRATEI) com a proteína induzida por estresse chamada esferulina do micetozoa *Physarum polycephalum* (Dunwell, 1998). Outras cupinas foram identificadas em diversos táxons, incluindo muitos exemplos em bactérias (Dunwell e Gane, 1998), sendo que coletivamente elas compartilham as seqüências consensos de histidinas e a estrutura de barril beta do domínio. Em uma estimativa inicial, previu-se que existiriam cerca de 18 diferentes sub-classes funcionais dentro das cupinas (Dunwell *et al.*, 2001), mas atualmente sabe-se que existe um número bem maior. De maneira geral, existem dois grandes grupos de cupinas, as que possuem um domínio simples de cupina, e as que possuem 2 ou mais domínios de cupinas (Dunwell *et al.*, 2004), sendo que as proteínas do tipo germininas e as germininas participam do grupo das monocupinas.

As proteínas do tipo germininas (GLPs – Germin-like proteins) e germininas constituem uma família grande e diversa de proteínas de plantas. O nome germinina vem da primeira proteína descrita, a germinina de trigo. Todas as GLPs e germininas possuem um tamanho aproximado de 220 aminoácidos, incluindo um peptídeo sinal putativo. As proteínas maduras possuem 3 regiões altamente conservadas chamadas de “boxes” A, B e C e uma região hipervariável limitada por duas cisteínas (Figura 21).

Em mais da metade das GLPs, mas não nas germininas, nessa região está presente o triplet de aminoácidos RGD. Nas germininas o triplet KGD é mais frequentemente encontrado (ou algumas vezes KGE), e o “box” B e C possuem 3 histidinas e um glutamato que podem estar envolvidos em ligação a metais (Bernier e Berna, 2001).

A Figura 21 ilustra a seqüência deduzida de aminoácidos do candidato EGLV1 (211 aminoácidos). Ela apresenta um sítio de clivagem de um peptídeo sinal predito pelo programa SignalP 3.0 (Bendtsen *et al.*, 2004) entre os aminoácidos 19 e 20 (SHAAV, o sítio se localiza entre os aminoácidos sublinhados); os “boxes” B e C possuem as 3 histidinas e o glutamato; o “box” A é idêntico ao consenso, apresentando uma cisteína dentro do mesmo e outra na vizinhança; um sítio de N-glicosilação foi predito pelo programa NetNGlyc 1.0 (Gupta *et al.*, 2004); e o tripeptídeo KGD está presente também. A presença de duas cisteínas próximas (possibilidade de ponte de dissulfeto), a presença de um peptídeo sinal, e de um sítio de N-glicosilação, podem em conjunto indicar uma proteína de secreção. O tripeptídeo KGD é semelhante ao RGD, que está presente em proteínas animais de adesão a matriz extracelular, como a fibronectina e vitronectina. Essas proteínas costumam interagir com integrinas e participam da troca de informação entre o meio externo e interno das células (Bernier e Berna, 2001).

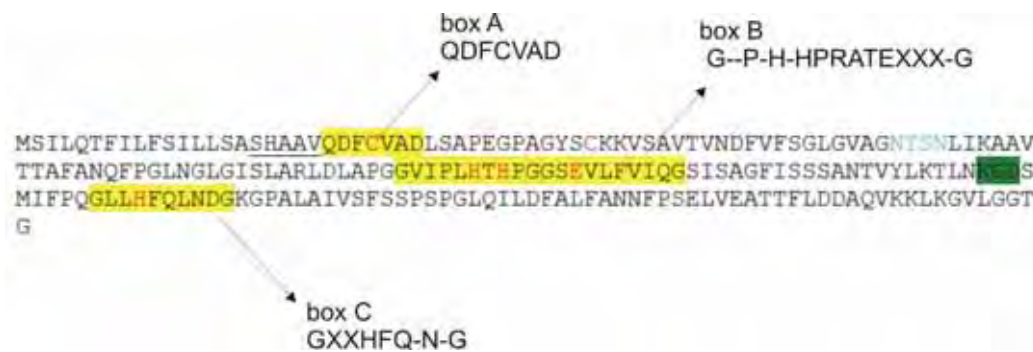


Figura 21 – Seqüência deduzida de aminoácidos do EST EGEQLV2201B04.g. As letras sublinhadas representam um sítio putativo de clivagem de um peptídeo sinal. Os quadros amarelos representam áreas moderadamente conservadas dos “boxes” A, B e C. Letras em azul representam um sítio putativo de N-glicosilação. O quadro verde mostra um tripeptídeo conservado. As letras em vermelho realçam as duas cisteínas, histidinas e um ácido glutâmico. – qualquer aminoácido; X – qualquer aminoácido hidrofóbico.

A maior parte das caracterizações bioquímicas, descritas na literatura, foi realizada empregando germininas, principalmente as de trigo e cevada, pois elas foram os primeiros membros identificados da família. As germininas são glicoproteínas oligoméricas (Lane, 1991), e a presença delas em matriz extracelular é bem documentada (Lane *et al.*, 1992; Lane, 1994). Em contraste, a maioria das GLPs conhecidas provém de projetos de sequenciamento, sendo a caracterização dessas proteínas esparsa. Dos poucos relatos bioquímicos (revisão em Bernier e Berna, 2001), todas as proteínas GLPs estudadas estão parcialmente associadas à matriz extracelular, variando a natureza de associação. Aparentemente, entre uma isoforma ou outra, a variação ocorre somente na cadeia lateral de oligossacarídeos das mesmas. Essa variação entre isoformas pode ser uma característica comum nessa família (Bernier e Berna, 2001).

A maioria das seqüências obtidas de GLPs via sequenciamento sistemático de diferentes espécies necessitam de caracterização do perfil de expressão. Mas, numa

visão geral, pode-se encontrar clones de cDNAs de GLPs em praticamente todos os órgãos e estágios de desenvolvimento das plantas, embora seja freqüente a identificação de múltiplos cDNAs provindos de uma mesma biblioteca. Nesse caso, eles geralmente se originam do mesmo gene. Em muitos casos, por mais que as GLPs estejam presentes em todos os tipos celulares, cada gene GLP é expresso de maneira específica e somente num subgrupo de células (Bernier e Berna, 2001).

É possível citar exemplos de genes codificando GLPs com o padrão de expressão preferencial em folhas, como no caso do gene *AtGER1* de *Arabidopsis thaliana* (Membré *et al.*, 1997); do gene *Zmglp1* em folhas novas de milho (Fan *et al.*, 2005); e do candidato EGLV1 validado neste estudo em *Eucalyptus grandis*. Há casos em que a expressão é específica de folhas como a do gene *PnGLP* em folhas novas de *Pharbitis nil* (Ono *et al.*, 1996); e a do gene *HvGLP1* também em folhas novas de *Hordeum vulgare* (Vallelian-Bindschedler *et al.*, 1998a).

