UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS

***CAMPUS* DE BOTUCATU**

Tânia Mara Bortoloto

Caracterização de uma Região Genômica Relacionada à Característica Lignotuber em Eucalipto.

Botucatu-SP

2011

Tânia Mara Bortoloto

Caracterização de uma Região Genômica Relacionada à Característica Lignotuber em Eucalipto.

Tese apresentada ao Instituto de Biociências, Universidade Estadual Paulista "Júlio de Mesquita Filho" – Campus de Botucatu, para a obtenção do título de Doutor em Ciências Biológicas – Área de Concentração Genética.

Orientador: Prof. Dr. Celso Luís Marino

Botucatu-SP

2011

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO DE AQUIS. E TRAT. DA INFORMAÇÃO
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ***ROSEMEIRE APARECIDA VICENTE***

Bortoloto, Tânia Mara.
Caracterização de uma região genômica relacionada à característica lignotuber em eucalipto / Tânia Mara Bortoloto. – Botucatu : [s.n.], 2011

Tese (doutorado) - Universidade Estadual Paulista, Instituto de Biociências de Botucatu
Orientador: Celso Luís Marino
Capes: 20203004

1. Melhoramento genético. 2. Eucalipto.

Palavras-chave: Estresse hídrico; Eucalipto; Lignotuber; Marcador molecular; Melhoramento genético.

AGRADECIMENTOS

A Deus por Sua Providência e pelo dom da vida.

Aos meus pais, que me ensinaram tudo o que sei e graças a eles me tornei quem sou hoje.

Ao meu irmão e amigo Alexandre pelo amor e amizade.

Ao Tomaz pelo amor, carinho, companheirismo, respeito e amizade.

Ao Prof. Celso Luiz Marino pela confiança e oportunidade.

À companhia Suzano Papel e Celulose por fornecer o material vegetal e pela colaboração no trabalho, especialmente ao Shinitiro, Giziane e Esteban que participaram ativamente nas discussões dos resultados e no andamento dessa tese.

Aos colegas do laboratório CAGEM: Júlio, Vanusa, Cíntia,Virginia, Leonardo, Andréia, Helena, Suzana, Karine, Aletéa, Paola, Cecília, Lídia, Maurício, e todos os que por lá passaram.

À Juliana pelas horas de bate-papo e pela dedicação, companheirismo e auxílio nesse trabalho e especialmente por sua amizade.

A todos os colegas do laboratório BIOGEM e do departamento de genética.

Ao amigo Bonsai pelo auxilio nas etapas desse trabalho, especialmente nas discussões e análise dos dados.

Ao amigo Douglas que sempre esteve disposto a auxiliar nas discussões e análise dos dados.

A toda a minha família avós e avôs (em memória), tios, tias, primos... E também a minha nova família que ganhei de presente da vida, Sr. Tomaz, Luiza, Guilherme e Titinha pela amizade e pela força.

A todos os amigos que passaram em minha vida, que deixaram um pouquinho de si e levaram um pouquinho de mim.

A todos que de alguma forma contribuíram para a elaboração desse trabalho.

À CAPES pela bolsa concedida.

Dedico este trabalho às pessoas que lutam diariamente ao meu lado com fé, amor, alegria, determinação e coragem, tornando os meus dias mais felizes e bonitos. Aos meus pais Valter e Mara, ao meu irmão Alexandre e ao meu esposo, companheiro e eterno namorado, Tomaz. Sem vocês eu não seria nada!

RESUMO

BORTOLOTO, T.M. **Caracterização de uma região genômica relacionada à característica lignotuber em eucalipto.** 2011. **80f.** Tese (Doutorado) – Instituto de Biociências de Botucatu, Universidade Estadual Paulista "Júlio de Mesquita Filho", Botucatu, 2011.

O grupo CAGEN em parceria com a empresa Suzano de Papel e Celulose vem desenvolvendo projetos de identificação de características relacionadas ao estresse abiótico em eucalipto. A presença de lignotuber é uma característica que confere a essas espécies maior tolerância a níveis de estresse. Em 2006 foi realizado um experimento que demonstrou uma correlação entre o aumento dos níveis de estresse e o aumento da formação de lignotuber em eucalipto, até atingir a proporção de 3:1, indicando a ação de um gene com efeito principal na formação do lignotuber. Através da técnica de BSA (*Bulked Segregant Analyses*) em conjunto com marcadores RAPD (*Random Amplified Polymorphic DNA*) foi possível detectar um marcador molecular que permitiu a identificação de indivíduos não portadores do lignotuber em eucalipto. Esse marcador foi convertido em SCAR e passou a ser chamado de *ELig* (*Eucalyptus* – Lignotuber). Diante do exposto, esse trabalho teve como objetivo a caracterização dessa região genômica relacionada à presença de lignotuber em eucalipto. O marcador *ELig* foi eficiente em identificar indivíduos de espécies de eucalipto sem lignotuber, podendo ser incorporado a programas de seleção assistida por marcadores moleculares. As espécies *Eucalyptus grandis* e *E. urophylla* apresentaram uma única cópia da sequência *ELig* em seus genomas. Essa sequência mostrou identidade a um *contig de EST* de eucalipto de 1918 nucleotídeos do banco de dados FORESTs-FAPESP, que apresenta o domínio de

proteínas *DnaJ*, uma grande família de proteínas com membros relacionados à resposta da planta contra estresse abiótico, incluindo proteínas *heat shock* e chaperonas. A identidade do *ELig* ao *contig* restringe-se à região 5' do *ELig*, nessa região não foi possível identificar o polimorfismo no sítio de pareamento dos *primers*, indicando que esse polimorfismo encontra-se na região 3' da sequência *ELig*. Através da técnica de *RACE 3'* (*Rapid Amplification of cDNA Ends*), foi possível verificar que a porção 3' do EST está completa. A sequência 3´ do *ELig*, que não apresenta identidade ao *contig*, apresenta-se após a região 3' *UTR* na proteína predita do *contig*. Com a análise do genoma completo do eucalipto, foi possível identificar no genoma da espécie uma região com identidade à sequência *ELig*, essa região também apresentou o domínio *DnaJ*. Análise nas proximidades dessa região permitiu identificar alguns genes candidatos para análises de expressão.

Palavras-chave: Estresse hídrico, Eucalipto, Lignotuber, Marcador Molecular, Melhoramento Genético.

ABSTRACT

CAGEN group in association with Suzano Pulp and Paper has developed projects to identify characteristics related to abiotic stress in Eucalyptus. Lignotuber presence is a resource which gives to these species higher levels of stress tolerance. An experiment was performed in 2006 which showed a correlation between increased stress levels and the increment of lignotuber formation, until 3:1 proportion, indicating the gene action with major effect on the lignotuber formation. Using BSA (Bulked Segregant Analyses) technique with RAPD markers (Random Amplified Polymorphic DNA) was detected a molecular marker that allowed the identification of individuals without lignotuber in Eucalyptus. This marker was converted into a SCAR (Sequenced Characterized Amplified Region) marker and was named Elig (Eucalyptus - Lignotuber). The purpose of this study is characterize this genomic region related to the lignotuber presence in Eucalyptus. The Elig marker was effective in identifying Eucalyptus without lignotuber, and it can be incorporated into marker-assisted selection programs. Eucalyptus grandis and E. urophylla presented a single copy sequence ELig in their genomes. This sequence showed identity with a eucalypt EST contig with1918 nucleotides on the FORESTs-FAPESP database, the contig displays the protein domain DnaJ, a large proteins family with members related to plant response against abiotic stress, this domain includes heat shock proteins and chaperones. The ELig identity with the contig was restricted in the 5 'ELig region, in which was not possible to identify the polymorphism in the primers pairing site, indicating that this polymorphism is found in the 3' ELig region sequence. Through the RACE 3' (Rapid Amplification of cDNA Ends) technique, we found that the 3' EST portion is complete. The 3’ Elig sequence

that has no contig identity, it is presented after the 3' UTR region of the predicted protein contig. By analyzing the Eucalyptus complete genome, was identified a region with sequence identity to ELig, this region also had the DnaJ domain. Analysis in the vicinity of this region identified some candidate genes for expression analysis.

Key words: Drought stress, Eucalyptus, Lignotuber, Molecular Marker, Genetic Improvement.

SUMÁRIO

RESUMO

ABSTRACT

LISTA DE TABELAS

LISTA DE FIGURAS

1. INTRODUÇÃO..... 16

2. OBJETIVOS..... 20

2.1. Geral..... 20

2.2. Específicos..... 20

3. MATERIAL E MÉTODOS..... 21

3.1. Material vegetal..... 21

3.2. Metodologias..... 21

3.2.1. Delineamento Experimental..... 21

3.2.2. Teste da eficiência do marcador *ELig*..... 22

3.2.3. Análise da região 5' da sequência *ELig*..... 25

3.2.4. Determinação do número de cópias por *Southern blot*..... 29

3.2.5. Técnica de *RACE 3'* (*Rapid Amplification of cDNA Ends*)..... 35

3.2.6. *Genome Walking*..... 39

3.2.7. Análises de Bioinformática..... 41

3.2.7.1. Comparação da Sequência *ELig* com banco de dados de EST..... 41

3.2.7.2. Comparação da Sequência *ELig* com o genoma completo de eucalipto..... 42

4. RESULTADOS E DISCUSSÃO..... 44

4.1 Análise da eficiência do uso do Marcador *ELig* em eucalipto..... 44

4.2. Análise das Sequências no banco de dados FOREST de eucalipto..... 48

4.3. Análise da região 5’ da sequência *ELig*..... 54

4.4. Número de cópias da região *ELig*..... 57

4.5. *RACE 3*`..... 59

4.6 *Genome Walking*……..... 61

4.7. Análise das sequências no *EucalyptusDB*..... 63

5. CONCLUSÕES..... 75

6. REFERÊNCIAS BIBLIOGRÁFICAS…… 76

LISTA DE TABELAS

Tabela 1: Teste da eficiência do marcador *ELig* com indivíduos de quatro espécies de eucalipto *E. grandis, E. urophylla, E. saligna* e *E. brassiana*..... 44

Tabela 2: Análise das proximidades da sequência com identidade ao *ELig* no genoma completo de *Eucalyptus grandis*..... 70

LISTA DE FIGURAS

Figura 1: Organograma com as principais etapas desenvolvidas ao longo do trabalho..... 22

Figura 2: *Primers* para amplificar a região 5' da sequência ELig. As barras azuis mostram a região em que as sequências apresentam identidade entre si..... 25

Figura 3: Esquema da metodologia *RACE* 3' (Kit 3´ *RACE System for Rapid Amplification of cDNA Ends* - Invitrogen®)……. 36

Figura 4: Esquema da metodologia *GenomeWalker*™ *Universal Kit* (Clontech Laboratories, Inc.)..... 39

Figura 5: Amplificação com o *primer ELig* em treze indivíduos *E. grandis* e treze *E. urophylla*..... 45

Figura 6: Esquema de ligação da marca *ELig* e o gene de efeito principal na formação do lignotuber em cruzamento de *E. urophylla* e *E. grandis* na produção de híbridos F2 de *E. urograndis*..... 47

Figura 7: Alinhamento da sequência do *singleton* com a sequência *ELig*..... 49

Figura 8: Alinhamento da sequência do *contig* com a sequência *ELig*..... 49

Figura 9: Identidade entre a predição de proteína do *contig* do EST com o domínio *DnaJ* do banco de dados de domínio de proteínas Pfam (Finn *et al.*, 2010)..... 51

Figura 10: Esquema da estrutura das proteínas com domínio *DnaJ* mostrando todos os seus *motifs* (adaptado de http://pfam.sanger.ac.uk/family/PF00226)..... 51

Figura 11: Gráfico com as avaliações realizadas em uma população F2 de *E. urograndis* submetida a estresse hídrico e nutricional, quanto a presença de lignotuber..... 54

Figura 12: Gel de agarose 1% com o produto da amplificação com as combinações de *primers* desenhados para amplificação da região final do *contig* e inicial do *ELig*..... 55

Figura 13: Alinhamento da sequência do *cluster* com a sequência *ELig* e com a sequência da amplificação do *primer EST-f* com o *primer EST-r1* (CP1)..... 56

Figura 14: Gel de agarose 1% com o produto de amplificação dos *primers EST-f* e *EST-r1* com DNA de quatro indivíduos de cada espécie *E. grandis*, *E. urophylla* e *E. saligna* 56

Figura 15: *Southern blot* da sequência *ELig* com o genoma de *E. grandis* e *E. urophylla*..... 58

Figura 16: Gel de agarose com os produtos de amplificação da técnica de RACE 3' utilizando os *primers* gene específicos (GSP) desenhados a partir da sequência *ELig*..... 59

Figura 17: Alinhamento das sequências do *contig* e *ELig* (ambas de *E. grandis*) com os fragmentos obtidos com a técnica de *RACE* 3'..... 60

Figura 18: Gel agarose 1% com o resultado da amplificação *nested* da técnica de *Genome Walking*..... 62

Figura 19: Alinhamento da sequência *ELig* com as sequências obtidas com a técnica de *Genome Walking* no sentido 5' (GW-5) e no sentido 3' (GW-3)..... 62

Figura 20: Alinhamento entre a região do genoma completo com identidade ao *ELig* e a sequência *ELig*..... 64

Figura 21: Alinhamento das sequências obtidas com a técnica de *Genome Walking* (*GW*), comparando-as com a sequência do genoma completo e a sequência *ELig*..... 65

Figura 22: Predição de ORFs a partir da sequência do *SUPER* do banco de dados *EucalyptusDB* que apresentou identidade à sequência *ELig*..... 66

Figura 23: Comparação das sequências do *SUPER* do banco de dados *EucalyptusDB* que apresentou identidade à sequência *ELig* com o *contig EST* e com a sequência *ELig*..... 66

Figura 24: Esquema do bloco gênico com os genes candidatos da vizinhança do *ELig*, usando uma distância de 100 Kb para cada sentido 5'e 3'..... 68

Figura 25: Genes candidatos do banco de dados de *Eucalyptus grandis* (*EucalyptusDB*) da vizinhança do *ELig*, usando uma distância de 100 Kb para cada sentido (5'e 3')..... 69

1. INTRODUÇÃO

O gênero *Eucalyptus* pertence à família Myrtaceae e inclui mais de 700 espécies, a maioria nativa da Austrália (Brooker, 2000 *apud* Grattapaglia e Kirst, 2008), com exceção de *E. urophylla* e *E. deglupta* que são originais do Timor e Papua Nova Guiné, respectivamente (Pryor, 1985). É um gênero florestal de interesse econômico plantado comercialmente nas zonas tropicais e subtropicais do globo. Hoje, o eucalipto assume grande importância na economia brasileira, sendo destinado principalmente à produção de celulose, papel e carvão vegetal utilizado nas siderúrgicas. O eucalipto também é muito utilizado na produção de lâminas, compensados, aglomerados, madeira serrada, móveis, além de outros produtos como o óleo essencial utilizado em produtos de limpeza, alimentos, perfumes e remédios e o mel produzido a partir do pólen de suas flores.

Segundo o relatório de março de 2011 da BRACELPA a produtividade média de madeira para celulose entre os anos de 2009 e 2010 sofreu um aumento de 24 para 44 m^3/ha/ano que representa um crescimento de 83%, e ainda indica que o país tem potencial para chegar à produtividade média de 70 m^3/ha/ano (BRACELPA, 2011). A produção de celulose apresentou um aumento de 5,6% do ano de 2009 para o ano de 2010, chegando a atingir a marca de 14 milhões de toneladas em 2010, marca que faz do país o quarto maior produtor de celulose do mundo. As exportações em 2010 chegaram a 6,8 bilhões de dólares com um saldo comercial de 4,9 bilhões, gerando 115 mil empregos diretos e 575 mil empregos indiretos (BRACELPA, 2011). Esse crescimento é devido a vários fatores ambientais como o clima e o solo e também a investimentos em pesquisa e desenvolvimento como a genética e a biotecnologia, além de

planejamentos sócio-ambientais, manejo florestal e organização do setor privado, dentre outros. A madeira utilizada na produção de papel é principalmente de eucalipto (81,6%) e pinus (17,2%) (BRACELPA, 2011).