Outro padrão interessante é que alguns genes apresentando expressão preferencial ou específica em folhas e outros não específicos, mas também presentes nas folhas apresentam, em geral, oscilação na quantidade de transcritos de acordo com o ciclo circadiano da planta. Em cevada, o gene *HvGLP1* é transcrito em maior quantidade no final do período de luz e em menor quantidade no final do período de escuridão (Vallelian-Bindschedler *et al.*, 1998a). Em *Pharbitis nil*, os transcritos do gene *PnGLP* se acumulam em maior quantidade no final do período de escuridão (Ono *et al.*, 1996), assim como gene *Zmglp1* de milho (Fan *et al.*, 2005). Os genes *Saglp* de *Sinapis alba* (Heintzen *et al.*, 1994) e *AtGER3* de *Arabidopsis thaliana* (Staiger *et al.*, 1999) possuem também um padrão circadiano de expressão apesar de não serem preferencial ou especificamente expressos em folhas. No caso do gene *Saglp*, o maior pico de transcrição foi detectado no final do período de luz, e em *AtGER3* no começo do

período da noite. A regulação desses genes em resposta ao ciclo circadiano levanta a questão se existem elementos regulatórios associados a esse tipo de resposta no promotor EGLV1.

Uma grande variedade de funções foi descoberta para a germininas e GLPs. As funções descritas incluem atividades enzimáticas específicas, atuação como receptores, e até como proteínas estruturais. Como exemplos de atividade enzimática associada pode-se citar a oxalato oxidase em germinina gf-2.8 de trigo (Lane, 1994, 2000) e em GLP de cevada, que reage a ataque do fungo *Erysiphe graminis* f. sp. *hordei* (Dumas *et al.*, 1995); a superóxido dismutase em nectarina I de tabaco (Carter *et al.*, 2000), e em Mn-SOD do musgo *Barbula unguiculata* (Yamahara *et al.*, 1999). Como exemplos de atuação como receptores é possível citar as proteínas ABP19/20 de pêssigo, prováveis receptores de auxina (Ohmiya *et al.* 2002), e as GLPs de raízes de pêra que atuam como receptores de ricadesinas, que são proteínas de superfície celular de bactérias do gênero *Agrobacterium* e *Rhizobium* (Swart *et al.*, 1994). Duas GLPs estruturais foram descritas em *Hordeum vulgare* reagindo a estresse de calor, peróxido de hidrogênio ou infecção por fungo. Através da sua insolubilização, essas GLPs alteram a composição da parede celular e aumentam a resistência a esses tipos de estresses (Vallelian-Bindschedler *et al.*, 1998a, b).

5.3) Os elementos regulatórios presentes no promotor EGLV1

Como realçado anteriormente, elementos do tipo DOFCOREZM (AAAG), que são sítios de ligação para os fatores de transcrição do tipo DOF, possuem inúmeras repetições (24 vezes) no promotor ELGV1. As proteínas Dof são uma família de fatores de transcrição que compartilham a afinidade pela seqüência AAAG localizada em diversos promotores de plantas (Yanagisawa *et al.* 1999). A ligação das proteínas Dof a diversos promotores de plantas sugere que as mesmas estejam associadas a genes que respondem a uma variedade de estímulos e/ou que são tecido-específicos. Interessantemente, todos os processos biológicos em que se especula serem mediados pelas proteínas Dof são específicos de plantas, como por exemplo, a regulação da expressão de genes fotossintéticos (Yanagisawa e Sheen, 1998), de genes endosperma-específicos (Vicente-Carbajosa *et al.*, 1997), de um oncogene de plantas e de um gene de resposta a patógenos (de Paolis *et al.*, 1996). As proteínas Dof1 e Dof2 de milho têm preferência de ligação por algumas regiões flangeadoras da seqüência AAAG, dentre elas uma adenina ou timina imediatamente anterior ao primeiro nucleotídeo do elemento *cis* ((T/A)AAAG), sendo que a presença de uma guanina diminui a força de ligação na mesma posição, e a de uma citosina abole totalmente a afinidade da proteína pelo DNA (Yanagisawa e Sheen, 1998). Outro fator preponderante é o aumento de afinidade das Dof pelo sítio AAAG quando ele apresenta múltiplas repetições, geralmente quando dispostas em seqüência. O candidato EGLV1 apresenta alguns elementos DOFCOREZM seqüenciais como, por exemplo, o motivo TAAAG repetido duas vezes (como esquematizado na Figura 20) dentro da região transcrita do gene EGLV1 (+ 32 e + 45), dentre outros.

Adicionalmente, os elementos GATA-box, CAATBOX1 e GT1CONSENSUS quando associados a elementos do tipo I-Box participam fortemente da mediação do padrão de expressão de inúmeros genes responsáveis a luz (Argüello-Astorga e Herrera-Estrella, 1996).

Por outro lado, a presença de elementos do tipo ROOTMOTIFTAPOX1, descritos como determinantes de padrão de expressão específica em raiz, não apresenta respaldo nas validações do candidato EGLV1 por RT-PCR ou qPCR e nem nos ensaios histoquímicos de GUS realizados no presente trabalho.

Elementos *cis*-regulatórios do tipo TATA-box foram preditos em regiões distantes do ponto de início de transcrição (entre -438 a -723) apresentando 6 repetições do TATA-box5 e uma do TATA-box4. O sítio putativo TATATATA presente na posição -25, embora não reconhecido pelo PLACE, aparenta ser o TATA-box funcionalmente mais plausível pelo seu posicionamento (~30 nucleotídeos anteriores ao ponto de início de transcrição).

A análise identificou também elementos associados à expressão em semente, como o EBOXBNNAPA, MYB2CONSENSUSAT e MYB1AT. Elementos associados à afinidade ao cobre (CURECORECR), de expressão específica em grãos de pólen (POLLEN1LELAT52), responsável a nodulação (NODCON1GM e NODCON2GM), ao ataque de fungo a raízes (OSE1ROOTNODULE e OSE2ROOTNODULE) e ao estresse biótico como os do tipo WKRY, também foram identificados. Contudo, alguns dos elementos descritos acima não possuíam os padrões de conservação descritos na literatura e indicativos de sua funcionalidade.

5.4) O candidato EGRT2

O candidato EGRT2 foi selecionado com base no perfil de expressão do gene *AtHAK5*, integrante da família dos genes *AtKT/HAK/KUPs* em *Arabidopsis thaliana* (Ahn *et al.*, 2004). Essa família é composta de treze membros distribuídos em 5 cromossomos (Maser *et al.*, 1997), os quais estão relacionados ao transporte e translocação de potássio. Os transportadores de potássio em *Arabidopsis* podem ser divididos em 3 grandes famílias gênicas a saber, KT/HAK/KUP, TRK/HKT e *Antiporte Cation Proton* (CPA), que juntos com os canais de potássio regulam o transporte e a translocação deste elemento (Gierth e Maser, 2007).

A expressão relativa do gene *AtHAK5* analisada por qPCR foi específica de raiz em estágio de plântula, e detectada em níveis quase imperceptíveis na planta em fase de florescimento. Contudo, uma grande indução de sua expressão em raiz, e em menor escala em folhas velhas de plantas em fase de florescimento, foi observada após um período de privação de potássio por 6 dias. Após 6 horas da suplementação de potássio, a expressão relativa do gene *AtHAK5* diminuiu para os níveis anteriores ao estresse (Ahn *et al.*, 2004). Em experimento de complementação em levedura mutante deficiente na capacidade de absorção de potássio, foi demonstrado que o gene *AtHAK5* complementa essa deficiência (Rubio *et al.*, 2000). Vale ressaltar que *AtHAK5* foi o único gene de toda a família *AtKT/HAK/KUPs* regulado pela deficiência de potássio em plântulas e também em plantas adultas.

Em eucalipto, os únicos genes descritos como relacionados ao transporte de potássio são os genes *EcKT1* e *EcKT2*, pertencentes à família TRK/HKT de transporte de K^+ e Na^+ (Liu *et al.*, 2001). Os seus cDNAs foram isolados de *Eucalyptus camaldulensis* e inseridos em oócitos de *Xenopus laevis*. As proteínas resultantes

aumentaram os níveis de potássio transportado para os oócitos na presença de sódio, mas também transportaram K^+ na ausência desse elemento, e vice-versa. Outra característica observada foi a indução da expressão do gene *EcKTs* na presença de soluções hipotônicas. Essa característica, em particular, pode ajudar o *Eucalyptus camaldulensis* a se adaptar ao seu habitat natural, que periodicamente é inundado, mantendo a homeostase de potássio. Cabe ressaltar que o candidato EGRT2 não apresenta nenhum grau de similaridade de nucleotídeos e nem de aminoácidos com os genes *EcKT1* e *EcKT2*.

6) Conclusão

Com base nos resultados obtidos no presente trabalho pode-se concluir que:

- A metodologia adotada para a seleção de genes candidatos com perfil de expressão específico em dado órgão/tecido mostrou-se pouco eficiente, atingindo taxas de validação de aproximadamente 10%.
- O fragmento de 1,2 kb isolado pela técnica de “genome walking” para o candidato EGLV1, apresenta atividade transcricional, sendo portanto funcionalmente ativo.
- Como demonstrado nos ensaios GUS, o promotor EGLV1 consegue modular a transcrição do gene repórter GUS preferencialmente em folhas de plantas de tabaco transgênicas. Esse resultado indica que os elementos *cis*-regulatórios necessários para dirigir preferencialmente a expressão gênica em folha estão presentes na referida seqüência.
- Os resultados de quantificação da expressão relativa do candidato EGLV1 em diferentes órgãos/tecidos de eucalipto corroboram o padrão de expressão preferencial em folhas.
- O candidato EGRT2 apresenta expressão específica em raiz, com exceção do caule de eucalipto propagado vegetativamente, em que se detectou uma presença mínima de transcritos.

7) Perspectivas

Durante a realização do presente trabalho foram disponibilizadas diferentes seqüências promotoras isoladas de genes com padrão de expressão específico em dado órgão de eucalipto. Uma dessas seqüências promotoras (de um gene com expressão em folha) foi caracterizada funcionalmente enquanto que as demais ainda requerem validação funcional. Os ensaios com tais objetivos estão sendo continuados por outros e serão disponibilizados.

Como iniciativa de uso da informação genômica para a identificação de regiões regulatórias em eucalipto, o presente trabalho tem um caráter pioneiro. A metodologia utilizada, apesar da baixa sensibilidade, mostrou-se viável e pode ainda ser mais bem trabalhada. Do ponto de vista aplicado, tais seqüências promotoras representam uma importante inovação biotecnológica para o setor florestal, pois a utilização das mesmas em cassetes de expressão visando a produção de transgênicos irá viabilizar a expressão racional do transgene em órgãos chaves da planta, livrando os demais órgãos da proteína heteróloga indesejada.

Tanto o promotor isolado do candidato EGLV1, quanto o promotor isolado do candidato EGRT2 poderão ser utilizados em aplicações que visem expressar preferencialmente proteínas de interesse em folhas ou em raízes de eucalipto, respectivamente. Conseqüentemente serão necessários mais testes submetendo a diferentes estresses bióticos/abióticos para determinar se eles são influenciados por eles, e assim determinar o melhor uso desses promotores.

8) Anexos

8.1) Loci, produtos gênicos, táxons e referências da Figura 02

A – SpHE (metaloendoprotease) de ouriço-do-mar *Strongylocentrotus purpuratus* (Wei *et al.*, 1995); B – DQA1 (proteína de histocompatibilidade) de *Homo sapiens* (Petronzelli *et al.*, 1995); C – β MHC (cadeia pesada da miosina) de *Rattus rattus* (Wright *et al.*, 1999); D – Lbc3 (leghemoglobina) de *Glycine max* (Stougaard *et al.*, 1987); E – Endo16 (proteína de adesão celular) de *S. purpuratus* (Yuh, Bolouri, e Davidson 1998, 2001); F – Forkhead (fator de transcrição winged-helix) do urocordata *Ciona intestinalis* (Di Gregório, Corbo, e Levine, 2001); G – APETALA3 (fator de transcrição MADS-box) de *Arabidopsis thaliana* (Hill *et al.*, 1998); H – DOMAD1 (fator de transcrição MADS-box) da orquídea *Dendrobium* cv Madame Thong-In (Yu, Yang, e Goh, 2002); I – Even-skipped (fator de transcrição homodomínio) de *D. melanogaster* (Sackerson, Fujioka, e Goto, 1999); J – GAL10 e GAL1 (enzimas metabólicas não relacionadas genealogicamente) de *Saccharomyces cerevisiae* (West, Yocum, e Ptashne, 1984). K – PAX6 (fator de transcrição paired-box) de *Mus musculus* (Kammandel *et al.*, 1999); L – Otx (fator de transcrição homodomínio) de *S. purpuratus* (Yuh *et al.*, 2002); M – Yp1 e Yp2 (proteínas parálogas yolk) de *D. melanogaster* (Chung *et al.*, 1996); N – Dlx6 e Dlx4 (fatores de transcrição homodomínio parálogos) de *Danio rerio* (Zerucha *et al.*, 2000; a estrutura éxon/ítron desse loci não é conhecida em detalhes; somente elementos regulatórios compartilhados são mostrados); O – APOC3 e APOA1 (proteínas carreadoras de lipídeos não relacionadas genealogicamente; somente elementos regulatórios compartilhados são mostrados) de *H. sapiens* (Li *et al.*, 1995; Naganawa *et al.*, 1997); P - Ftz e Scr (fatores de transcrição homodomínio parálogos) de *D. melanogaster* (Calhoun, Stathopoulos, e Levine, 2002);

Q – BM5 (proteína sinalizadora) de *M. musculus* (DiLeone, Russel, e Kingsley, 1998; a posição do éxon 3 não é precisamente conhecida; padrões de “splicing” são omitidos para simplificar o quadro).

8.2) Tampão CTAB estoque (100 ml) – Extração de RNA:

2% CTAB (cationic hexadecyl trimethyl ammonium bromide), 2% PVP (polyvinylpyrrolidone K 40), 100 mM Tris-HCl, pH 8.0, 25 mM EDTA, 2 M NaCl, q.s.p. água Milli-Q. Autoclavar o tampão.

8.3) Gel de agarose desnaturante 1% para RNA (40 ml):

Adiciona-se 0.4 g de agarose, 4 ml MOPS 10X, 2 ml formaldeído, 34 ml água deionizada tratada com DEPC. Ferver no microondas até dissolver completamente a agarose e adicionar 1 µl de brometo de etídio (10 mg/ml).

Tampão de carregamento desnaturante:

Para cada amostra: 2 µl de MOPS 10X, 3 µl formaldeído, 10 µl formamida, 2 µl de tampão de carregamento 1X, 3 µl de RNA total. Desnaturar as amostras com o tampão por 15 min a 65°C e dar um choque térmico no gelo.