No Brasil, os produtores de eucalipto sempre se preocuparam com a inclusão de espécies e populações diferentes do gênero, para aumentar a diversidade dessa cultura e isso possibilitou ao país possuir hoje um dos maiores bancos de germoplasma do gênero *Eucalyptus* do mundo, sendo o mais diverso fora dos locais de origem dessas espécies.

Em climas tropicais, como no Brasil, as espécies mais plantadas são *E. grandis*, *E. urophylla* e *E. saligna* (SBS, 2011). O híbrido *E. urograndis* originado do cruzamento entre as espécies *E. grandis* e *E. urophylla*, também é muito utilizado pelas empresas brasileiras. O eucalipto apresenta uma série de características que fizeram deste gênero a planta exótica mais plantada no Brasil, isso inclui capacidade de adaptação às condições ambientais e rápido crescimento.

Hoje, a cultura do eucalipto no Brasil está se expandindo intensamente para o norte e o nordeste do país, regiões onde os biomas predominantes possuem longos períodos de seca e os solos geralmente apresentam escassez de nutrientes. Desta forma, há a necessidade de identificar características adaptativas que aumentem as chances de sobrevivência das plantas nessas condições e introduzi-las no programa de melhoramento do gênero. Diante desse cenário, podemos destacar a tolerância a períodos prolongados de secas e a capacidade de armazenamento de nutrientes.

O Centro de Análises Genômicas (CAGEN) do Instituto de Biociências da UNESP de Botucatu - SP, em parceria com a empresa Suzano Papel e Celulose, vem desenvolvendo projetos que visam à busca e introdução de características silviculturais

em culturas de eucalipto, dando ênfase à resposta da planta a situações de estresses bióticos e abióticos. Dentre essas características, destacamos a presença de lignotuber.

Os lignotuberes são protuberâncias caulinares encontradas nas axilas dos cotilédones de algumas espécies vegetais, essas estruturas apresentam uma grande quantidade de parênquima de reserva de carboidratos e outros nutrientes, são formadas de inúmeras gemas dormentes que possuem um papel importante nos processos de rebrota após distúrbios como fogo ou corte (Bamber e Mullete,1978; Whittock *et al.*, 2003; Canadell e Lopes-Soria, 1998; Sampson, 1998). Após o corte do dossel, as gemas presentes no lignotuber são estimuladas devido à quebra da dominância apical (dos Reis e Reis, 1997). A origem dessa estrutura é controversa, mas acredita-se que sua fixação nas espécies de eucalipto possa estar relacionada à seleção natural causada pelo fogo das queimadas naturais que ocorrem nas florestas da Austrália, de onde o eucalipto é originado (Cheal *et al*. 1979 *apud* Wouters, 2004).

Essa estrutura está presente em quase todas as espécies do gênero *Eucalyptus* (95%) como é o caso de *E. urophylla*, *E. brassiana* e *E. saligna*. Outras espécies não apresentam o órgão, como é o caso de *E. grandis* (Eldridge, 1994).

A presença de lignotuber em eucalipto é influenciada pelo ambiente. Whittock *et al*. (2003) mostraram que houve uma menor formação de lignotuber nas plantas de *E. globulus* em ambiente mais úmido, indicando que o órgão é formado preferencialmente em condições de estresse hídrico.

Martins (2006) avaliou a segregação do caráter em uma população F2 de híbridos *E. urograndis*. Esta população F2 foi originada a partir da autofecundação de uma população F1 formada a partir do cruzamento entre um clone de *E. grandis* e um clone de *E. urophylla*. As plantas F2 foram submetidas a condições de estresse hídrico e

nutricional o que ocasionou um aumento na produção de lignotuber nos indivíduos. A proporção de plantas com lignotuber após o estresse foi de 3:1 (qui quadrado 2,42; p>0,1) indicando se tratar de uma característica com ação de um gene de efeito principal.

No mesmo trabalho, Martins (2006) utilizando marcadores *Random Amplified Polymorphic DNA* (RAPD) combinados com a técnica de *Bulked Segregant Analyses* (BSA), pôde identificar um marcador molecular com tamanho de 874nt (nucleotídeos) presente nos indivíduos da espécie *E. grandis* e nos híbridos *E. urograndis* que não apresentaram lignotuber. Essa marca foi convertida em *Sequence Characterized Amplified Region* (SCAR). Através da ferramenta BLASTn (*Basic Local Alignment Search Tool*) (Stephen *et al.*, 1997), a sequência SCAR foi utilizada como sonda no banco de sequências de ESTs (*Expressed Sequence Tag*) de eucalipto (FORESTs-FAPESP – *Eucalyptus Genome Sequencing Project Consortium* - Vicentini *et al.*, 2005). A sequência do marcador SCAR apresentou identidade a uma sequência de EST desse banco, indicando que esse marcador encontra-se numa região expressa do genoma de *Eucalyptus.*

2. OBJETIVOS

2.1. Geral

Caracterização de uma região genômica relacionada à presença de lignotuber em eucalipto.

2.2. Específicos

- Validar a eficiência do marcador *ELig* na identificação de indivíduos portadores de lignotuber em eucalipto.

- Determinar o número de cópias dessa região no genoma de *E. grandis* e *E. urophylla*.

- Determinar a sequência completa dessa região no genoma de eucalipto através de técnicas moleculares.

- Caracterização dessa região através da análise e comparação das sequências obtidas em bancos de dados.

3. MATERIAL E MÉTODOS

3.1. Material Vegetal

Todo o material vegetal utilizado neste trabalho foi fornecido pela empresa Suzano Papel e Celulose.

Folhas frescas de árvores das espécies *E. grandis*, *E. urophylla*, *E. saligna* e *E. brassiana* foram utilizadas para extração de DNA.

Para extração de RNA dos diferentes tecidos (raiz, caule e lignotuber) foram utilizadas plântulas de *E. grandis* e *E. urophylla*.

3.2. Metodologias

3.2.1. Delineamento Experimental

O organograma abaixo (Figura 1) ilustra os procedimentos principais realizados no decorrer do trabalho.

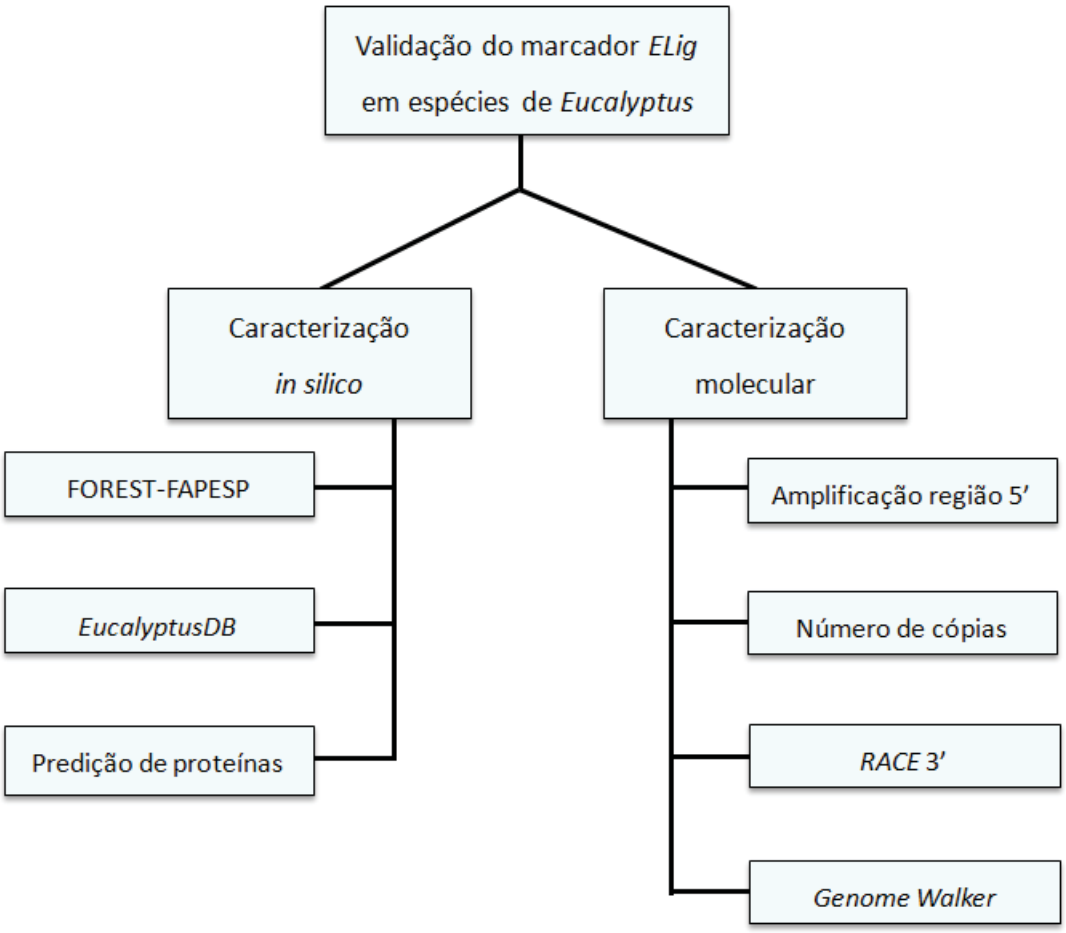

Figura 1: Organograma com as principais etapas desenvolvidas ao longo do trabalho.

3.2.2. Teste da eficiência do marcador *ELig*

Para o teste da eficiência do marcador *ELig* em identificar indivíduos de eucalipto com lignotuber, foram utilizados DNAs de 77 indivíduos da espécie *E. urophylla*, 69 indivíduos da espécie *E. grandis*, 10 indivíduos de *E. saligna* e oito indivíduos de *E. brasiana*. Os indivíduos que compõem essas espécies pertencem ao banco de germoplasma da empresa Suzano Papel e Celulose. Os eucaliptos que compõem esse material são considerados espécies puras, mas não se descarta a hipótese de existir indivíduos híbridos originários do próprio programa de melhoramento da

empresa. Algumas árvores podem manter características típicas de uma espécie, mas ser um híbrido, em diferentes graus, com outra espécie, isso pode ocorrer até mesmo nos locais de origem natural dos eucaliptos.

- Extração de DNA

Para o teste de eficiência do marcador, a extração de DNA foi realizada através do protocolo descrito por Ferreira e Grattapaglia (1995) com modificações:

Em um tubo foram macerados, aproximadamente, 150 mg de folhas com nitrogênio líquido, em seguida, foram adicionados 700 µl de solução de extração (2% CTAB (brometo de cetil trimetil amônio), 2% PVP (polivinilpirrolidona K 40), 100 mM Tris-HCl pH=8,0, 25 mM EDTA, 2 M NaCl (cloreto de sódio) e 0,2% de mercapto-etanol), aplicado em capela de exaustão e misturado até que todo o material macerado entrasse em contato com a solução. Esta mistura foi incubada à temperatura de 65 ºC por 45 a 60 min. Em seguida, foi realizada a primeira extração, adicionando-se 600µl de solução CIA (clorofórmio – álcool isoamílico 24:1). Esta solução foi misturada com o auxílio de vortex e os tubos levados à centrífuga refrigerada a 10 ºC, com rotação de 12.000 rpm durante 5 min.

Após a centrifugação, a solução apresentou três fases, a fase superior (aquosa) foi transferida para um novo tubo (±500 µl) e a esta solução foram adicionados 100 µl de CTAB 5% (5% CTAB, 2% PVP, 100 mM Tris-HCl pH=8,0, 25 mM EDTA, 2 M NaCl). Foi realizada então a segunda extração adicionando-se 600 µl de CIA. Essa solução foi homogeneizada em vortex e centrifugada nas mesmas condições anteriores. Após a centrifugação a solução apresentou duas fases separadas por uma película de

proteínas. A fase superior (± 500 µl) foi transferida para um novo tubo onde foi adicionado 400 µl de álcool isopropílico gelado. O material foi mantido à temperatura de -20 ºC por um período de aproximadamente 10 h. As amostras foram centrifugadas nas condições anteriores.

Após a centrifugação foi possível visualizar um precipitado no fundo dos tubos, o sobrenadante foi descartado e o precipitado lavado com 500 µl de etanol 95% por 2 min, o etanol foi descartado e, após secagem em centrífuga a vácuo, o precipitado foi eluído em 80µl de tampão TE (pH=8,0) contendo 10 ng/µl de RNAse.

As amostras foram levadas ao banho-maria a 37 ºC para a digestão do RNA por no mínimo 40 min ou até completa dissolução do precipitado. Os DNAs foram diluídos com água ultra pura na concentração de 50 ng/µl.

-Reações de PCR

Os DNAs extraídos foram submetidos à reação de PCR com os *primers ELig*. A reação (10 µl) foi composta de: 10 mM Tris-HCl, pH=8,0, 1,5 mM $MgCl_2$, 0,5 µM de cada *primer*, 1 mM de dNTP (Invitrogen), 25 ng de DNA genômico e 0.5 U de *Taq* DNA polimerase (Invitrogen). A amplificação com os *primers ELig* foi realizada em um termociclador *Mastercycler personal Eppendorf* ®, com a seguintes condições: 5 min a 94 ºC por 1 ciclo, 30 s a 94 ºC, 1 min a 64 ºC e 1 min a 72 ºC por 35 ciclos, com um ciclo final de 7 min a 72 ºC. Os produtos de amplificação foram separados em gel de agarose 1% corado com brometo de etídio.

3.2.3. Análise da região 5’ da sequência *ELig*

Para analisar a região 5’da sequência *ELig* e comparar com a sequência do *contig*, alguns *primers* foram desenhados para realização de uma reação de PCR, sendo que o *primer* senso foi desenhado na sequência do *contig* e os *primers* reversos foram desenhados na sequência genômica, possibilitando amplificar parte da sequência do *contig*. Foi desenhado um *primer* senso (*f*) e três *primers* reversos (*r1*, *r2* e *r3*) conforme a Figura 2. Essa análise possibilitou também a investigação da sequência do sítio de pareamento do *primer* RAPD na região 5’ do *ELig*.

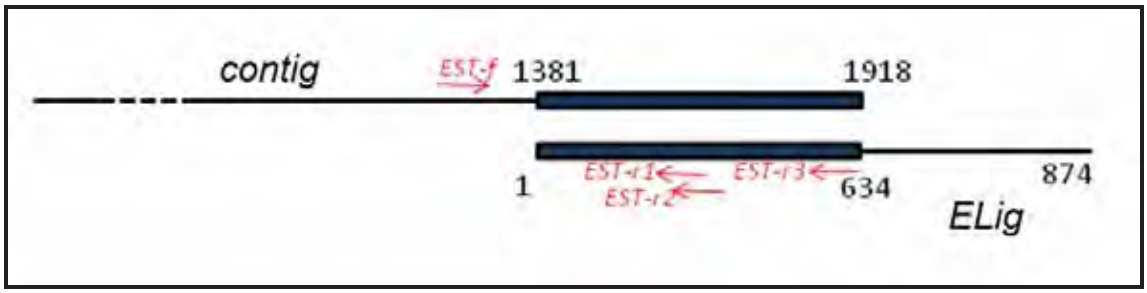

Figura 2: *Primers* para amplificar a região 5’ da sequência ELig. As barras azuis mostram a região em que as sequências apresentam identidade entre si. As setas vermelhas indicam as posições dos *primers* desenhados para amplificar essa região. O *primer EST-f* foi utilizado em reação de PCR com os *primers* reversos *EST-r1*, *EST-r2* e *EST-r3*. Os números acima e abaixo das barras indicam as posições dos nucleotídeos das sequências.