8.4) Reação de qPCR:

MIX	Volume (µl)
SyBr Green Mix	5,8
Oligonucleotídeos (senso e reverso) mix	0,42 [0,4 µM]
H ₂ O deionizada (tratada DEPC) autocl.	3,28
cDNA (60ng/µl)	1,0
Volume total	10.5

8.5) Tampão CTAB estoque (50 ml) – Extração de DNA:

2% CTAB (cationic hexadecyl trimethyl ammonium bromide), 1% PVP (polyvinylpyrrolidone K 40), 100 mM Tris-HCl, pH 8.0, 20 mM EDTA, 1,4 M NaCl, 0,2 % 2-mercaptoetanol (adicionar somente na hora do uso) e completar com água Milli-Q.

8.6) Ciclagem de PCR utilizada na técnica do *Genome Walking*:

Termociclador utilizado – Perkin Elmer GeneAmp PCR System 9600

PCR primário:

- 7 ciclos: 94°C - 2 sec
72°C - 3 min
- 32 ciclos: 94°C - 2 sec
67°C - 3 min
- 67°C por mais 4 min.

PCR secundário:

- 5 ciclos: 94°C - 2 sec
72°C - 3 min
- 20 ciclos: 94°C - 2 sec
67°C - 3 min
- 67°C por mais 4 min.

8.7) Mini-preparação de plasmídeos por lise alcalina

Referência: Kit Colunas Qiagen, modificado

- 1) Inocular uma colônia isolada em 3 ml de meio LB contendo o antibiótico apropriado. Após incubação durante toda a noite a 37°C, transferir aproximadamente 1,4 ml para o tubo de microcentrífuga de 1,5 ml.

- 2) Centrifugar a 14000 rpm durante 60 s, descartar o sobrenadante. Adicionar outros 1,4 ml da cultura de bactérias e centrifugar outros 60s. Ressuspender as bactérias em 300 µl de P1 utilizando vortex.
- 3) Adicionar 300 µl de P2. Agitar por inversão suave do tubo e incubar à temperatura ambiente durante 5 min. Não prolongar este período em excesso.
- 4) Adicionar 300 µl de P3. É importante agitar cada tubo por inversão suave logo após a adição de P3. Nesta etapa deve ser formado um precipitado branco.
- 5) Incubar as amostras no gelo por 20 min.
- 6) Centrifugar a 14000 rpm durante 30 min e coletar com cuidado 700 a 800 µl do sobrenadante para um novo tubo. Evitar a coleta do precipitado que se forma na superfície da solução.
- 7) Centrifugar, se necessário, por 15 min novamente as amostras e coletar 650 a 750 µl do sobrenadante para um novo tubo.
- 8) Adicionar 400 µl de isopropanol e incubar a -20°C por 1 hora.
- 9) Centrifugar a 14000 rpm durante 30 min à temperatura ambiente.
- 10) Descartar o sobrenadante, adicionar 700 µl de etanol 70%, inverter o tubo com cuidado e centrifugar a 14000 rpm durante 5 min.
- 11) Descartar o etanol, secar completamente o DNA e ressuspender em 30 µl de água deionizada autoclavada.

Soluções:

P1) Tris-HCl (pH 8.0) 50 mM, EDTA (pH 8.0) 10 mM e RNase 100 µg/ml. Estocar a 4°C uma vez adicionada a RNase)

P2) NaOH 200 mM, SDS 1%

P3) Acetato de potássio 3M, ajustar o pH para 5.5 com ácido acético glacial. Armazenar a 4°C.

8.8) Mix da reação de sequenciamento:

MIX	Volume (μ l)
Save Money	2.0
Big Dye v.3.1	1.5
DNA molde	350 ng
Oligo	1.0 (0.5 μ M)
H ₂ O	completar para 10.0
Volume total	10.0

8.9) Produção de células competentes de *Agrobacterium tumefaciens*

- 1) Inoculou-se uma colônia isolada em 3 ml de meio LB contendo rifampicina (100 mg/l) e estreptomicina (50 mg/l);
- 2) Incubou-se o pré-inoculo a 28°C, com agitação (100-150 rpm), por 12 a 16 horas (A_{600} entre 0.5-1.0);
- 3) Transferiu-se 2 ml da cultura para 50 ml de meio LB, em um Erlenmeyer estéril de 250 ml;
- 4) Incubou-se o inoculo a 28°C, com agitação (100-150 rpm), por 12 a 16 horas (A_{600} entre 0.5-1.0);
- 5) Transferiu-se a cultura para um tubo e incubando-a no gelo por 15 min.
- 6) As células foram centrifugadas a 3000 rpm, por 15 min a 4°C;
- 7) Descartou-se o sobrenadante e delicadamente ressuspendeu-se (com micropipeta) as células em 1 ml de CaCl₂ 20 mM (mantido a 4°C);
- 8) Distribuiu-se a suspensão de células em alíquotas de 100 μ l, em tubos de microcentrífuga previamente resfriados.

8.10) Transformação do vetor binário em células competentes de *A. tumefaciens*

- 1) Adicionou-se aproximadamente 1 μ g de DNA plasmidial por alíquota de 100 μ l de células competentes e misturou-se delicadamente.

- 2) Incubou-se a o tubo no gelo por 30 min;
- 3) Incubaram-se as células em nitrogênio líquido até congelamento (cerca de 2 min);
- 4) Transferiu-se imediatamente para 37°C e incubou-se por 5 min;
- 5) Adicionou-se 1 ml de meio LB, homogeneizou-se delicadamente com a micropipeta e incubou-se por 2 horas a 28°C;
- 6) Plaqueou-se 150 µl da suspensão de células em meio LB ágar contendo rifampicina (100 mg/l), estreptomicina (50 mg/l) e canamicina (100 mg/l).
- 7) Incubou-se a 28°C até o crescimento de colônias isoladas (2 dias). Coletaram-se algumas colônias transformadas para conferir a integridade do vetor transferido.

8.11) Crescimento de plântulas de tabaco *in vitro*

Sementes de tabaco selvagem foram esterilizadas pelo seguinte processo:

- 1) Incubação das sementes em álcool 70% por 1 min;
- 2) Lavagem em água deionizada autoclavada por 3 vezes;
- 3) Incubação em hipoclorito de sódio 50% com Tween-20 por 20 min;
- 4) Lavagem em água deionizada autoclavada por 5 vezes;
- 5) Semeadura em placa com meio Murashige & Skoog (MS) a 4.4 g/l, pH 5.7 e phytigel a 2.29 g/l.

As plântulas foram transferidas isoladamente para germiboxes maiores após algumas semanas de germinação.