Esses *primers* foram utilizados para realização de PCR com o DNA de indivíduos *E. grandis*, *E. urophylla* e *E. saligna*. Os fragmentos resultantes foram recortados do gel de agarose, purificados, clonados e sequenciados conforme os protocolos abaixo.

- Purificação do DNA de bandas de Gel de agarose (Kit Amersham®)

As bandas de interesse foram recortada do gel de agarose e para cada 10 mg de gel, foi adicionado 10 µl de tampão de captura. Os tubos foram agitados em vortex e a mistura foi incubada a 60 ºC por 5-10 min até dissolver completamente a agarose. Em seguida, as amostras foram transferidas para as colunas, incubadas por 1min à temperatura ambiente e centrifugada a 13.000 rpm por 30 s. As colunas foram lavadas com tampão de lavagem e centrifugadas. O DNA purificado foi eluído em TE (pH=8,0).

- Clonagem

Os DNAs de interesse foram clonados em bactéria *Escherichia coli* da cepa DH5α com o vetor *pGEM®-T Easy vector sistem* (Promega®).

Para ligação do fragmento de DNA ao vetor foi utilizada a proporção 3:1 (três partes de DNA para uma parte de vetor). Foi adicionado 8 µl do produto da amplificação purificado, 3 U de T4 DNA ligase; 50 mg de vetor , tampão da enzima ligase e água deionizada para completar 10 µl de reação. Essa reação foi incubada por 20 h a 16 ºC.

Para a transformação da *E. coli*, foram misturados 10 µl de produto de ligação a 50 µl de células competentes. A mistura foi incubada em gelo por 30 min. Foi submetido um choque térmico a 42 ºC por 1 min e 30 s, as amostras foram transferidas para gelo por 5 min Foram adicionados 600 µl de meio de cultura líquido *Circle Grow*® (CG) e incubados por agitação suave por 50 min a 37 ºC. As amostras foram espalhadas em placa de petri com 25 ml de meio sólido CG. Para a seleção das colônias com o vetor foi utilizado o antibiótico ampicilina (100 µg/ml), e para a seleção das bactérias

com o vetor com o DNA alvo foi utilizado IPTG (Isopropil beta D Thiogalactopyranoside) (0,4 mM) em conjunto com X-galactose (XGAL) (0,1 μg/ml).

As placas foram incubadas a 37 ºC de um dia para o outro. Colônias brancas foram selecionadas e transferidas para meio líquido e mantidas em estufa a 37 ºC de um dia para outro, e posteriormente foi realizada a extração do DNA plasmidial através do protocolo que segue.

As células selecionadas foram transferidas para um novo tubo e centrifugadas. Em seguida, foram eluídas em solução contendo Tris-HCl (50 mM, pH=8,0), EDTA (10 mM, pH=8,0), RNAse (100 μg/ml) e homogeneizadas. A lise alcalina foi realizada com uma solução de NaOH (200 mM) e SDS (1%) por 5 min A reação foi neutralizada com uma solução de acetato de potássio (3 M, pH=5,5), as amostras foram incubadas em gelo por 20 min e então foram centrifugadas por 30 min a 14.000 rpm. O sobrenadante foi transferido para um novo tubo e foi adicionado isopropanol, nesta etapa, as amostras foram incubadas a -20 ºC por pelo menos uma hora. As amostras foram centrifugadas novamente e o precipitado de DNA foi lavado com etanol 70% e, após secagem, foi eluído em água deionizada autoclavada. O DNA extraído foi quantificado em gel de agarose a 1%.

Para confirmar a presença do inserto no plasmídeo, foi realizada um reação de digestão com a enzima *EcoRI* que possui sítio de restrição próximo ao local de inserção do DNA. A reação constou de: 400-500 ng de DNA, 2 unidades da enzima *EcoRI* (Invitrogen®), tampão da enzima, o volume foi completado para 30 μl com água ultra pura autoclavada.

A reação foi incubada a 37 ºC por 2 horas. A visualização do produto da digestão foi realizada em gel de agarose a 1% corado com brometo de etídio.

Culturas permanentes dos clones sequenciados foram feitas transferindo-se para um novo tubo 850 µl da cultura de bactéria e 150 µl de glicerol. As culturas foram armazenadas a -80 ºC.

- Sequenciamento

Para a reação de sequenciamento foi utilizado 100 ng de DNA plasmidial, tampão de sequenciamento (200 mM Tris–HCl; 5 mM MgCl2 pH=9); *Big Dye; primer* M13 (1 µM); a reação foi completada com água deionizada para um volume final de 10 µl.

As condições de amplificação foram as seguintes: extensão inicial de 96 ºC por 2 min; 40 ciclos de 96 ºC por 30 s, 55 ºC por 30 s, 60 ºC por 4min

Antes do sequenciamento, a reação foi purificada com 80 µl de isopropanol 75%, por 15 min, a placa foi centrifugada por 45 min a 4.000 rpm, o álcool foi descartado e foi adicionado etanol 70% para lavagem dos precipitados, foi realizada uma nova centrifugação nas condições anteriores por 10 min. O álcool foi descartado e os precipitados foram secos.

A reação foi eluída em 3 µl de tampão de carregamento desnaturante e levada para sequenciamento em um sequenciador 3100 *Applied Biotechnology.*

As sequências obtidas foram alinhadas e analisadas no software ChromasPro.

3.2.4. Determinação do número de cópias por *Southern blot*

A técnica de *Southern blot* foi utilizada para identificar o número de cópias da região de interesse. Para a realização da técnica, utilizou-se o *Kit Amersham® AlkPhos Direct™ Labeling and Detection Systems*. A descrição das etapas da técnica está a seguir.

- Extração de DNA

Para realização da técnica de *Southern blot* é necessário um DNA livre de impurezas, que permita uma digestão eficiente com enzimas de restrição. Para se obter uma maior qualidade na extração do DNA, foi utilizado o protocolo descrito por Weising *et al.*, (2005), com algumas modificações.

Foi extraído DNA de folhas frescas de duas árvores uma da espécie *E. grandis* e outra de *E. urophylla*.

O tampão de extração (2% CTAB, 1,4 M NaCl, 20 mM EDTA pH=8,0, 100 mM Tris-HCl (pH 8,0) e 1% β mercaptoetanol), foi aquecido a temperatura de 60 ºC e 12 ml desse tampão foi separado em tubos com capacidade de 50 ml onde foi adicionado 1 g de material foliar macerado em cadinho na presença de nitrogênio líquido. Essa mistura foi mantida sob agitação a 60 ºC por 60 min, 12 ml de CIA foram adicionados. As amostras foram deixadas em temperatura ambiente por 20 min sob agitação leve. As amostras foram centrifugadas a 4.000 rpm por 20 min e a fase superior (aquosa) foi transferida para um novo tubo onde adicionou-se 2 ml de CTAB 5% (5% CTAB, 2% PVP, 100 mM Tris-HCl pH=8,0, 25 mM EDTA, 2 M NaCl) e repetiu-se a lavagem com CIA. A fase superior foi transferida para um novo tubo onde

foi realizada digestão de RNAs com a adição de 6 µl de RNAse a 20 mg/ml e incubação por 30 min a 37 ºC. Para a precipitação do DNA, foram adicionados 0,8 volumes de isopropanol e centrifugado por 20 min a 5.000 rpm. O sobrenadante foi descartado e 10 ml de etanol 70% foram utilizados para lavagem do precipitado de DNA com centrifugação de 20 min a 5.000 rpm. O etanol foi descartado, o DNA foi seco à temperatura ambiente e depois foi eluído em 500 µl de TE (pH=8,0).

O DNA extraído foi quantificado em espectrofotômetro NanoDrop (ND1000®), e a qualidade do DNA foi verificada em eletroforese em gel de agarose a 1%, corado com brometo de etídio. Comprovada a qualidade do DNA este foi digerido com as enzimas de restrição selecionadas.

- Restrição enzimática

Para realização da técnica de *Southern blot* é necessário utilizar enzimas de restrição que não cortem a sequência de DNA que será usada como sonda. O programa pDraw 32 1.1.100 foi utilizado para analisar a sequência *ELig* e selecionar as enzimas. Foi verificada a presença de sítio de restrição para as seguintes enzimas: *BamHI*, *EcoRI*, *HindIII*, *NdeI*, *PstI*, *KpnI* e *XbaI*.

A técnica exige uma digestão completa do DNA, o eucalipto possui muitos compostos secundários que atrapalham a enzima de restrição, desta forma, a reação de digestão precisou ser padronizada para obtenção de uma digestão ideal. Para isso, foram realizados testes de digestões com as enzimas selecionadas em diferentes condições de concentração de DNA e enzima, em banho-maria a 37 ºC por 16 horas. O protocolo

com o melhor resultado de digestão de DNA e que foi utilizado para a técnica está descrito abaixo.

O DNA foi digerido em duplicata com três diferentes endonucleases: *EcoRI*, *NdeI* e *XbaI*. Foi utilizado para a reação de digestão: 20 µg de DNA; 60 unidades de enzima (Biolabs – New England®); tampão específico para a enzima selecionada; 0,1 mg de soro de albumina bovina (Invitrogen®); o volume foi completado para 1 ml com água ultra pura autoclavada. Após incubação a 37 ºC por 16 horas as enzimas foram inativadas a 65 ºC por 20 min. Para verificar se o processo de digestão das amostras de DNA ocorreu adequadamente, 1 µg das amostras de DNA digerido foi submetido à eletroforese em gel de agarose a 0,7%, corado com brometo de etídio.

Ao DNA digerido foram adicionados 0,1 volumes de acetato de amônio 7,8 M gelado e 2,5 volumes de etanol absoluto gelado. As amostras foram deixadas a -20 ºC até o dia seguinte e depois centrifugadas a 10.000 rpm por 60 min a 4 ºC. Em seguida, descartou-se o sobrenadante, o precipitado foi lavado em 1 ml de etanol 70% gelado e centrifugado a 10.000 rpm por 30 min a 4 ºC. O etanol foi descartado, o precipitado foi seco em centrífuga a vácuo e posteriormente eluído em 40 µl de TE (pH=8,0). As amostras foram armazenadas a -20 ºC até a sua utilização.

Adicionou-se 8 µl de Sacarose (40%) a 40 µl de cada amostra digerida. Utilizou-se 10 ng do DNA purificado da banda polimórfica como controle positivo.

As amostras foram aplicadas em gel de agarose a 1% corado com brometo de etídio e submetidas à eletroforese por 15 horas a 35 Volts para separação ideal dos fragmentos digeridos.

A visualização do gel foi realizada em um transiluminador de luz UV (*Molecular Imager ChemiDoc XRS Bio-Rad Laboratories, Inc*.) e foi fotografado com régua ao lado do marcador de peso molecular com zero na altura dos poços.

- Southern blot

Após visualizar e fotografar o gel, este foi cortado retirando-se as partes não utilizadas e a parte do marcador de peso molecular, deixando apenas as amostras e o controle positivo. O gel foi medido e fez-se um pequeno corte no canto esquerdo inferior para permitir sua posterior orientação. Em seguida, realizou-se a depurinação do gel através da incubação em HCl 0,25 M sob agitação suave por 10 min; lavando-o a seguir com água ultra pura. Seguiu-se a desnaturação do gel por 30 min sob agitação suave em NaCl 1,5 M , NaOH 0,5 M; lavando-o após com água ultra pura. Finalmente, o gel foi neutralizado em solução de Tris HCl 0,5 M pH=7,5 , NaCl 1,5 M por 30 min sob agitação suave.

O DNA do gel foi transferido para uma membrana (*Amersham Hybond*™- *N+*) através de capilaridade. Os papéis utilizados na transferência do DNA foram recortados de acordo com as dimensões do gel (14 x 16 cm).

Em uma vasilha colocou-se 500 ml da solução de transferência 20X SSC (NaCL 3 M; $Na_3C_6H_5O_7$ x $2H_2O$ 0,3 M, pH=7,0) e montou-se sobre uma plataforma uma ponte com o papel de cromatografia 3MM (*Whatman® 3MM CHR Chromatography Paper*). A ponte de papel foi molhada e as pontas do papel foram deixadas em contato com a solução 20X SSC. O gel foi colocado cuidadosamente sobre a ponte e sobre ele a membrana (*Amersham Hybond*™- *N+)* cortada nas dimensões do

gel e corretamente identificada. Sobre a membrana foram colocadas três folhas, já recortadas, do papel absorvente (*Amersham* Hybond™ Blotting Paper) e finalmente 6 cm de papel absorvente comum. O processo de transferência, que promove por capilaridade a passagem do DNA contido no gel para a membrana, ocorreu em um período de 20 h.

Terminada a transferência, a membrana protegida por um envelope de papel absorvente (*Amersham Hybond™ Blotting Paper*) e papel alumínio foi mantida a 80ºC por 2 h para a fixação do DNA à membrana e secagem da mesma.

- Obtenção da sonda

Para obtenção da sonda, a sequência *ELig* foi amplificada utilizando-se o par de *primer ELig* senso e reverso em uma reação contendo DNA de *E. grandis.*

A reação constou de 200 ng de DNA; 10 mM Tris-HCl, pH=8,0; 1,5 mM de $MgCl_2$; 0,5 mM de dNTPs; 0,25 µM de cada *primer* senso e reverso e 2 unidades de Taq DNA polimerase (Invitrogen®), o volume foi completado para 20µl com água ultra pura autoclavada.

A amplificação foi realizada em um aparelho termociclador utilizando as mesmas condições apresentadas anteriormente. As amostras amplificadas foram separadas por eletroforese em gel de agarose a 1% corado com brometo de etídio.

O produto amplificado de 874 nt (*ELig*) foi recortado do gel e purificado. O DNA purificado foi eluído em 50 µl de água ultra pura autoclavada e quantificado em

espectrofotômetro NanoDrop (ND1000®). A amostra que apresentou a melhor qualidade e maior quantidade de DNA foi selecionada para ser utilizada como sonda.

A sonda selecionada foi diluída na concentração de 10 ng/µl com água estéril. Em seguida, adicionou-se a 10 µl da sonda já diluída: 10 µl de *Reaction buffer*, 2 µl de *Labelling Reagent* e 10 µl de *Cross- linker* diluído (1 *cross linker* : 4 água estéril), todos reagentes do Kit *Amersham AlkPhos Direct™ Labeling and Detection Systems.*

- Hibridização da membrana

A membrana foi colocada na garrafa de hibridização contendo o tampão (NaCl 5 M; *Blocking reagent* 4%; *Hybridization buffer* q.s.p 40 ml) previamente aquecido a 55 ºC, e foi deixada por 15 min a mesma temperatura. Acrescentou-se 200 ng da sonda marcada ao tampão (5 ng/ml de tampão) e foi hibridizada a 55 ºC por 21 h sob agitação.

A solução anterior foi removida e foram adicionados 300 ml do 1º tampão de lavagem (uréia 2 M, SDS 0,1%; fosfato de sódio 50 mM, pH 7.0; NaCl 0,15 M; $MgCl_2$ 1 mM; *Blocking reagent* 0,2%) já aquecido a 55 ºC à garrafa de hibridização contendo a membrana, deixando sob agitação por 10 min na mesma temperatura. A lavagem foi repetida com mais 300 ml do 1º tampão.