8.12) Elementos *cis*-regulatórios putativos do promotor EGLV1 – base PLACE

Elementos <i>cis</i>	Repetições	Elementos <i>cis</i>	Repetições
-10PEHVPSBD	4	MYBCOREATCYCB1	1
-300ELEMENT	2	MYBST1	1
2SSEEDPROTBANAPA	1	MYCATERD1	2
ABRELATERD1	1	MYCATRD22	2
ABRERATCAL	1	MYCCONSENSUSAT	10
ACGTATERD1	4	NODCON1GM	3
ANAERO3CONSENSUS	1	NODCON2GM	9
ARFAT	1	OSE1ROOTNODULE	3
ARR1AT	17	OSE2ROOTNODULE	9
BIHD1OS	5	P1BS	2
CAATBOX1	16	POLASIG1	6
CACGTGMOTIF	2	POLASIG2	2
CACFTTPPCA1	12	POLASIG3	2
CANBNNAPA	2	POLLENILELAT52	11
CAREOSREPI	1	PREATPRODH	1
CCA1ATLHCB1	2	PRECONSCRHSP70A	1
CCAATBOX1	5	PROLAMINBOXOSGLUB1	1
CPBCSPOR	2	PYRIMIDINEBOXOSRAMY1A	1
CURECORECR	8	RAV1AAT	1
DOFCOREZM	24	REALPHALGLHCB21	2
E2FCONSENSUS	1	ROOTMOTIFTAPOX1	9
EBOXBNNAPA	10	RYREPEATBNNAPA	2
EECCRCAH1	4	RYREPEATLEGUMINBOX	1
ELRECOREPCR1	1	SEBFCONSSTPR10A	2
ERELEE4	1	SEF1MOTIF	1
EVENINGAT	1	SEF4MOTIFGM7S	1
GATABOX	9	SORLIP1AT	1

GTICONSENSUS	11	SREATMSD	1
GTIGMSCAM4	2	SURECOREATSULTR11	3
GTGANTG10	9	TAAAGSTKST1	4
IBOX	2	TATABOX4	1
IBOXCORE	4	TATABOX5	6
IBOXCORENT	2	TATABOXOSPAL	1
INRNTPSADB	1	TBOXATGAPB	1
INTRONLOWER	2	TELOBOXATEEF1AA1	1
LIBOXATPDF1	1	UP2ATMSD	1
LTRE1HVBLT49	2	WBOXATNPR1	2
MYB1AT	4	WBOXHVIS01	1
MYB2CONSENSUSAT	1	WBOXNTERF3	3
MYBATRD22	1	WRKY71OS	8
MYBCORE	1		

9) Referências

- AHN, S.J.; SHIN, R.; SCHACHTMAN, D.P. Expression of KT/KUP Genes in Arabidopsis and the role of root hairs in K⁺ Uptake. **Plant Physiol.**, v.134, n.3, p.1135-1145, mar. 2004.
- ALTSCHUL, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res.**, v.25, p.3389-3402, 1997.
- ARAGÃO F.J.L. *et al.* Metodologias para transformação genética de Plantas-Modelo. Circular Técnica 15, ISSN 1516-4349, **EMBRAPA**, Brasília, DF, 2002.
- ARGÜELLO-ASTORGA, G.R.; HERRERA-ESTRELLA, L.R. Ancestral multipartite plant promoters have structural features correlating with specific phototransduction pathways. **Plant Physiol.**, v.112, p.1151–1166, 1996.
- ARKHIPOVA, I.R. Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. **Genetics**, v.139, p.1359-1369, 1995.
- AYRE, B.G.; BLAIR, J.E.; TURGEON, R. Functional and Phylogenetic Analyses of a conserved regulatory program in the phloem of minor veins. **Plant Physiol.**, v.133, p.1-11, 2003.
- BENDTSEN, J.D. *et al.* Improved prediction of signal peptides: SignalP 3.0. **J. Mol. Biol.**, v.340, p.783-795, 2004.
- BERNIER, F.; BERNA, A. Germins and germin-like proteins: Plant do-all proteins. But what do they do exactly? **Plant Physiol. Biochem.**, v.39(7-8), p.545-554, 2001.
- BRACELPA – Associação Brasileira de Celulose e Papel. Relatório de Desempenho do Setor de Celulose e Papel. Projeções 2007.

BRANDALISE, M. **Isolamento e caracterização de promotores tecido específico de raiz e folha de *Coffea arabica***. 2007. 141f. Tese (Doutorado) – Instituto de Biociências, Universidade Estadual Paulista, Botucatu.

BREATHNACH, R.; CHAMBON, P. Organization and expression of eukaryotic split genes coding for proteins. **Annu. Rev. Biochem.**,v.50, p.349-383, 1981.

BUCHER, P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. **J. Mol. Biol.**, v.212, p.563-578, 1990.

BURKE, T.W.; KADONAGA, J.T. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. **Genes & Dev.**,v.10, p.711-724, 1996.

BURKE, T.W.; KADONAGA, J.T. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAF_{II}60 of *Drosophila*. **Genes & Dev.**, v.11, p.3020-3031, 1997.

BUTLER, J.E.F.; KADONAGA, J.T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. **Genes & Dev.**, v.16, p.2583-2592, 2002.

CALHOUN, V.C.; STATHOPOULOS, A.; LEVINE, M. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. **Proc. Natl. Acad. Sci. USA**,v.99, p.9243-9247, 2002.

CAREY, M.; SMALE, S.T. Transcription regulation in eukaryotes: concepts, strategies, and techniques. **Cold Spring Harbor Laboratory Press**, Cold Spring Harbor, Nova Yorque, 2000.

- CARTER, C.; THORNBURG, R.W. Purification and characterization as a germin-like, manganese superoxide dismutase implicated in the defense of floral reproductive tissues. **J. Biol. Chem.**, v.275, n.47, p. 36726–36733, 2000.
- CHANG, S.; PURYEAR, J.; CAIRNEY, J. A simple and efficient method for isolating RNA from pine trees. **Plant Mol. Rep.**, v.11, p.113-116, 1993.
- CHAUMONT, F. *et al.* Aquaporins constitute a large and highly divergent protein family in maize. **Plant Physiol.**, v.125(3), p.1206-1215, 2001.
- CHEN, K.; RAJEWSKY, N. The evolution of gene regulation by transcription factors and microRNAs. **Nature**, v.8, p.93-103, 2007.
- CHUNG, Y.D. *et al.* Identification of ovarian enhancer-binding factors which bind to ovarian enhancer 1 of the *Drosophila* genes yp1 and yp2. **Mol. Gen. Genet.**, v.251, p.347-351, 1996.
- COLLART, M.A.; STRUHL, K. NOT1 (CDC39), NOT2 (CDC36), NOT3, e NOT4 encode a global-negative regulator of transcription that differentially affects TATA-element utilization. **Genes & Dev.**, v.8, p.525-537, 1994.
- CORDEN, J., *et al.* Promoter sequences of eukaryotic protein-coding genes. **Science**, v.209, p.1406-1414, 1980.
- DEVIC M., *et al.* Efficient PCR walking on plant genomic DNA. **Plant Physiol. and Biochem.**, v.35, p.1-9, 1997.
- DILEONE, R.J.; RUSSELL, L.B.; KINGSLEY, D.M. An extensive 3' regulatory region controls expression of Bmp5 in specific anatomical structures of the mouse embryo. **Genetics**, v.148, p.401-408, 1998.
- DUMAS, B.; FREYSSINET, G.; PALLET, K.E. Tissue-specific expression of germin-like oxalate oxidase during development and fungal infection of barley seedlings. **Plant Physiol.**, v.107, p.1091-1096, 1995.

- DUNWELL, J.M. Cupins: a new superfamily of functionally diverse proteins that include germins and plant seed storage proteins. **Biotechnol. Genet. Engin. Rev.**, v.15, p.1–32, 1998.
- DUNWELL, J.M.; GANE, P.J. Microbial relatives of seed storage proteins: conservation of motifs in a functionally diverse superfamily of enzymes. **J. Mol. Evol.**, v.46, p.147–154, 1998.
- DUNWELL, J.M.; KHURI, S.; GANE, P.J. Microbial relatives of the seed storage proteins of higher plants: conservation of structure, and diversification of function during evolution of the cupin superfamily. **Microbiol. Mol. Biol. Rev.**, v.64, p.153–179, 2000.
- DUNWELL, J.M. *et al.* Evolution of functional diversity in the cupin superfamily. **Trends Biochem. Sci.**, v.26, p.740–745, 2001.
- DUNWELL, J.M.; PURVIS, A.; KHURI, S. Cupins: the most functionally diverse protein superfamily? **Phytochemistry**, v.65, p.7-17, 2004.
- FAN, Z. *et al.* Cloning and expression analysis of Zmglp1, a new germin-like protein gene in maize. **Biochemical and Biophysical Research Communications**, v.331, p.1257-1263, 2005.
- FERREIRA, E.M.; GRATTAPAGLIA, D. Extração de DNA genômico total de plantas. Em: Introdução ao uso de marcadores moleculares em análise genética. **Embrapa**, p.121-139, 3^a ed., 1998.
- FILLATTI, J.J. *et al.* Agrobacterium-mediated transformation and regeneration of Populus. **Mol. Gen. Genet.**, v.206, p.192-199, 1987.
- GANE, P.J.; DUNWELL, J.M.; WARWICKER, J. Homology modeling based on the structure of vicilins predicts a histidine cluster in the active site of oxalate oxidase. **J. Mol. Evol.**, v.46, p.488-493, 1998.

- GIERTH, M.; MASER, P. Potassium transporters in plants – Involvement in K⁺ acquisition, redistribution and homeostasis. **FEBS Letters**, v.581, p.2348-2356, 2007.
- GOICOECHEA, M. *et al.* EgMYB2, a new transcription activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. **The Plant J.**, v.43, p.553-567, 2005.
- GOLDBERG, M.L. Sequence analysis of *Drosophila* histone genes. Ph.D.dissertation, **Stanford University**, Califórnia, 1979.
- GONZÁLEZ, E. R. *et al.* Production of transgenic *Eucalyptus grandis* X *E. urophylla* using the sonication-assisted Agrobacterium transformation (SAAT) system. **Funct. Plant Biol.**, v.29, p.97-102, 2002.
- GREEN, J. *et al.* Analysis of the expression patterns of the *Arabidopsis thaliana* tubulin-1 and *Zea mays* ubiquitin-1 promoters in rice plants in association with nematode infection. **Physiological and Molecular Plant Pathology**, v.60, p.197-205, 2002.
- GREGÓRIO, A. D.; CORBO, J.C.; LEVINE, M. The regulation of forkhead/HNF-3 beta expression in the *Ciona* embryo. **Dev. Biol.**, v.229, p.31-43, 2001.
- GRIFFITHS, A.J.F., *et al.* Regulation of gene transcription In: Modern Genetic Analysis. 3rd edition, **W.H. Freeman and Company**, New York, NY, 2000.
- GRIFFITHS, A.J.F. *et al.* Introdução à genética. **Editores Guanabara Koogan S.A.**, 8 ed, 2006.
- GUPTA, R.; JUNG, E.; BRUNAK, S. Prediction of N-glycosylation sites in human proteins. In preparation – www.cbs.dtu.dk/services/netnglyc, 2004.
- HEINTZEN, C. *et al.* Circadian oscillations of a transcript encoding a germin-like protein that is associated with cell walls in young leaves of the long-day plant *Sinapis alba* L. **Plant Physiol.**, v.106, p.905-915, 1994.

- HIGO, K. *et al.* Plant cis-acting regulatory DNA elements (PLACE) database. **Nucleic Acids Res.**, v.27, p.297-300, 1999.
- HILL, T.A. *et al.* Discrete spatial and temporal cis-acting elements regulate transcription of the *Arabidopsis* floral homeotic gene APETALA3. **Development**, v.125, p.1711-1721, 1998.
- HO, C.K. *et al.* Agrobacterium tumefaciens mediated transformation of *Eucalyptus camaldulensis* and production of transgenic plants. **Plant Cell Rep.**, v.17, p.675-680, 1998.
- HOCHHEIMER, A., TJIAN, R.. Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. **Genes & Dev.**, v.17, p.1309-20, 2003.
- HOSHINO, A.A. **Isolamento e Caracterização de Promotores Tecido-específicos a partir das Informações do SUCEST (“Sugarcane Expressed Sequence Tags”)**.2007.143f. Tese (Doutorado) – Instituto de Biociências – Universidade Estadual Paulista, Botucatu.
- HUANG, N. *et al.* The tissue-specific activity of a rice beta-glucanase promoter (*Gns9*) is used to select rice transformants. **Plant Sci.**, v.161, p.589-595, 2001.
- HULTMARK, D.; KLEMENZ, R.; GEHRING, W. J. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. **Cell**, v.44, p.429-438, 1986.
- IYER, V. R. *et al.* Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. **Nature**, v.409, p.533-538, 2001.
- JAVAHERY, R. *et al.* DNA sequence requirements for transcriptional initiator activity in mammalian cells. **Mol. Cell. Biol.**, v.14, p.116-127, 1994.

JEFFERSON, R. A.; KAVANAGH, T. A.; BEVAN, M. W. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. **EMBO J.**, v.6(13), p.3901-3907, 1987.

KAMMANDEL, B. *et al.* Distinct cis-essential modules direct the time-space pattern of the Pax6 gene activity. **Dev. Biol.**, v.205, p.79-97, 1999.

KANZAKI, H. *et al.* Overexpression of the wasabi defensin gene confers enhanced resistance to blast fungus (*Magnaporthe grisea*) in transgenic rice. **Theor. Appl. Gen.**, v.105, p.809-814, 2002.

KAYO, T. Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys. **Proc. Natl. Acad. Sci. USA**, v.98, p.5093-5098, 2001.

KAZAN, K. *et al.* Expressions of a pathogenesis-related peroxidase of *Stylosanthes humilis* in transgenic tobacco and canola and its effect on disease development. **Plant Sci.**, v.136, p.207-217, 1998.

KONIECZNY, A.; AUSUBEL, F.M. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based makers. **The Plant J.**, v.4, n.2, p.403-410, 1993.

KORIMBOCUS, J. *et al.* Improved detection of Sugarcane yellow leaf virus using a real-time fluorescent (TaqMan) RT-PCR assay. **J. Virol. Meth.**, v.103, p.109-20, 2002.

KRAUSKOPF, E.; HARRIS, P. J.; PUTTERIL, J. The cellulose synthase gene *PrCESA10* is involved in cellulose biosynthesis in developing tracheids of the gymnosperm *Pinus radiata*. **Gene**, v.350(2), p.107-116, 2005.

KUTACH, A.K.; KADONAGA, J.T. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. **Mol. Cell. Biol.**, v.20, p.4754-4764, 2000.