Em seguida, a membrana foi colocada em um recipiente limpo e incubada em 100 ml do 2º Tampão de Lavagem (Tris base 0,1 M; NaCl 0,05 M; $MgCl_2$ 2 mM) sob agitação por 5 min a temperatura ambiente. Uma segunda lavagem com 100 ml do 2º tampão foi realizada.

Terminada a lavagem, o excesso de líquido foi retirado da membrana e esta foi colocada em um saco plástico aberto. Sobre a superfície da membrana foram espalhados 5 ml da solução de detecção (*Kit Amersham AlkPhos Direct™ Labeling and Detection Systems*), em seguida fechou-se o plástico deixando a solução agir por 5 min. O excesso de líquido foi retirado, o plástico foi selado e a membrana colocada dentro do cassete de raios X. Uma chapa de raios X corretamente identificada foi colocada em cima da membrana por 10 min. Após o tempo de exposição necessário, a chapa foi revelada por 4 min, em seguida fixada por 15 min e por fim lavada em água ultra pura.

3.2.5. Técnica de *RACE 3'* (*Rapid Amplification of cDNA Ends*)

A técnica de *RACE* 3' foi realizada com o objetivo de elucidar a região 3' transcrita da região de interesse.

O protocolo foi realizado com o uso do kit 3´ *RACE System for Rapid Amplification of cDNA Ends* (Invitrogen®). Um esquema simplificado da técnica está representado na Figura 3.

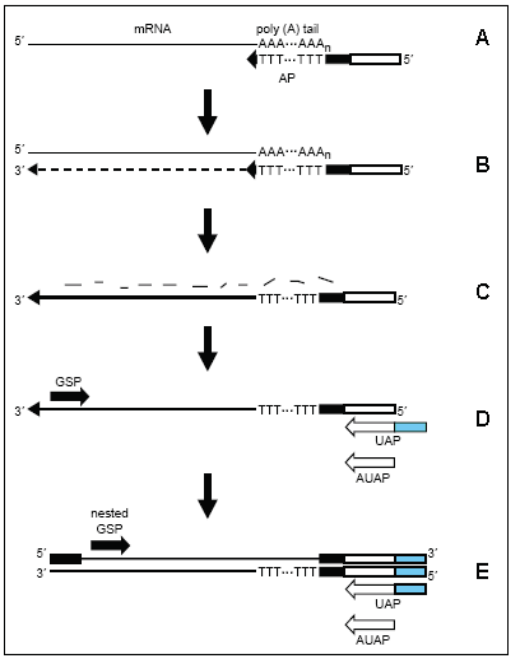

Figura 3: Esquema da metodologia *RACE* 3' (Kit 3´ *RACE System for Rapid Amplification of cDNA Ends* - Invitrogen®). **Em A:** pareamento do *primer* adaptador (AP) com cauda dT ao RNA mensageiro. **Em B:** extensão da fita complementar a partir de AP utilizando *SuperScript™ II RT*; **Em C:** degradação da fita molde de RNA utilizando RNase H; **Em D:** PCR para amplificação do cDNA utilizando os *primers* GSP 1 (*Primer* Gene específico) e o UAP ou AUAP; **Em E:** PCR para amplificação do produto primário, utilizando os *primers* UAP ou AUAP e GSP 2 (*nested*).

A partir da sequência *ELig* foram desenhados *primers* gene - específicos externos e internos, GSLig1 e GSLig2, respectivamente. Para o desenho dos *primers* utilizou-se a ferramenta OligoAnalyser 3.1 (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx). O *primer* GSLig1 foi composto de 23nt e o GSLig2 por 29nt com temperaturas de pareamento de 61 ºC.

- Extração de RNA

Foram extraídos RNAs de diferentes tecidos vegetais de plântula de *E. grandis* (raiz e caule) e *E. urophylla* (lignotuber) através do protocolo descrito por Chang, *et al.* (1993) com modificações.

A 200 mg de tecido macerado em nitrogênio líquido foi adicionado 1 ml de tampão de extração CTAB pré aquecido a 65 ºC (2% CTAB; 2% PVP; 100 mM Tris-

HCl pH=8,0; 25 mM EDTA; 2 M NaCl), o macerado foi mantido a 65 ºC por 15 min Após atingir a temperatura ambiente, foi adicionado 1 ml de CIA (clorofórmio – álcool isoamílico 24:1) e foi centrifugado a 10.000 rpm por 10 min. O sobrenadante foi coletado e uma nova lavagem com CIA foi realizada nas mesmas condições anteriores. Foi adicionado um volume de cloreto de lítio (LiCl 5 M), as amostras foram deixadas a 4 ºC de um dia para outro para precipitação. No dia seguinte foi centrifugado por 30 min a 12.000 rpm a 4 ºC. O sobrenadante foi descartado e o precipitado formado foi eluído em 200 µl e TE-SDS (10 mM Tris-HCl pH=8,0; 1 mM EDTA pH=8,0; 1% SDS (dodecil sulfato de sódio)) e adicionado 100 µl de NaCl 5 M e 300 µl de isopropanol. As amostras foram armazenadas por 30 min a -20 ºC para precipitação, após esse período foi centrifugado a 12.000 rpm por 20 min a 4 ºC. O sobrenadante foi descartado e o precipitado formado foi lavado com 1 ml de etanol 70%, as amostras foram centrifugadas a 10.000 rpm (4 ºC) por 5 min. O precipitado foi eluído em 30 µl de água DEPC autoclavada e o RNA foi armazenado em freezer a -70 ºC.

- RACE 3'

Para transcrever o RNA em cDNA é necessário primeiramente eliminar todo DNA da amostra através da digestão com DNase, para isso 1 µg de RNA foi tratado com 1 unidade da enzima DNase (DNase I, RNase *free*) e foi incubado a 37 ºC. A enzima DNase foi inativada à temperatura de 65 ºC por 10 min na presença de EDTA (2,5 mM).

Em tubos estéreis adicionou-se 1 µg do RNA tratado, *primer* adaptador AP (0,83 µM) e volume final de 12 µl completados com água deionizada autoclavada. Essa

mistura foi mantida a 70 ºC por 10 min e em seguida levada ao gelo por 1 min. Foi adicionado então uma mistura contendo Tris-HCl (10 mM, pH=8,0); $MgCl_2$ (2,5 mM); dNTP (0,5 mM); DTT (0,1 M) e *SuperScript*™ *II RT* (200 unidades), essa mistura (volume final de 20 µl) foi incubada a 42 ºC por 55 min e inativada a 70 ºC por 15 min Foi levada ao gelo para adição de RNase H (2 unidades) e incubada por 20 min a 37 ºC para a digestão da fita molde de RNA.

O cDNA obtido foi submetido a reações de amplificação utilizando os *primers* gene específicos. Na primeira reação foi utilizado o *primer* GSP externo (*GSLig1*) e na segunda o *primer* GSP interno (*GSLig2*).

As reações foram realizadas em um volume final de 50 µl e continham Tris-HCl (10 mM, pH=8,0); $MgCL_2$ (1,5 mM); dNTP (0,2 mM); *primer GSLig* (0,2 µM); *primer* AUAP (0,2 µM); enzima Taq DNA polimerase do tipo *hot start* (Invitrogen® - 1 unidade) e 2 µl de cDNA.

O ciclo de amplificação foi o seguinte: extensão inicial de 94 ºC; 35 ciclos de 94 ºC por 3 min, 60 ºC por 45 seg, 72 ºC por 1 min; e uma extensão final de 72 ºC por 7 min.

Para a reação *nested*, o DNA amplificado na primeira reação foi diluído em água deionizada na proporção 1:50. Foi utilizado 1 µl do DNA amplificado diluído, a reação ocorreu com o *primer* interno *GSLig2*, nas mesmas condições anteriores.

As bandas selecionadas foram cortadas do gel de agarose e purificadas. O DNA foi eluído com 50 µl de tampão TE. Os fragmentos obtidos foram clonados sequenciados e as sequências obtidas foram analisadas com o auxilio do software ChromasPro e ferramentas de BLAST (Stephen *et al.*, 1997).

3.2.6. *Genome Walking*

Para a realização da técnica de *Genome Walking*, foi utilizado o Kit *GenomeWalker*™ Universal Kit (Clonthec). Um esquema resumido da técnica é apresentado na Figura 4:

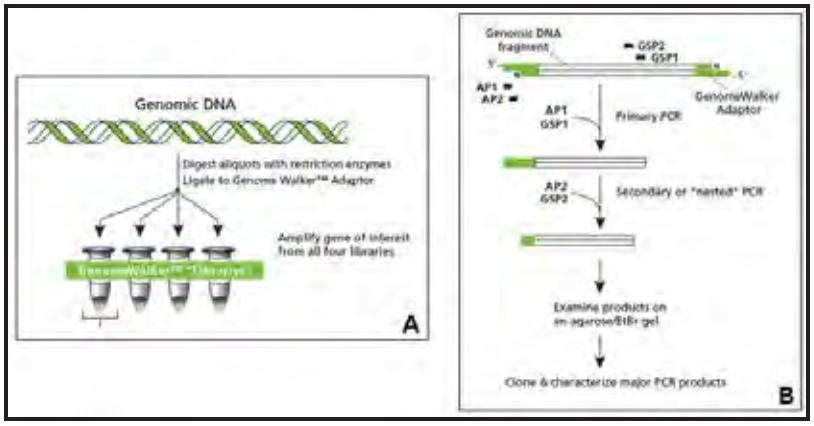

Figura 4: Esquema da metodologia *GenomeWalker*™ *Universal Kit* (Clontech Laboratories, Inc.). **Em A:** digestão do DNA genômico por enzimas de restrição e ligação dos adaptadores. **Em B:** reação de PCR para amplificação da região alvo no sentido 5', através do uso dos *primers* GSP 1 e AP 1 (PCR primário) e GSP 2 e AP 2 (PCR *nested*).

O material vegetal selecionado para o desenvolvimento da técnica foi DNA extraído de folhas um indivíduo de *E. grandis* pertencente ao programa de melhoramento da empresa Suzano Papel e Celulose.

Os *primers* foram desenhados na sequência *ELig* para estender as sequências no sentido 5'e 3'. De acordo com as normas do fabricante: temperatura variando de 62-65 ºC e os tamanhos variaram de 23 a 29 nucleotídeos.

Para construir as bibliotecas de DNA digerido foram utilizadas as enzimas de restrição *EcoRV*, *DraI*, *PruII* e *SspI*.

Para a digestão, foi utilizado 2.500 ng de DNA, 80 unidades de enzima de restrição e tampão próprio de cada enzima num volume total de 100 µl, essa mistura foi incubada a 37 ºC por 2 h, nessa etapa a reação foi misturada em vortex e deixada em incubação a 37 ºC por 16-18 h.

Para purificação do DNA digerido foi adicionado 100 µl de fenol, e a mistura foi centrifugada por 1 min a 13.000 rpm a fase superior foi transferida para outro tubo onde foi adicionado 95 µl de clorofórmio, foi centrifugado nas mesmas condições anteriores e a fase superior foi transferida para outro tubo. Para precipitação do DNA foi adicionado 190 µl de etanol 95%; 9,5 µl de Acetato de sódio (3 M, pH=4,5) e 20 µg de glicogênio. A mistura foi levada ao vortex e posteriormente foi centrifugada por 15 min a 4 ºC e 14.000 rpm. O sobrenadante foi descartado e o pellet foi lavado com 100 µl de etanol 80% gelado, foi centrifugado por 10 min a 14.000 rpm o álcool foi descartado e após a secagem do precipitado, este foi ressuspendido em 20 µl de TE (pH=7,5). O DNA purificado foi quantificado em espectrofotômetro NanoDrop (ND-1000®).

Para a ligação dos adaptadores ao DNA digerido e purificado foi utilizado 100 ng desse DNA, 6 µM de adaptador 3 unidades de enzima T4 DNA ligase e o tampão recomendados para a enzima, num volume final de 8 µl. A mistura foi incubada a 16 ºC por 15 horas e para inativação da enzima, esta mistura foi incubada a 70 ºC por 5 min. Foi adicionado 72 µl de TE (pH=7,5) ao produto de ligação.

Foi realizada reação de PCR com as quatro bibliotecas confeccionadas e nos dois sentidos do DNA (5’ e 3’) com cada um dos *primers* separadamente. Foram realizadas duas reações para cada sentido (5’e 3’) com *primer* interno e externo.

A reação de PCR foi composta de 1 µl da biblioteca de DNA digerido; 2 mM de $MgCl_2$; 0,4 mM de dNTP; 0,2 mM de *primer* AP1 (do kit); 0,2 mM de *primer*

específico; 1,5 unidades de Taq Platinum (Invitrogen®); 10 mM Tris-HCl (pH=8,0) e água deionizada para completar o volume de 50 µl.

A reação foi realizada em termociclador nas seguintes condições: sete ciclos de 25 s a 94 ºC e 3 min a 72 ºC; 32 ciclos de 25 s a 94 ºC e 3 min a 67 ºC e posteriormente uma extensão final a 67 ºC por 10 min.

Após a reação, foi separado 1 µl do produto da reação de PCR para realizar a reação *nested* com o *primer* AP2 (do kit) e o *primer* específico 2. Os demais reagentes, assim como as condições de amplificação foram utilizados nas mesmas condições anteriores.

Para visualização do produto da reação, foi realizado um gel de agarose 1% corado com brometo de etídio. As bandas selecionadas foram recortadas e purificadas. O DNA obtido foi quantificado em espectrofotômetro NanoDrop (ND-1000®).

Os fragmentos obtidos foram clonados e sequenciados e as sequências obtidas foram analisadas com o auxilio do software Chromas Pro e ferramentas de BLAST (Stephen *et al.*, 1997).

3.2.7. Análises de Bioinformática

3.2.7.1. Comparação da Sequência *ELig* com banco de dados de EST

A sequência *ELig* foi comparada no banco de dados de EST de eucalipto FOREST-FAPESP (*Eucalyptus Genome Sequencing Project Consortium* - Vicentini *et al*., 2005), através da ferramenta de BLASTn (*Basic Local Alignment Search Tool*) (Stephen *et al.*, 1997). As sequências de nucleotídeos com identidade ao *ELig* foram

alinhadas com o *ELig* para comparação, para o alinhamento foi utilizada a ferramenta BLASTn 2seq, o software ClustalX e ChromasPro.

A sequência de nucleotídeos do *contig*, com identidade ao *ELig*, foi traduzida em possíveis *ORFs* (*Open Reading Frames*) através da ferramenta *GENE FINDING in Eukariota* (*FGENESH HMM based Gene structure prediction*, utilizando a opção de Dicot plants *Arabdopsis* disponível em http://linux1.softberry.com/all.htm). As proteínas preditas foram utilizadas para comparação com banco de dados públicos de domínios de proteínas (pFam – Finn *et al.*, 2010).

3.2.7.2. Comparação da Sequência *ELig* com o genoma completo de eucalipto

A sequência *ELig* e o *contig* do EST foram utilizadas como sonda para análise no genoma completo do eucalipto do banco de dados *phytozome* (disponível em http://www.phytozome.net/eucalyptus.php) através da ferramenta BLAST (Stephen *et al.*, 1997). Esse banco de dados apresenta o genoma 8x mapeado do *Eucalyptus grandis BRASUZ1*. Possui cerca de 690 Mbp agrupados em 4952 *scaffolds*. As sequências de nucleotídeos com identidade ao *ELig* e ao *contig* EST foram alinhadas para comparação, para o alinhamento foi utilizada a ferramenta BLASTn 2seq (Stephen *et al.*, 1997), o software ClustalX e ChromasPro.