- LACORTE, C.; ROMANO, E. Capítulo 6: Transferência de vetores para *Agrobacterium*. Em: BRASILEIRO, A.C.M. & CARNEIRO, V.T.C. Manual de Transformação Genética de Plantas. **EMBRAPA** – Brasília, DF, 1998.
- LAGRANGE, T., *et al.* New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. **Genes & Dev.**, v.12, p.34-44, 1998.
- LANE, B.G. Cellular desiccation and hydration: developmentally regulated proteins, and the maturation and germination of seed embryos. **FASEB J.**, v.5:, p.2893-2901, 1991.
- LANE, B.G. *et al.* Germin isoforms are discrete temporal markers of wheat development. **Eur. J. Biochem.**, v.209, p.961-969, 1992.
- LANE, B.G. Oxalate, germin, and the extracellular matrix of higher plants. **FASEB J.**, 8: 294-301, 1994.
- LANE, B.G. Oxalate oxidase and differentiating surface structure in wheat: germins. **Biochem. J.**, v.49, p.309-321, 2000.
- LAUVERGEAT, V. *et al.* The vascular expression pattern directed by the *Eucalyptus gunnii* cinnamyl alcohol dehydrogenase EgCAD2 promoter is conserved among woody and herbaceous plant species. **Plant Mol. Biol.**, v.50, p.497-509, 2002.
- LEITE, S.M.M. **Variabilidade genética em uma população-base de *Eucalyptus grandis* Hill ex-Maiden através do marcador RAPD (polimorfismo de DNA amplificado ao acaso)**. 1998. 84f.Dissertação (Mestrado) – Faculdade de Ciências Agrônomicas, Universidade Estadual Paulista, Botucatu.
- LEMON, B.; TJIAN, R. Orchestrated response: a symphony of transcription factors for gene control. **Genes & Dev.**, v.14, p.2551-2569, 2000.

- LETTICE, L.A. *et al.* Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. **Proc. Natl. Acad. Sci. USA**, v.99, p.7548-7553, 2002.
- LEVY, M.; EDELBAUM, O.; SELA, I. Tobacco mosaic virus regulates the expression of its own resistance gene N¹. **Plant Physiol.**, v.135, p1-6, 2004.
- LEWIN, B. Genes VII, **Oxford University Press**, Oxford, 2000.
- LI, L. *et al.* Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. **Proc. Natl. Acad. Sci. USA**, v.100, p.4939-4944, 2003.
- LI, W.W. *et al.* Common genetic variation in the promoter of the human apo CIII gene abolishes regulation by insulin and may contribute to hypertriglyceridemia. **J. Clin. Invest.**, v.96, p.2601-2605, 1995.
- LIU, W. *et al.* Characterization of two HKT1 homologues from *Eucalyptus camaldulensis* that display intrinsic Osmosensing capability. **Plant Physiol.**, v.127, p.283-294, 2001.
- LIVAK, K.; SCHMITTGEN, T.D. Analysis of Relative Gene Expression Data Using Real Time Quantitative PCR and the 2^{- $\Delta\Delta C_t$} Method. **Methods**, v.25, p.402-408, 2001.
- LO, K.; SMALE, S.T. Generality of a functional initiator consensus sequence. **Gene**, v.182, p.13-22, 1996.
- LUDWIKOW, A.; GALLOIS, P.; SADOWSKI, J. Ozone-Induced oxidative stress response in *Arabidopsis*: transcription profiling by microarray approach. **Cell Mol. Biol. Lett.**, v.9(4b), p.829-842, 2004.
- MASER, P. *et al.* Phylogenetic relationships within cation transporter families of *Arabidopsis*. **Plant Physiol.**, v.35, p.459-469, 1997.
- MEMBRÉ, N. cDNA sequence, genomic organization and differential expression of three *Arabidopsis* genes for germin/oxalate oxidase-like proteins. **Plant Mol. Biol.**, v.35, p.459-469, 1997.

- MERKLE, S. A.; NAIRN J. Hardwood Tree Biotechnology. **In Vitro Cell. Dev. Biol. Plant**, v.41, p.602–619, 2005.
- MODY, M., *et al.* Genomewide gene expression profiles of the developing mouse hippocampus. **Proc. Natl. Acad. Sci. USA**, v.98, p.8862–8867, 2001.
- MOQTADERI, Z., *et al.* TBP-associated factors are not generally required for transcriptional activation in yeast. **Nature**, v.383, p.188-191, 1996.
- MOQTADERI, Z.; KEAVENEY, M.; STRUHL, K. The histone H3-like TAF is broadly required for transcriptional in yeast. **Mol. Cell**, v.2, p.675-682, 1998.
- MOREAU, C. *et al.* A genomic approach to investigate developmental cell death in woody tissues of *Populus* trees. **Genome Biol.**, v.6, p.R34, 2005.
- MULLINS, K.V. *et al.* Regeneration and transformation of *Eucalyptus camaldulensis*. **Plant Cell Rep.**, v.16, p.787-791, 1997.
- MURASHIGE T.; SKOOG F. A revised medium for rapid growth and bioassays with tobacco cultures. **Physiol. Plant**, v.15, p.473-497, 1962.
- NAGANAWA, S. *et al.* Intestinal transcription and synthesis of apolipoprotein AI is regulated by five natural polymorphisms upstream of the apolipoprotein CIII gene. **J. Clin. Invest.**, v.99, p.1958-1965, 1997.
- NEUTEBOOM, L.W. *et al.* Characterization and tissue-regulated expression of genes involved in pineapple (*Ananas comosus* L.) root development. **Plant Sci.**, v.163, p.1021-1035, 2002.
- NEZNANOV, N.; UMEZAWA, A.; OSHIMA, R.G. A regulatory element within a coding exon modulates keratin 18 gene expression in transgenic mice. **J. Biol. Chem.**, v.272, p.27549-27557, 1997.
- OHLER, U.; NIEMANN, H. Identification and analysis of eukaryotic promoters: recent computational approaches. **Trends in Genetics**, v.17, n.2, p.56- 60, 2001.

- OHMIYA, A. Characterization of ABP19/20, sequence homologues of germin-like protein in *Prunus persica* L. **Plant Science**, v.163, p.683-689, 2002.
- OHTSUKI, S.; LEVINE, M.; CAI, H.N. Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. **Genes & Dev.**, v.12, p.547-556, 1998.
- OKAMOTO, M. *et al.* High-Affinity Nitrate Transport in Roots of Arabidopsis Depends on Expression of the NAR2-Like Gene AtNRT3.1. **Plant Physiol.**, v.140, p.1036-46, 2006.
- ONO, M. Transient increase in the level of mRNA for a germin-like protein in leaves of short-day plant *Pharbitis nil* during the photoperiodic induction of flowering. **Plant Cell Physiol.**, v.37, n.6, p. 855-861, 1996.
- PAOLIS A. D., *et al.* A rolB regulatory factor belongs to a new class of single zinc finger plant proteins. **Plant J**, v.10: p.215-223, 1996.
- PETRONZELLI, F. *et al.* Polymorphism in the upstream regulatory region of DQA1 gene in the Italian population. **Tissue Antigens**, v.45, p.258-263, 1995.
- PAUX, E. *et al.* Identification of genes preferentially expressed during wood formation in Eucalyptus. **Plant Mol. Biol.**, v.55(2), p.263-80, 2004.
- PRIBNOW, D. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. **Proc. Natl. Acad. Sci.**, v.72, p.784-788, 1975a.
- PRIBNOW, D. Bacteriophage T7 early promoters: Nucleotide sequences of two RNA polymerase binding sites. **J. Mol. Biol.**, v.99: 419-443, 1975b.
- PURNELL, B.A.; EMANUEL, P.A.; GILMOUR, D.S. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. **Genes & Dev.**, v.8: p.830-842, 1994.