A sequência de nucleotídeos do *scaffold*, com identidade ao *ELig*, foi traduzida em possíveis *ORFs* (*Open Reading Frames*) através da ferramenta *GENE FINDING in Eukariota* (*FGENESH HMM based Gene structure prediction*, utilizando a opção de Dicot plants *Arabdopsis* disponível em http://linux1.softberry.com/all.htm). As

proteínas preditas foram utilizadas para comparação com banco de dados públicos de domínios de proteínas (pFam - Finn *et al.*, 2010).

Para análise e identificação de genes candidatos foram selecionados 100 mil nucleotídeos *upstream* e *downstream* da sequência com identidade ao *ELig* no banco de dados *phytozome*. Essa sequência de 200 mil nucleotídeos foi utilizada para a predição de proteínas como descrito anteriormente.

O banco de dados do genoma de eucalipto *phytozome* não está finalizado, apresentando-se em fase de organização pós sequenciamento, as sequências ainda não estão completamente ordenadas em cromossomos, mas em *scaffolds*.

4. RESULTADOS E DISCUSSÃO

4.1 Análise da eficiência do uso do Marcador *ELig* em eucalipto

A sequência amplificada pelo par de *primers* SCAR, desenvolvido por Martins (2006), foi denomidade *ELig* (*Eucalyptus* – Lignotuber).

Com o objetivo de avaliar a eficiência do marcador *ELig* na identificação de indivíduos portadores de lignotuber em eucalipto, o marcador foi testado em 164 indivíduos de quatro espécies de eucalipto. Os resultados são apresentados na tabela 1.

Tabela 1: Teste da eficiência do marcador *ELig* com indivíduos de quatro espécies de eucalipto *E. grandis*, *E. urophylla*, *E. saligna* e *E. brassiana*.

Espécie	Lignotuber	Genotipados	Com *ELig*	Sem *ELig*	Eficiência
E. urophylla	+	77	1	76	98,7%
E. grandis	-	69	61	8	88,4%
E. saligna	+	10	1	9	90%
E. brassiana	+	8	0	8	100%
	Total	164			93,9%

Legenda: o sinal + indica as espécies que apresentam lignotuber; o sinal - indica a espécie onde não há a formação do órgão; Com *ELig*: número de indivíduos com a marca; Sem *ELig*: número de indivíduos que não apresentaram a marca; Total: número total de indivíduos genotipados.

O *ELig* foi eficiente em identificar plantas portadoras de lignotuber em 93,9% do total de plantas analisadas, onde indivíduos sem lignotuber apresentaram a marca

(61) e indivíduos com lignotuber não apresentaram a marca (93) (Figura 5). Em 6,1% dos indivíduos a marca demonstrou não ser eficiente em identificar esses indivíduos.

A espécie *E. grandis* não apresenta lignotuber, dos 69 indivíduos analisados, 61 apresentaram a marca (88,4%). As demais espécies analisadas apresentam lignotuber e o esperado é não apresentarem o marcador; em *E. urophylla* a eficiência do marcador foi de 98,7%, uma vez que de 77 indivíduos genotipados, somente um apresentou a marca; em *E. saligna* a eficiência do marcador foi de 90% com dez indivíduos genotipados e *E. brassiana* a eficiência foi de 100%, onde nenhum dos 8 indivíduos analisados apresentou a marca.

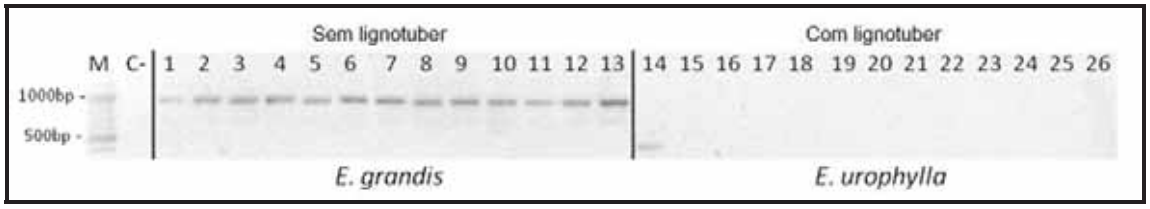

Figura 5: Amplificação com o *primer ELig* em treze indivíduos *E. grandis* e treze *E. urophylla*. As canaletas de 1 a 13 contem amplificação com DNA de indivíduos da espécie *E. grandis* e de 14 a 26 amplificação com DNA de indivíduos *E. urophylla*. Podemos observar a eficiência do marcador em identificar os indivíduos da espécie *E. grandis*, que não apresentam lignotuber. Legenda: M: marcador molecular com tamanhos de fragmentos conhecidos *ladder* 1Kbp (Fermentas®); C-: controle negativo.

Um dos problemas em se trabalhar com espécies "puras" de eucalipto no Brasil é a facilidade de ocorrência de cruzamentos interespecíficos de forma natural nos bancos de germoplasma dessas espécies. É comum que espécies isoladas geograficamente na Austrália sejam plantadas próximas no território brasileiro, possibilitando a formação de híbridos, por exemplo, a espécie *E. urophylla* foi plantada em Rio Claro, na sua introdução no país, ao lado das espécies *E. tereticornis* e *E. saligna*. É muito provável que populações formadas com sementes de *E. urophylla* sejam híbridas entre essas espécies, existindo ainda a possibilidade de outras espécies,

como o *E. grandis*, cruzarem com o *E. urophylla*, já que a espécie floresce 10 meses ao ano (Pryor, 1971). Diante disso não há como mensurar o nível de hibridação dessas espécies no germoplasma do Brasil e das empresas de papel e celulose. Uma alternativa para solucionar esse problema, seria a busca de indivíduos puros das espécies analisadas no centro de origem, ou seja, nas florestas naturais onde as espécies ainda se mantêm geneticamente puras, no entanto, a hibridação ocorre, em baixa porcentagem, nas zonas limítrofes entre duas espécies.

As informações da presença ou ausência de lignotuber em uma determinada espécie de eucalipto é outro fator importante a ser discutido, uma vez que essas informações são encontradas em livros textos e evidenciadas na rotina de programas de melhoramento, mas não existem trabalhos de genética de população considerando a presença ou ausência de lignotuber em todos os indivíduos pertencentes a uma espécie.

O marcador *ELig* foi testado com alguns indivíduos híbridos *E. urograndis*, porém como nesses indivíduos o lignotuber aparece apenas sob condições de estresse, não foi possível determinar qual o estresse necessário para formação do órgão, desta forma, os indivíduos sem lignotuber, podem não ter apresentado o órgão pelo fato do estresse não ter atingido o limiar necessário para aquele individuo, ou seja ele não apresentou o órgão por fatores ambientais e não por fatores genéticos. Portanto não temos como afirmar se um híbrido *E. urograndis* que ainda não apresentou o órgão se ele não apresentará em condições maiores de estresse e isso poderia ocasionar em falsas recombinações da marca *ELig* com a característica lignotuber. Já os híbridos *E. urograndis* que apresentam lignotuber, podem ser indivíduos heterozigotos para o gene de efeito principal na formação do órgão, cujo marcador *ELig* está etiquetando (Figura 6). Desta forma cerca de dois terços dos indivíduos híbridos com o lignotuber do cruzamento de um *E. grandis* com *E. urophylla* apresentariam a marca. Por isso o

marcador não foi avaliado nesses indivíduos, pois não seria eficiente na marcação da característica, assim, escolhemos as espécies puras para a análise do marcador.

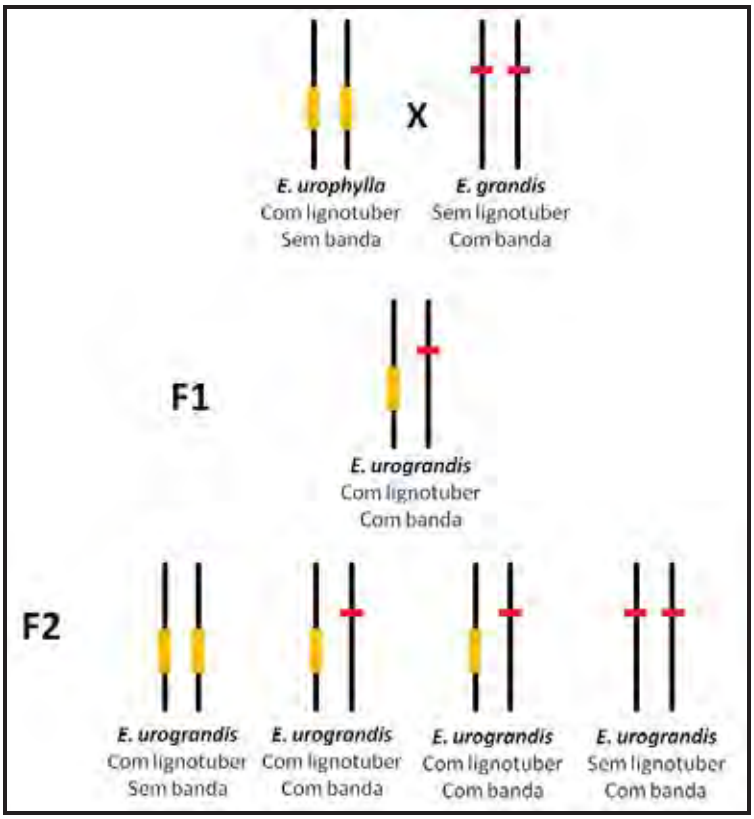

Figura 6: Esquema de ligação da marca *ELig* e o gene de efeito principal na formação do lignotuber em cruzamento de *E. urophylla* e *E. grandis* na produção de híbridos F2 de *E. urograndis*. Em amarelo é esquematizado o gene de efeito principal na formação do lignotuber e em vermelho a marca *ELig*. As barras verticais pretas indicam os cromossomos. Observe que em 2/3 dos híbridos *E. urograndis* com lignotuber é esperada a presença da marca.

A genotipagem com a marca *ELig* demonstrou uma alta correlação entre a ausência do fragmento amplificado e a presença do lignotuber, indicando que na região genômica de pareamento dos *primers* nos indivíduos com lignotuber exista uma mutação, ou deleção dessa região, que impede a amplificação do *ELig*. Diante do exposto, o marcador *ELig* é um marcador eficiente para identificar indivíduos da espécie *E. grandis* que não apresentam a característica lignotuber, estando ausente nas

espécies de eucalipto que apresentam a característica, isso faz do marcador uma ferramenta interessante para o uso em melhoramento genético e seleção assistida por marcador molecular (SAM).

4.2. Análise das Sequências no banco de dados FOREST de eucalipto

Através da ferramenta de BLASTn (Stephen *et al.*, 1997), no banco de dados de EST, foi possível determinar que a sequência *ELig* apresenta identidade com um *contig* (1918 bp) e um *singleton* (827 pb) de *Eucalyptus.*

O *singleton* apresentou identidade à sequência *ELig* em duas regiões, a primeira que inclui os nucleotídeos de 1 a 266 (barra rosa da Figura 7) e a segunda dos nucleotídeos 262 a 423 (barra azul da Figura 7). Na primeira região (1 a 266) a identidade ocorreu com a sequência *ELig* invertida (*plus/minus*) e foi da posição do nucleotídeo 396 a 662 da sequência *ELig* (E-value de $2e^{-136}$ e identidade de 99%). Na segunda região de identidade (262 a 423) a identidade ocorreu com as sequências no mesmo sentido (*plus/plus*) e foi da posição do nucleotídeo 248 a 409 da sequência *ELig* (E-value de $7e^{-79}$ e identidade de 99%). Este *singleton* é formado por um transcrito de plântula de *E. grandis* cultivada totalmente no escuro (SL7). A Figura 7 apresenta o alinhamento entre a sequência *ELig* e o *singleton*.

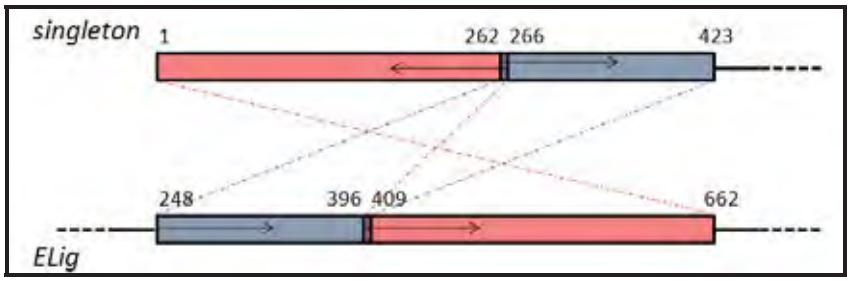

Figura 7: Alinhamento da sequência do *singleton* com a sequência *ELig*. As barras grossas com a mesma cor indicam as regiões com identidade entre as sequências. As setas indicam a direção de alinhamento. As linhas pontilhadas indicam onde começa e termina a identidade. Os números indicam as posições dos nucleotídeos das sequências

O alinhamento do *ELig* com o *contig* ocorreu entre a região 5' da sequência *ELig* e a região 3'do *contig*, conforme esquematizado na Figura 8.

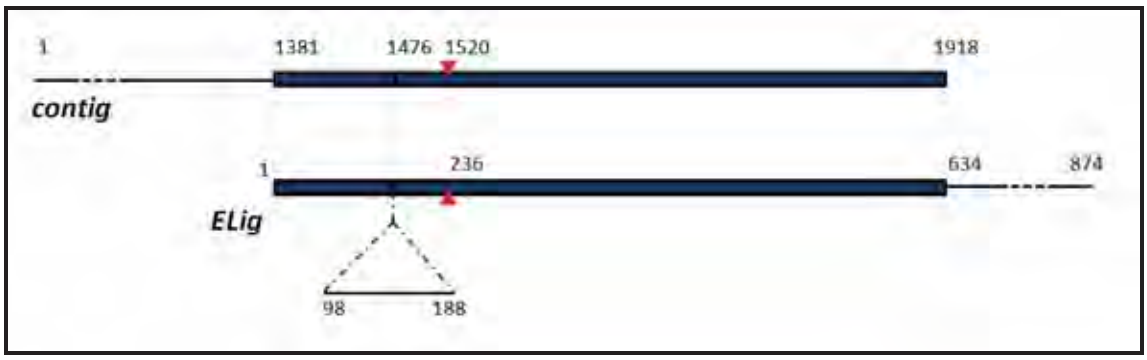

Figura 8: Alinhamento da sequência do *contig* com a sequência *ELig*. As barras azuis mostram a região em que as sequências apresentam identidade entre si e os traços pretos mostram as regiões sem identidade entre as sequências. O triângulo vermelho mostra o códon de parada de leitura (*STOP codon*) presente nas sequências. Os números acima e abaixo das barras indicam as posições dos nucleotídeos das sequências.

O *contig* apresentou 97% de identidade ao *ELig*, onde 440nt de 449nt foram idênticos (*e-value* 0.0). Esse *contig* é formado pelas seguintes bibliotecas: Plântulas de *E. camaldulensis* cultivadas no escuro (SL8); Madeira de *E. grandis* (WD2); Plântulas de *E. grandis* cultivadas no escuro e expostas a luz por 3 horas antes da extração de

RNA (SL1); Plântulas de *E. urophylla* cultivadas no escuro (SL6); Raízes de mudas de *E. grandis* de viveiro (RT3).

A região 5' da sequência *ELig* (nucleotídeo 1 ao 97) apresentou identidade de 100% à região do nucleotídeo 1381 a 1476 do *contig*, a região do nucleotídeo 98 ao 188 da sequência *ELig* não apresenta similaridade com o *contig*, indicando se tratar de uma região não transcrita da sequência, um *intron*, a região do nucleotídeo 189 a 634 da sequência *ELig* apresentou identidade de 99% à sequência do *contig* da posição do nucleotídeo 1477 até a última base 1918 na posição 3', as sequências não apresentaram nenhum polimorfismo *indel*, apresentando apenas 3 substituições entre as sequências (283 *ELig* C/T 1567 *contig*; 319 *ELig* C/T 1603 *contig*; 344 *ELig* A/G 1628 *contig*). Foi detectada a presença de um códon de parada de leitura (*STOP codon*) na posição do nucleotídeo 236 na sequência *ELig*, que corresponde à posição do nucleotídeo 1520 na sequência do *contig*, indicando se tratar realmente da região final de um gene.

A sequência do *contig*, com identidade ao *ELig*, foi traduzida em possíveis *ORFs* (*Open Reading Frames*) para comparação com banco de dados públicos de domínios de proteínas (Finn *et al.*, 2010), apresentando grande similaridade com o domínio das família de proteínas *DnaJ* (Bit-score de 28,7 e E-value de $6{,}2e^{-07}$) (Figura 9).

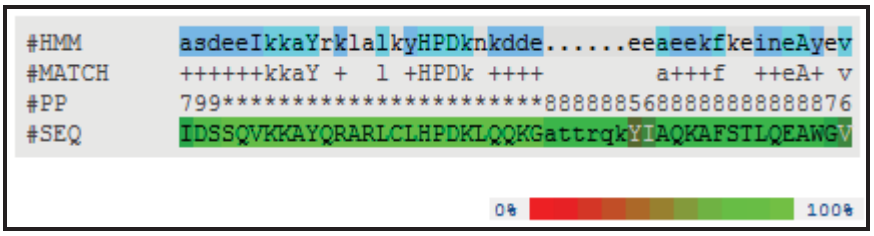

Figura 9: Identidade entre a predição de proteína do *contig* do EST com o domínio *DnaJ* do banco de dados de domínio de proteínas Pfam (Finn *et al.*, 2010). Quanto mais verde a barra, maior a similaridade entre as sequências. Legenda: #HMM – sequência consenso depositada no banco de dados; #MATCH – equivalência entre as sequências *query* e a sequência HMM; #PP – grau de confiança do alinhamento de cada resíduo individual; #SEQ – sequência utilizada como *query*.

Estruturalmente, a proteína *DnaJ* consiste de um domínio N-terminal conservado (chamado de *"J' domain"*) com cerca de 70 aminoácidos, uma região rica em glicina (*glycine-rich – "G' domain"*) de cerca de 30 resíduos, um domínio central contem quatro repetições do *motif* CXXCXGXG (*"CRR' domain"*) e uma região C-terminal com 120 a 170 resíduos (Finn *et al.*, 2010). A estrutura do domínio é mostrada na Figura 10:

Figura 10: Esquema da estrutura das proteínas com domínio *DnaJ* mostrando todos os seus *motifs* (adaptado de http://pfam.sanger.ac.uk/family/PF00226):

Proteínas da família *DnaJ* apresentam em geral funções de chaperonas e participam em vários processos da maquinaria intracelular (Cheetham e Caplan, 1998; Miernyk, 2001; Walsh *et al.*, 2004). O domínio *DnaJ* é altamente conservado e está presente em várias proteínas procarióticas e eucarióticas dentre os diversos grupos de seres vivos. É uma família grande e pode ser o maior domínio entre as chaperonas (Cheetham e Caplan, 1998).

As proteínas da família *DnaJ* podem se associar a diversos sistemas de proteínas *hsp* (*heat shock proteins*) (Walsh *et al.*, 2004). Em seres procariontes, a proteína interage com a chaperona *hsp70-like DnaK* (Cyr *et al*., 1994). O domínio "J" das proteínas *DnaJ* media a interação com proteínas *DnaK* e é constituído de quatro hélices, a segunda hélice possui uma superfície carregada, que inclui pelo menos um par de resíduos básicos que são essenciais para a interação com o domínio de ATPase da *hsp70* (Bork *et al*., 1992).

Há diversos mecanismos envolvidos para garantir a sobrevivência celular sob condições hostis. Os genes ativados por estresse respondem aos diferentes estresses ambientais através da modulação da expressão de genes de proteínas *hsp* (Mukhopadhyay *et al.*, 2003). "Hsps/chaperonas" são expressas nas plantas em casos de estresse causado por altas temperaturas, e em respostas a outros estresses ambientais, como os de salinidade, de frio, hídrico e oxidativo (Macário e Macário, 1999; Aoki *et al.*, 2002; Wang *et al*., 2004). Elas possuem um papel crucial em proteger as plantas contra o estresse e no restabelecimento da homeostase celular (Wang, *et al*., 2004). O sistema "Hsp70/chaperonas" é responsável pelo *folding* de proteínas, translocação e degradação em processos celulares normais, estabilizar proteínas e membranas, assistir o *refolding* de proteínas em condições de estresse podendo se ligar a componentes nucleares e proteger esses componentes contra a desnaturação ou agregação (Caplan *et al.*, 1993; Cheetham e Caplan, 1998; Macário e Macário, 1999; Aoki *et al.*, 2002; Wang, *et al*., 2004).

Algumas proteínas *hsp* contribuem para o aumento da resistência da planta contra estresse hídrico. Em tabaco, por exemplo, num experimento com plantas transgênicas, a super expressão de *hsp70* contribuiu para o aumento na tolerância ao estresse hídrico, uma vez que as plantas com o transgene senso *NtHSP70-1* sobreviviam

a condições de seca mais severas quando comparadas as plantas sem o transgene e com o transgene anti-senso (Cho e Hong, 2006).

A exposição da planta a condições de estresse como altas temperaturas e estresse hídrico levam a modificações fisiológicas da planta, como o acúmulo de proteínas desnaturadas (Lee e Schofft, 1996; Cho e Hong, 2006). Em melão, foi observado que em estágios mais tardios de estresse hídrico, proteínas chaperonas têm a sua expressão aumentada (Yoshimura *et al*., 2008). Esse aumento na expressão tem uma função importante na proteção da estrutura das proteínas (Hammond e Helenius, 1995; Cho e Hong, 2006), protegendo as proteínas das células radiculares da desnaturação (Yoshimura *et al*., 2008).

Genes induzidos pela seca estão relacionados a mecanismos que conferem a estabilização estrutural celular com a proteção da integridade da membrana celular e a preservação da estrutura das proteínas, protegendo, desta forma, as macromoléculas e membranas celulares contra o estresse causado pela seca (Cho e Hong, 2006). Os estresses abióticos, como seca podem levar a ativação de vias metabólicas de resposta a estresse o que resulta na indução de diversos genes de defesa ao organismo da planta para a produção de proteínas *hsp*. Plantas com níveis elevados de proteínas da família de hsp exibem uma maior tolerância ao déficit de água (Alvim *et al*., 2004; Cho e Hong, 2006).

Martins (2006) demonstrou que há uma forte relação do estresse hídrico e a formação de lignotuber em eucalipto, uma vez que ao aumentar os níveis de estresse e diminuir a quantidade de água fornecida às plantas, o desenvolvimento de lignotuber nas plantas também foi aumentado, atingindo a proporção de três plantas com lignotuber para uma sem o órgão como mostra o gráfico na Figura 11. Segundo a

literatura, o estresse hídrico também tem importante função no aumento na produção de proteínas *hsp* aumentando a tolerância dessas plantas ao estresse (Cho e Hong, 2006). Existe, portanto, forte indício de que a região *ELig* tenha relação com o domínio de proteínas *DnaJ*, especialmente proteínas *hsp*, pois tanto essas proteínas quanto a formação do lignotuber, apresentam referências de estresse abiótico, principalmente estresse hídrico.

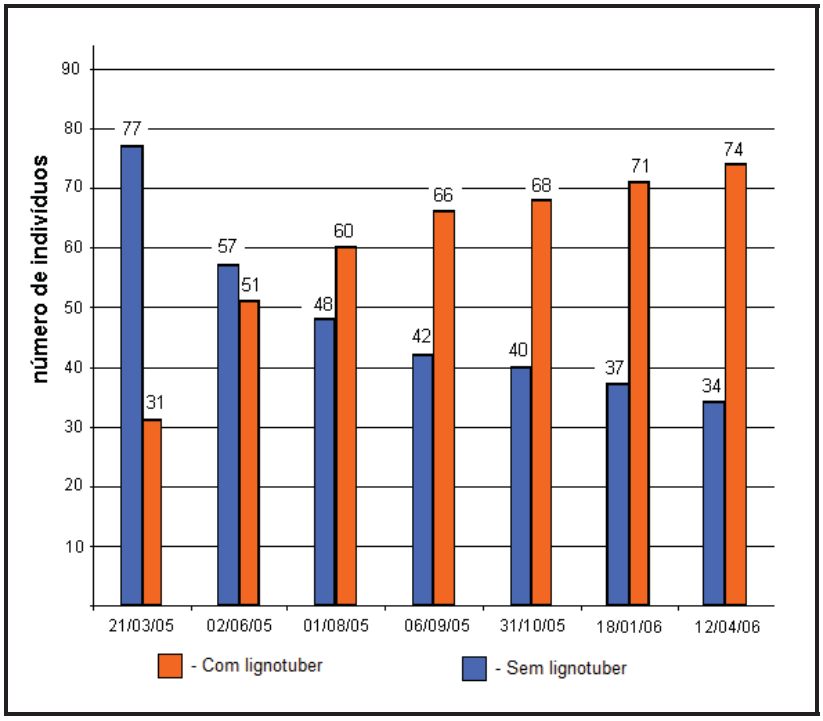

Figura 11: Gráfico com as avaliações realizadas em uma população F2 de *Eucalyptus urograndis*, submetida a estresse hídrico e nutricional, quanto à presença de lignotuber (Adaptado de Martins, 2006).

4.3. Análise da região 5' da sequência *ELig*

Foi realizada reação de PCR com os *primers* confeccionados para amplificar a região 5'da sequência *ELig* e compará-la à sequência do *contig* de EST utilizando dois indivíduos de *E. grandis*. O produto da amplificação está na Figura 12.

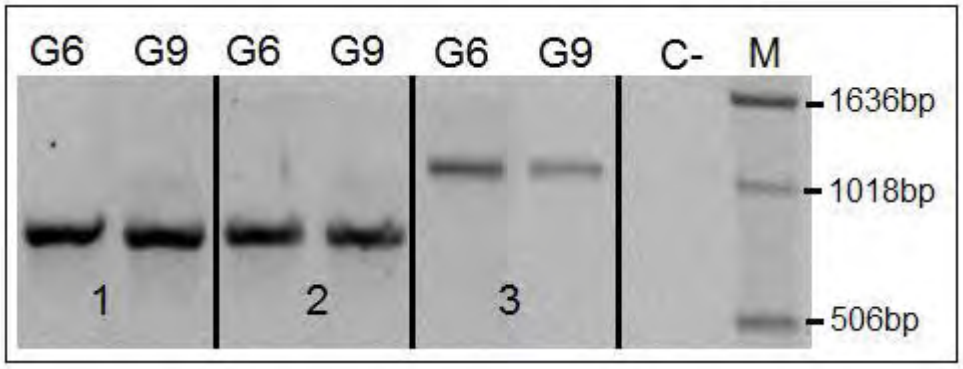

Figura 12: Gel de agarose 1% com o produto da amplificação com as combinações de *primers* desenhados para amplificação da região final do *contig* e inicial do *ELig*. **Em 1**: *primer EST-f* com *primer EST-r1*. **Em 2:** *primer EST-f* com *primer EST-r2*. **Em 3:** *primer EST-f* com *primer EST-r3*. **Legenda:** G6 e G9: indivíduos de *E. grandis*; C-: controle negativo; M: marcador de tamanhos de DNAs conhecido *ladder* 1Kb (Invitrogem®).

As bandas obtidas com a amplificação dos *primer EST-f* com *EST-r1* (combinação de *primers* 1 – CP-1) foram recortadas, purificadas, clonadas e sequenciadas. A sequência obtida foi comparada com as sequências do *contig* e com o *ELig*. Na região 5' a sequência CP-1 apresentou identidade à sequência do *contig* (identidade de 98% (573/580); *E-value*=0,0) e na região 3' a sequência obtida apresentou identidade à sequência *ELig*. Na posição 1298 da sequência do *contig* correspondente a posição 21 da sequência CP-1 há uma região de 103 nucleotídeos sem identidade entre as sequências, indicando se tratar de uma região não transcrita da sequência (*intron*). Isso indica que o *contig* é continuidade da sequência *ELig* no DNA genômico. Os alinhamentos são mostrados na Figura 13.

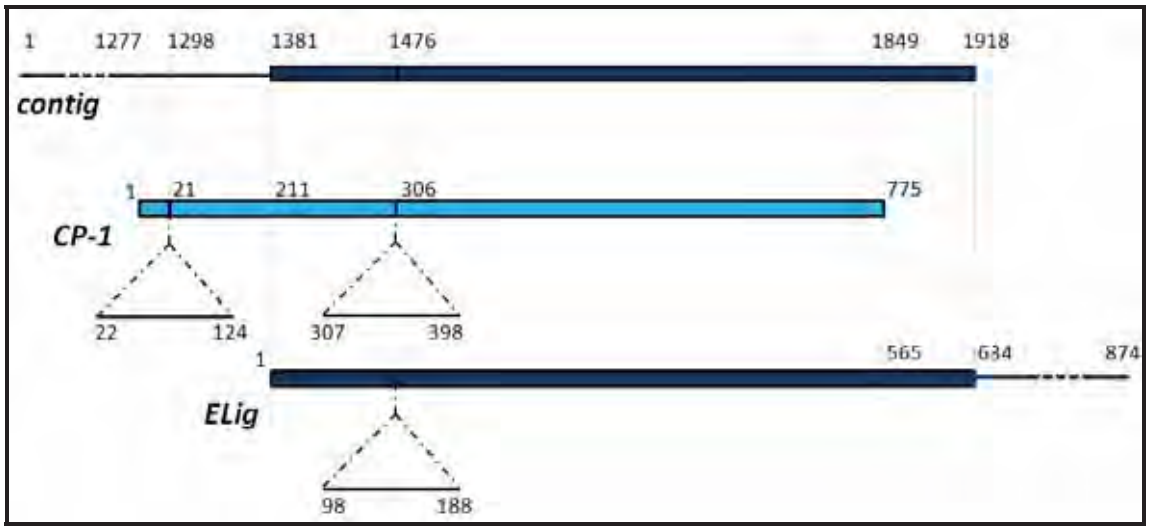

Figura 13: Alinhamento da sequência do *contig* com a sequência *ELig* e com a sequência da amplificação do *primer EST-f* com o *primer EST-r1* (CP-1). As barras mais grossas (barras azuis) mostram a região em que as sequências apresentam identidade entre si. Os números indicam as posições dos nucleotídeos nas sequências.

Esses *primers* foram analisados em reação de PCR com DNA de quatro indivíduos diferentes das espécies *E. grandis*, *E. urophylla* e *E. saligna*. Os resultados da amplificação são mostrados na Figura 14.

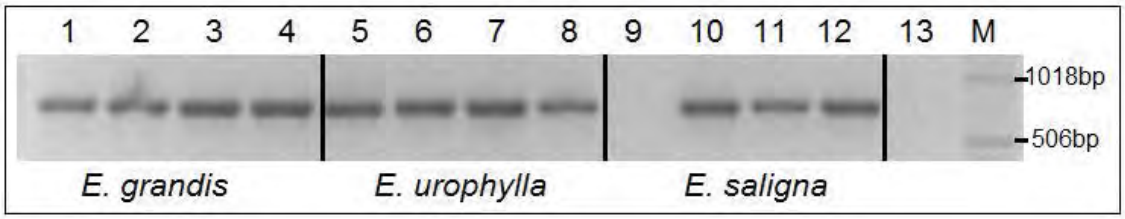

Figura 14: Gel de agarose 1% com o produto de amplificação dos *primers EST-f* e *EST-r1* com DNA de quatro indivíduos de cada espécie *E. grandis* (canaletas 1-4), *E. urophylla* (5-8) e *E. saligna* (9-12). O indivíduo 9 não apresentou amplificação do DNA. Na canaleta 13 encontra-se o controle negativo e na canaleta M encontra-se o padrão de DNA de tamanhos conhecido ladder 1Kb (Invitrogem®).