- RANCE, I. *et al.* Combination of viral promoter sequences to generate highly active promoters for heterologous therapeutic protein over-expression in plants. **Plant Sci.** v.162, p.833-842, 2002.
- ROZEN, S.; SKALETSKY, H.J. Primer3 on the WWW for general users and for biologist programmers. Em: Krawetz S., Miserner S. (eds). **Bioinformatics Methods and Protocols: Methods in Molecular Biology.** Humana Press, p.365-86, 2000.
- RUBIO, F.; SANTA-MARIA, G. E.; RODRIGUEZ-NAVARRO, A. Cloning of Arabidopsis and barley cDNAs encoding HAK potassium transporters in root and shoot cells. **Physiol. Plant**, v.109, p.34-43, 2000.
- SANDRELLI, F. *et al.* Molecular dissection of the 5' region of no-on-transientA of *Drosophila melanogaster* reveals cis-regulation by adjacent dGpi1 sequences. **Genetics**, v.157, p.765-775, 2001.
- SACKERSON, C.; FUJIOKA, M.; GOTO, T. The even-skipped locus is contained in a 16-Kb chromatin domain. **Dev. Biol.**, v.211, p.39-52, 1999.
- SARTORETTO, L.M.; BRASILEIRO, A.C.M.; BARREUTO CID, L.P. Biolistic transformation of *Eucalyptus grandis* X *E. urophylla* callus. **Funct. Plant Biol.**, 29: 917-924, 2002.
- SÉGUIN, A.; LAPOINTE, G.; CHAREST, P.J. Chapter 8: Transgenic Trees. Em: Bruce A. & Palfreyman J. W. Forest Products Biotechnology. **Taylor & Francis Ltd.**, p.287-303, 1998.
- SERRANO, R.; CULIANZ-MACIÁ, F.; MORENO, V. Genetic engineering of salt and drought tolerance with yeast regulatory genes. **Scientia Horticulturae**, v.78, p.261-269, 1999.
- SHEN, Y.G. *et al.* AhCMO, regulated by stresses in *Atriplex hortensis*, can improve drought tolerance in transgenic tobacco. **Theor. Appl. Genet.**, v.105, p.815-821, 2002.

- SMALE, S.T. *et al.* The initiator element: A paradigm for core promoter heterogeneity within metazoan protein-coding genes. **Cold Spring Harb. Symp. Quant. Biol.**, v.58, p.21-31, 1998.
- STAIGER, D.; KLAUS, A.; TREPP, G. The Atger3 promoter confers circadian clock-regulated transcription with peak expression at the beginning of the night. **Plant Mol. Biol.**, v.40, p.873-882, 1999.
- STOUGAARD, J. *et al.* 5' analysis of the soybean leghaemoglobin lbc3 gene: regulatory elements required for promoter activity and organ specificity. **EMBO J.**, v.6, p.3565-3569, 1987.
- STRUHL, K. Constitutive and inducible *Saccharomyces cerevisiae* promoters: Evidence for two distinct molecular mechanisms. **Mol. Cell. Biol.**, v.6, p.3847-3853, 1986.
- STRUHL, K. Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. **Cell**, v.49, p.295-297, 1987.
- SWART, S. *et al.* Purification and partial characterization of a glycoprotein from pea (*Pisum sativum*) with receptor activity for rhicadhesin, an attachment of *Rhizobiaceae*. **Plant Mol. Biol.**, v.24, p.171-183, 1994.
- TOURNIER, V. *et al.* An efficient procedure to stably introduce genes into an economically important pulp tree (*Eucalyptus grandis* X *E. urophylla*). **Transgenic Res.**, v.12, p.403-411, 2003.
- VALLELIAN-BINDSCHEDLER, L. *et al.* Structure, expression and localization of a germin-like protein in barley (*Hordeum vulgare* L.) that is insolubilized in stressed leaves. **Plant Mol. Biol.**, v.37, p.297-308, 1998a.

- VALLELIAN-BINDSCHEDLER, L. *et al.* Heat-induced resistance in barley to powdery mildew (*Blumeria graminis* f. sp. *hordei*) is associated with burst of active oxygen species. **Physiol. Mol. Plant Pathol.**, v.52, p.185-199, 1998b.
- VICENTE-CARBAJOSA, J. *et al.* A maize zinc-finger protein binds the prolamins box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. **Proc. Nat. Acad. Sci.**, v.94, p.7685-7690, 1997.
- VICENTINI, R., *et al.* *In silico* evaluation of the *Eucalyptus* transcriptome. **Genet. and Mol. Biol.**, v.28, n.3 (suppl), p.487-495, 2005.
- WEI, Z. *et al.* Characterization of the SpHE promoter that is spatially regulated along the animal-vegetal axis of the sea urchin embryo. **Dev. Biol.**, v.171, p.195-211, 1995.
- WEINZIERL, R.O.J. Mechanisms of gene expression. **Imperial College Press**, Londres, 1999.
- WEST, R.J.; YOCUM, R.; PTASHNE, M. *Saccharomyces cerevisiae* GAL1-GAL10 divergent promoter region: location and function of the upstream activating sequence UAS_G. **Mol. Cell. Biol.**, v.4, p.2467-2478, 1984.
- WHITE, K. P. *et al.* Microarray analysis of *Drosophila* development during metamorphosis. **Science**, v.286, p.2179-2184, 1999.
- WRAY, G.A. *et al.* The evolution of transcriptional regulation in Eukaryotes. **Mol. Biol. Evol.**, v.20, n.9, p.1377-1419, 2003.
- WRIGHT, C.E., *et al.* *In vivo* regulation of β -MGC gene in rodent heart: role of T₃ and evidence for an upstream enhancer. **Am. J. Physiol.**, v.276, p.C883-C891, 1999.
- YAMAHARA, T. *et al.* Isolation of a germin-like protein with manganese superoxide dismutase activity from cells of a moss, *Barbula unguilata*. **J. Biol. Chem.**, v.274, p.33274-33278, 1999.

YANAGISAWA, S.; SCHMIDT, R.J. Diversity and similarity among recognition sequences of Dof transcription factors. **The Plant J.**, v.17, p.209-214, 1999.

YANAGISAWA, S.; SHEEN, J. Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. **Plant Cell**, v.10, p.75-89, 1998.

YU, H.; YANG, S.H.; GOH, C.J. Spatial and temporal expression of the orchid floral homeotic gene *DOMADS1* is mediated by its upstream regulatory regions. **Plant Mol. Biol.**, v.49, p.225-237, 2002.

YUH, C.H.; BOLOURI, H.; DAVIDSON, E.H. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin. **Science**, v.279, p.1896-1902, 1998.

YUH, C.H.; BOLOURI, H.; DAVIDSON, E.H. Cis-regulatory logic in the *endo16* gene: switching from a specification to a differentiation mode of control. **Development**, v.128, p.617-629, 2001.

YUH, C.H. *et al.* Patchy interspecific sequence similarities efficiently identify positive cis-regulatory elements in the sea urchin. **Dev. Biol.**, v.246, p.148-161, 2002.

ZERUCHA, T. *et al.* A highly conserved enhancer in the *Dlx5/Dlx6* region is the site of cross-regulatory interactions between *Dlx* genes in the embryonic forebrain. **J. Neuro.**, v.20, p.709-721, 2000.

ZHENG, Z.; MURAI, N. A distal promoter regions of the rice seed storage protein glutelin gene enhanced quantitative gene expression. **Plant Science**, v.128, p.59-65, 1997.