As bandas obtidas com a amplificação desses *primers* foram recortadas e sequenciadas em dois indivíduos de *E. saligna*, três de *E. urophylla* e quatro de *E. grandis* e as sequências obtidas foram analisadas. Não houve diferenças significativas nas sequências de DNA dos indivíduos das espécies utilizadas. A sequência do sítio de

pareamento do *primer ELig* na região 5’ é idêntica em todos os indivíduos das espécies analisadas, indicando que o polimorfismo está no sítio de pareamento da região 3’da sequência.

4.4. Número de cópias da sequência *ELig*

A técnica de *Southern blot* foi realizada para determinar o número de cópias da sequência *ELig* no genoma das espécies *E. grandis* e *E. urophylla*.

O produto da amplificação do DNA de *E. grandis* com o *primer ELig* foi utilizado como sonda. Para a realização da técnica, foram realizadas análises *in silico* no intuito de selecionar enzimas de restrição que não contenham sítios nessa sequência. As enzimas analisadas foram *AluI*, *DraI*, *HindIII*, *BamHI*, *EcoRI*, *NdeI*, *XbaI* e *PstI*. As enzimas *AluI*, *DraI* e *HindIII* apresentaram respectivamente 3, 2 e 1 sítios de restrição, já as enzimas *BamHI*, *PstI*, *EcoRI*, *NdeI* e *XbaI* não apresentam sítio de restrição na sequência *ELig*. Destas, as três últimas foram utilizadas para a análise de *Southern blot* (Figura 15A).

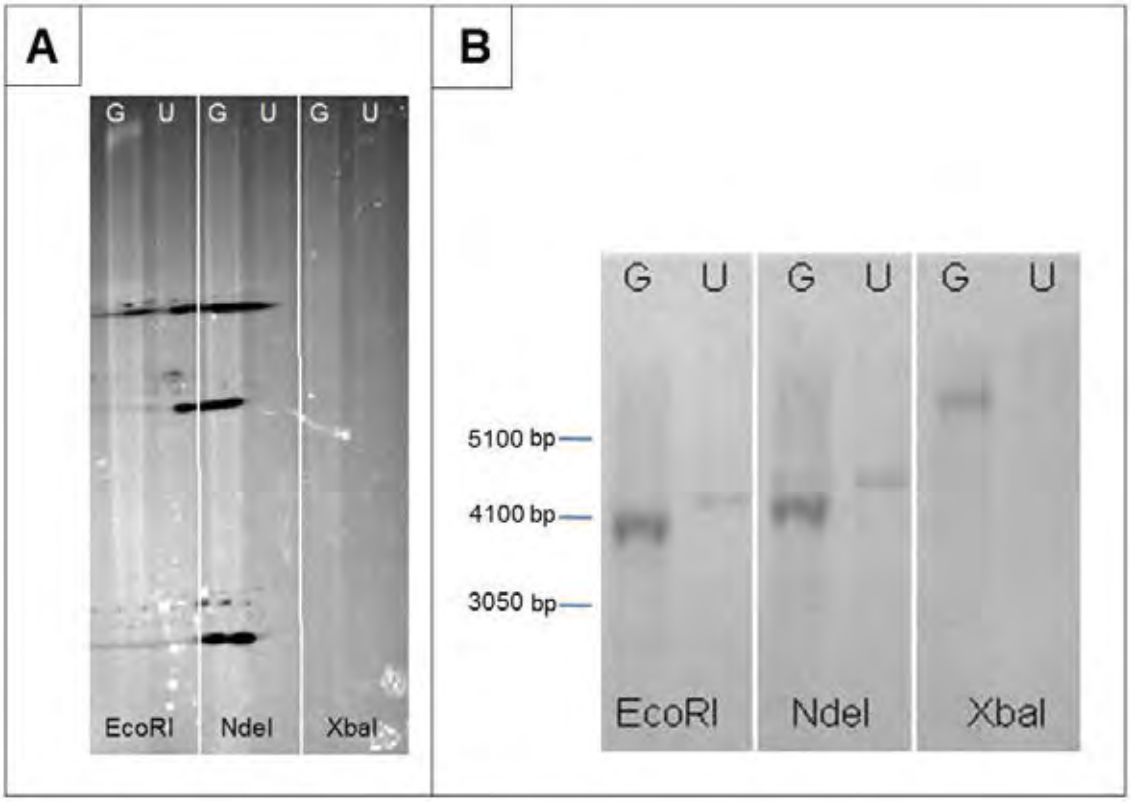

Figura 15: *Southern blot* da sequência *ELig* com o genoma de *E. grandis* e *E. urophylla*. **Em A:** Gel de agarose 1% com produto da digestão do DNA genomico de *E. grandis* (G) e *E. urophylla* (U) com as enzimas de restrição *EcoRI*, *NdeI* e *XbaI*. **Em B:** Chapa revelada após 15 minutos de exposição ao sinal quimioluminescente. Os resultados mostram a presença de apenas uma cópia da região *EgLig* no genoma de *E. grandis* (G) e *E. urophylla* (U). À direita estão indicados os tamanhos dos fragmentos em pares de base (bp).

O DNA de *E. grandis* e *E. urophylla* foi digerido com as enzimas de restrição *EcoRI*, *NdeI* e *XbaI* (em duplicata) (Figura 15A) e hibridizado com a sequência *ELig*. O resultado foi a presença de uma única banda em cada ensaio (Figura 15B), indicando que há apenas uma cópia dessa região no genoma de *E. grandis* e *E. urophylla*. As bandas apresentadas pela espécie *E. urophylla* foram maiores que as bandas apresentadas por *E. grandis*, com as três enzimas utilizadas, indicando que há diferenças significativas no genoma dessas duas espécies na região estudada.

4.5. *RACE 3`*

Somente a região 5’ da sequência *ELig* apresentou identidade ao *contig* (Figura 8), desta forma fez-se necessário analisar a região 3’ para a obtenção da sequência completa do gene.

A técnica foi realizada com RNA total extraído de caule e raiz de *E. grandis* e de lignotuber de *E.urophylla*. A Figura 16 apresenta os resultados das amplificações dos cDNAs com os *primers* GSP.

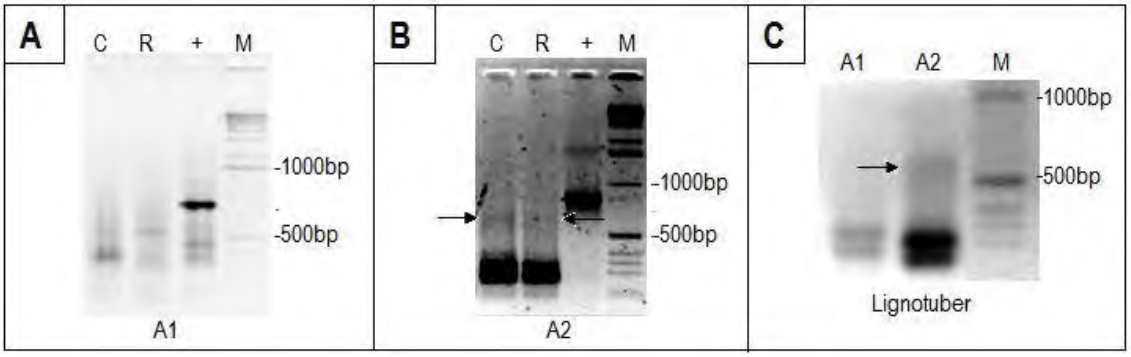

Figura 16: Gel de agarose com os produtos de amplificação da técnica de *RACE 3’* utilizando os *primers* gene específicos (GSP) desenhados a partir da sequência *ELig*. **Em A:** Primeira amplificação com o *primer* GSP interno com cDNA de caule e raiz de *E. grandis*. **Em B:** Amplificação *nested* com o *primer* GSP externo com caule e raiz de *E. grandis*. **Em C:** Amplificações com cDNA de lignotuber de *E. urophylla*. Legenda: C – caule; R – raiz; + – controle positivo; M – marcador molecular ladder 1Kb (Invitrogen®); A1 – amplificação primária; A2 – amplificação *nested*. As setas indicam as bandas maiores que foram recortadas e sequenciadas em cada ensaio.

A banda com maior peso molecular de cada um dos ensaios foi selecionada e sequenciada. O alinhamento com todas as sequências obtidas é mostrado na Figura 17:

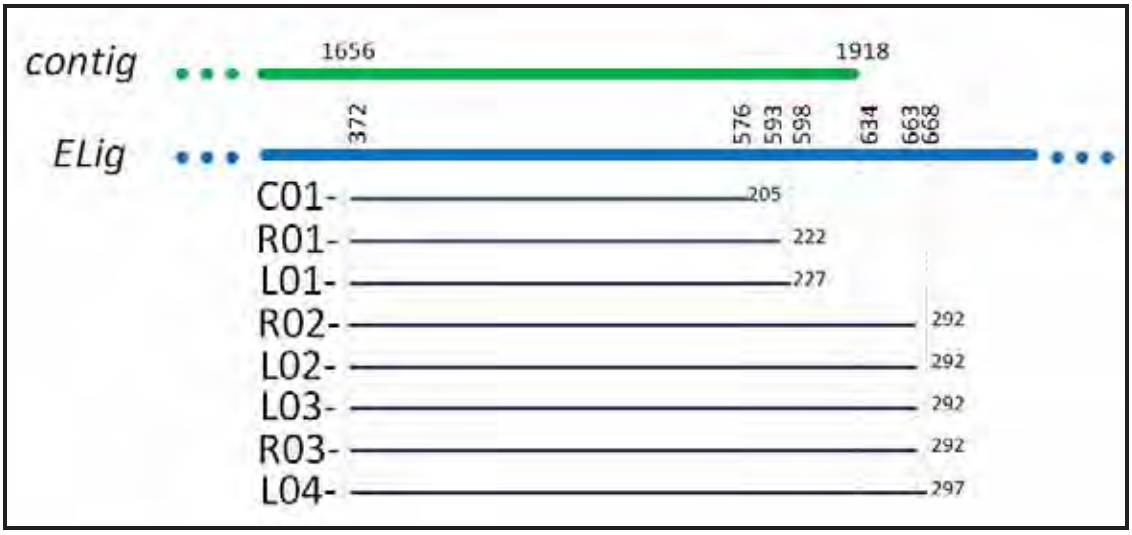

Figura 17: Alinhamento das sequências do *contig* e *ELig* (ambas de *E. grandis*) com os fragmentos obtidos com a técnica de *RACE* 3'. As sequências estão no sentido 5' – 3'. A barra azul representa a sequência do *ELig*, na posição 372nt é onde foi desenhado o *primer* gene específico externo (GSP-2); a barra verde representa a sequência do *contig* que apresenta identidade ao *ELig*. As barras mais finas indicam as sequências obtidas com a técnica de *RACE* 3'. Os números indicam a posição dos nucleotídeos das sequência descritas. Legenda: C: sequência derivada de cDNA de caule de *E. grandis*; R: sequências derivadas de cDNA de raiz de *E. grandis*; L: sequências derivadas de cDNA de lignotuber de *E. urophylla*.

Com cDNA de caule, foi obtida apenas uma sequência de 205nt (nucleotídeos). Com cDNA de raiz e lignotuber foram obtidas sequências de diferentes tamanhos, variando de 222 nt (raiz) a 297 nt (lignotuber).

Foram encontradas diferenças entre as sequências dos tecidos analisados somente no comprimento das sequências. As sequências de raiz tiveram identidade entre si apresentando apenas diferença no tamanho das sequências (222 nt e 292 nt), o mesmo ocorreu com as sequências de cDNA de lignotuber (227 nt, 292 nt e 297 nt).

As sequências apresentaram identidade com o *ELig* e com o *contig*. As sequências L2, L3, R2 e R3 na região 3' apresentaram 28 nucleotídeos além da sequência do *contig* e a L4 apresentou 33 (Figura 17). Essa região possui a mesma sequência que o *ELig*.

Todas as sequências obtidas com a técnica de *RACE* 3' apresentaram cauda poliA na porção 3', indicando que essa região seja responsável pela transcrição de RNAs de diferentes tamanhos na 3'UTR (*Untranslated Region*). Polimorfismo de comprimento na região 3' UTR indicam regulação na expressão de genes. Alguns elementos regulatórios podem participar desse procedimento, os *riboswitch* são exemplos de elementos regulatórios que controlam a formação de transcritos com comprimentos alternativos na regiao 3' UTR, que afetam o acúmulo de mRNA e a produção de proteínas (Wachter *et al.*, 2007).

4.6. *Genome Walking*

A técnica de *Genome Walking* foi realizada nos sentidos 5' e 3', a partir de *primers* confeccionados nas posições de 222 nt (para o sentido 5') e 372 nt (para o sentido 3') da sequência *ELig*. A Figura 18 apresenta o resultado da amplificação *nested* das bibliotecas de digestão de DNA genômico que apresentaram bandas com os *primers GSP* no sentido 5'e 3'. A amplificação com as demais bibliotecas não apresentaram bandas, sendo visualizado um arrasto no gel. As bandas indicadas foram recortadas, clonadas e sequenciadas.

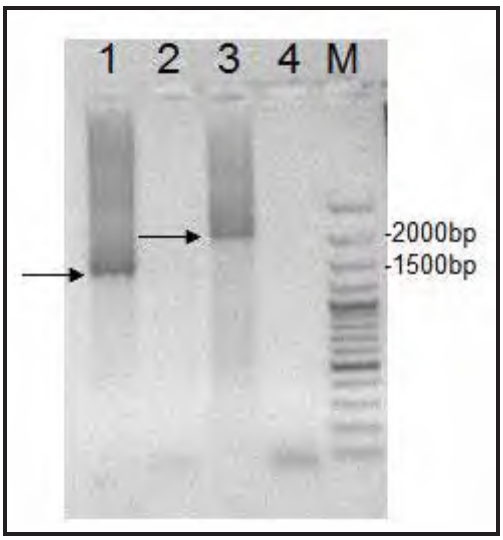

Figura 18: Gel agarose 1% com o resultado da amplificação *nested* da técnica de *Genome Walking*. As sentas indicam as bandas que foram recortadas, clonadas e sequenciadas. Legenda: 1 – produto da técnica *Genome Walking* no sentido 5' (biblioteca de *EcoRV*); 2 – controle negativo do sentido 5'; 3 – produto da técnica *Genome Walking* no sentido 3' (biblioteca de *PruII*); 4 – controle negativo do sentido 3'; M – marcador molecular com padrão de fragmentos de tamanhos conhecidos de DNA (ladder 100 bp Fermentas®).

No sentido 5' foram obtidas duas sequência com a técnica, uma de 661 nt e outra de 678 nt, porém o alongamento não foi suficiente para se formar o consenso entre elas, umas vez que a banda recortada do gel possuía cerca de 1400 nt de tamanho (Figura 18). A sequência de 661 nt (*GW5'*) apresentou identidade com o *ELig* (Figura 19), dessa sequência, 439 nt foram estendidos além da sequência *ELig*. Na posição 222 da dequência *GW5'* foi possível detectar a sequência do sítio de pareamento do *primer* RAPD.

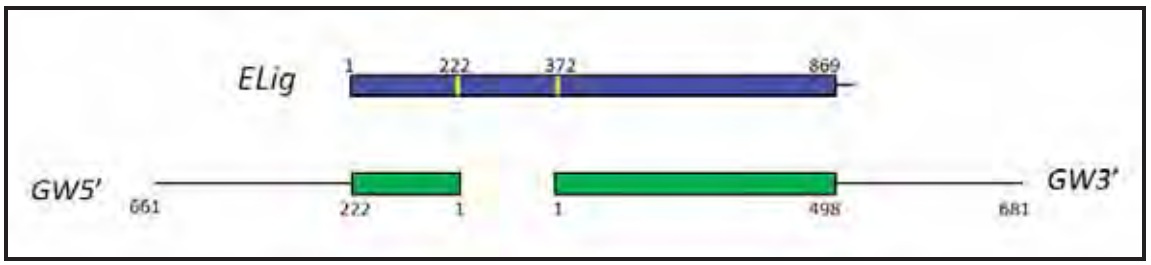

Figura 19: Alinhamento da sequência *ELig* com as sequências obtidas com a técnica de *Genome Walking* no sentido 5' (*GW5'*) e no sentido 3' (*GW3'*). A barra azul representa a sequência *ELig* e as barras verdes representam as sequências obtidas com *Genome Walking*, as barras grossas indicam as regiões onde as sequências apresentam identidade entre elas. Os números indicam as posições em pares de base das sequências.

No sentido 3' também foram obtidas duas sequência, uma de 663 nt e outra de 680 nt, porém o alongamento não foi suficiente para se formar o consenso entre elas, umas vez que a banda recortada do gel possuía cerca de 2.000 nt de tamanho (Figura 18). A sequência de 680 nt (*GW3'*) apresentou identidade com o *ELig* (Figura 19), dessa sequência, foi possível estender 182 nt além da sequência *ELig*. Comparando as duas sequências, foi possível identificar três inserções na região final da sequência *ELig* totalizando 32 nucleotídeos, além disso, o sítio de pareamento do *primer* RAPD, está presente na sequência *GW3'*, porém apenas as seis primeiras bases, as quatro últimas bases são diferentes, o que mostra que a região 3' da sequência *ELig* não é totalmente conservada.

Para conseguir juntar as sequências formando um consenso, seria necessário desenhar pares de *primers* na extremidade das sequências para alongá-las, porém, no período em que estávamos realizando a técnica de *Genome Walking*, o genoma completo de *Eucalyptus* foi disponibilizado, permitindo a análise dessas regiões *in silico*. As análises serão apresentadas abaixo.

4.7. Análise das sequências no *EucalyptusDB*

Foi realizada uma busca de sequências no banco de dados *Phytozome* de *Eucalyptus grandis* através da ferramenta de *BLAST* (Stephen *et al.*, 1997) utilizando a sequência *ELig* como *query*, isso possibilitou a identificação de um único *scaffold* com identidade a sequência *ELig* (*Bit-score* de 1310 e *E-value* de 0,0). A identificação de um único *scaffold*, confirma os resultados obtidos com o *Southern blot*, onde foi encontrada apenas uma cópia dessa sequência no genoma de *E. grandis* e *E. urophylla*.

A identidade encontrada entre o *scaffold* e a sequência *ELig* foi de 95% (dos 873 nt da sequência *ELig*, 822 nt foram idênticos; E-value=0,0). Da posição 1 até a posição 779 do *ELig*, foi identificada uma única substituição "G/A" na posição 295 e uma inserção de "AA" na posição 679, a partir da posição 740 as sequências apresentam menor identidade com mais substituições e *indels*, foram encontrados 3 GAPs (total de 32 nucleotídeos) nessa região (Figura 20). O sítio do *primer* RAPD foi localizado inteiro na posição 5' da sequência *ELig*, porém na posição 3' apenas os 6 primeiros nucleotídeos do *primer* estavam presentes na sequência do banco de dados, os outros quatro nucleotídeos estão diferentes nas duas sequências (Figura 20). Confirmando nossa hipótese inicial de que o polimorfismo está presente na posição 3'das sequências.

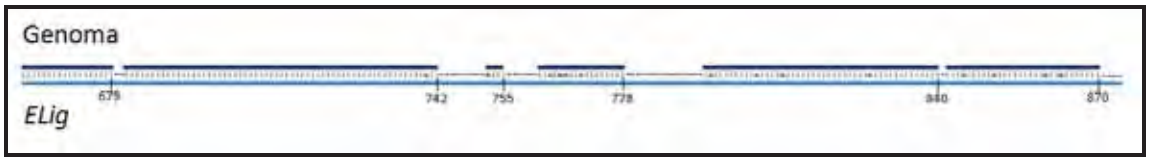

Figura 20: Alinhamento entre a região do genoma completo com identidade ao *ELig* e a sequência *ELig*. Mostrando a região 3' do *ELig* onde ocorre maior diferença entre as duas sequências, os números indicam a posição do nucleotídeo na sequência *ELig*. Observar os últimos quatro nucleotídeos, presentes na sequência *ELig* e ausentes na sequuência do genoma, fazem parte do sítio de pareamento do *primer* RAPD. Legenda: n: substituição de base; - GAP.

- Genome Walking

Com o genoma completo publicado, foi possível alinhar as sequências obtidas com a técnica de *Genome Walking* e o marcador *ELig*. O alinhamento das sequências está apresentado na Figura 21.

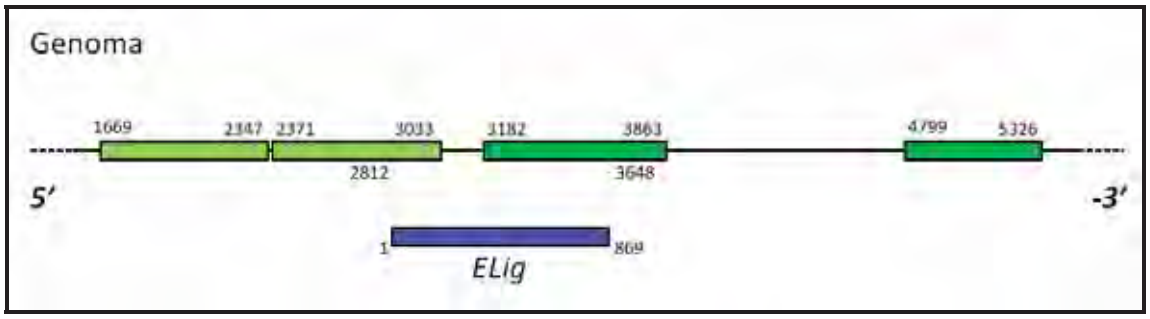

Figura 21: Alinhamento das sequências obtidas com a técnica de *Genome Walking* (*GW*), comparando-as com a sequência do genoma completo e a sequência *ELig*. As barras verdes indicam as sequências *GW* sobre a sequência do genoma de eucalipto. As duas barras da esquerda indicam as sequências *GW5'* e as duas da direita as sequências *GW3'*. A barra azul representa a sequência *ELig*. Os números indicam as posições dos nucleotídeos nas sequências.

As sequências obtidas com o *Genome Walking* apresentaram-se alinhadas conforme o esperado, as duas sequências da região 5' estavam distantes entre si 23 nucleotídeos (Figura 21) e tem identidade de 98% com a sequência do genoma (num total de 1343 nucleotídeos; E-value de 0,0). As duas sequências da região 3' estavam distantes entre si 935 nucleotídeos (Figura 21) e tem identidade de 99% com a sequência do genoma (total de 1210 nucleotídeos; E-value de 0,0).

- Predição de Proteínas

A sequência do *scaffold* do banco de dados *Phytozome* que apresentou identidade a sequência *ELig*, foi traduzida em possíveis *ORFs* (*Open Reading Frames*) para comparação com banco de dados públicos de domínios de proteínas (Finn *et al.* 2010), apresentando grande identidade com o domínio das família de proteínas *DnaJ* (Bit-score de 26,7 e E-value de $2,6e^{-06}$). O gene predito apresenta 3448nt e está dividido em seis exons e codifica uma proteína com 537 aminoácidos (Figura 22).

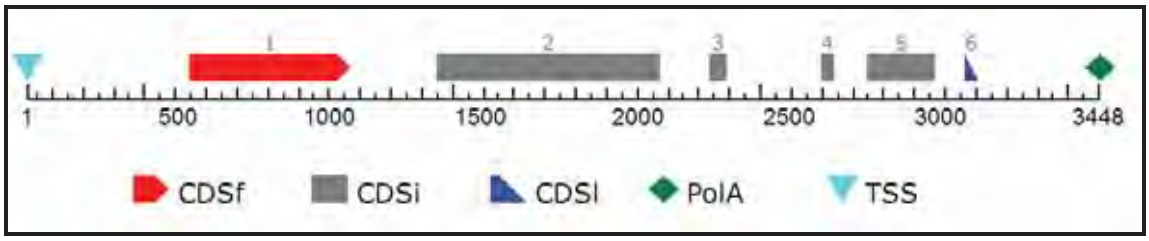

Figura 22: Predição de *ORFs* a partir da sequência do *scaffold* do banco de dados *Phytozome* que apresentou identidade à sequência *ELig*. Legenda: CDSf – (First Coding Segment) posição do primeiro exon; CDSi – (Internal Coding Segment) posição dos exons internos; CDSl – (Last Coding Segment) posição do último exon; PolA – posição da cauda poliA; TSS – posição de início da transcrição.

A sequência do *scaffold* que apresentou identidade com a sequência *ELig* foi comparada com a sequência do *contig* do banco de dados FOREST (Vicentini *et al.*, 2005). O alinhamento das sequências com a predição de proteínas estão representados na Figura 23.

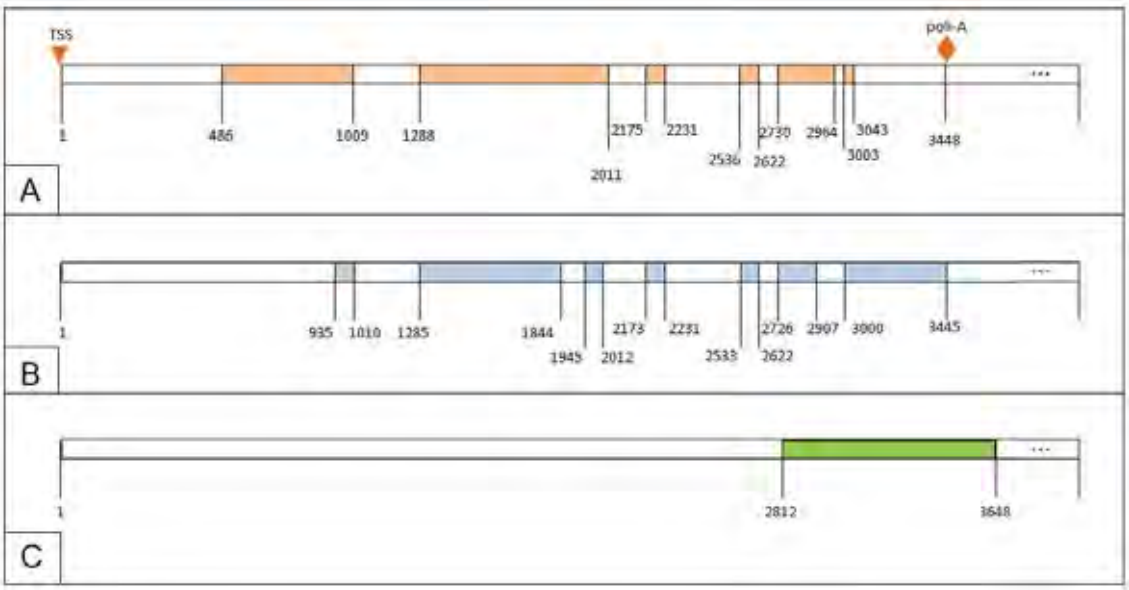

Figura 23: Comparação das sequências do *scaffold* do banco de dados *Phytozome* que apresentou identidade à sequência *ELig* com o *contig EST* e com a sequência *ELig*. **Em A**: predição da proteína obtida no software *Softberry* (http://softberry.com/) coma sequência do *scaffold*. **Em B**: alinhamento com a sequência do EST obtida no banco de dados do FOREST (Vicentini *et al.*, 2005). **Em C**: alinhamento com a sequência *ELig*. As barras coloridas indicam a região onde há similaridade entre as sequências comparadas. Observando as duas primeiras barras podemos observar alta similaridade entre a proteína predita pelo Softberry e o *contig*. Legenda: PoliA – Posição da cauda poliA; TSS – Posição de início da transcrição. Obs.: As sequências estão apresentadas no sentido 5'-3'.

Com essa análise foi possível observar que a sequência *ELig* está na região final do gene predito nos dois últimos exons incluindo a região onde será adicionada a cauda poli A e ainda uma região externa do gene e é nessa região que se encontra a origem do polimorfismo inicial detectado (*primer* RADP) (Figura 23C).

Também foi possível observar a alta similaridade da proteína predita a partir da sequência do genoma com a proteína predita a partir do *contig*, ambas apresentaram o domínio de proteínas *DnaJ*, porém com algumas diferenças nos tamanhos e números de exons (Figura 23A e 23B).

Com a sequência completa dessa região genômica, é possível desenhar *primers* que cubram toda essa região e avaliar se existem polimorfismos entre indivíduos das diferentes espécies de eucalipto e do híbrido *E. urograndis* com segregação para a presença do lignotuber e verificar a correlação destes polimorfismos com a presença do lignotuber.

As regiões *upstream* e *downstream* da sequência do *scaffold* do banco de dados *Phytozome* com identidade à sequência *ELig* também foram analisadas quanto a presença de genes, uma região de cem mil pares de base para cada sentido foi utilizada na análise. Foram observados 36 genes nessa região (Figura 24).

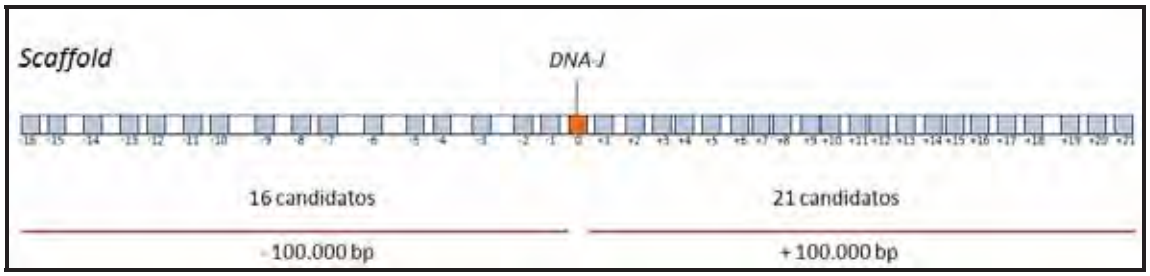

Figura 24: Esquema do bloco gênico com os genes candidatos da vizinhança do *ELig*, usando uma distância de 100 Kb para cada sentido 5'e 3'. O *ELig* está esquematizado em laranja e os genes candidatos em azul.

O gene *ELig – DnaJ* foi nomeado como zero e a partir dele no sentido 5' os genes receberam a nomenclatura de -1 a -16, sendo -1 o gene predito mais próximo do gene *ELig – DnaJ* e -16 o gene mais distante. E no sentido 3' os genes receberam a nomenclatura +1 a +21, sendo o gene +1 mais próximo do gene *ELig – DnaJ* e +21 o mais distante.

As sequências de proteínas preditas foram comparadas com banco de dados NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) e banco de domínios de proteínas através da ferramenta *blastp*. Dos 36 genes preditos, 24 não apresentaram domínios de proteínas, e nem similaridade com sequências de proteínas nos bancos de dados analisados. Os que apresentaram sequências com domínios de proteínas conservados, ou similaridade a sequências de proteínas depositadas nestes bancos de dados estão descritos na Tabela 2 e na Figura 25.

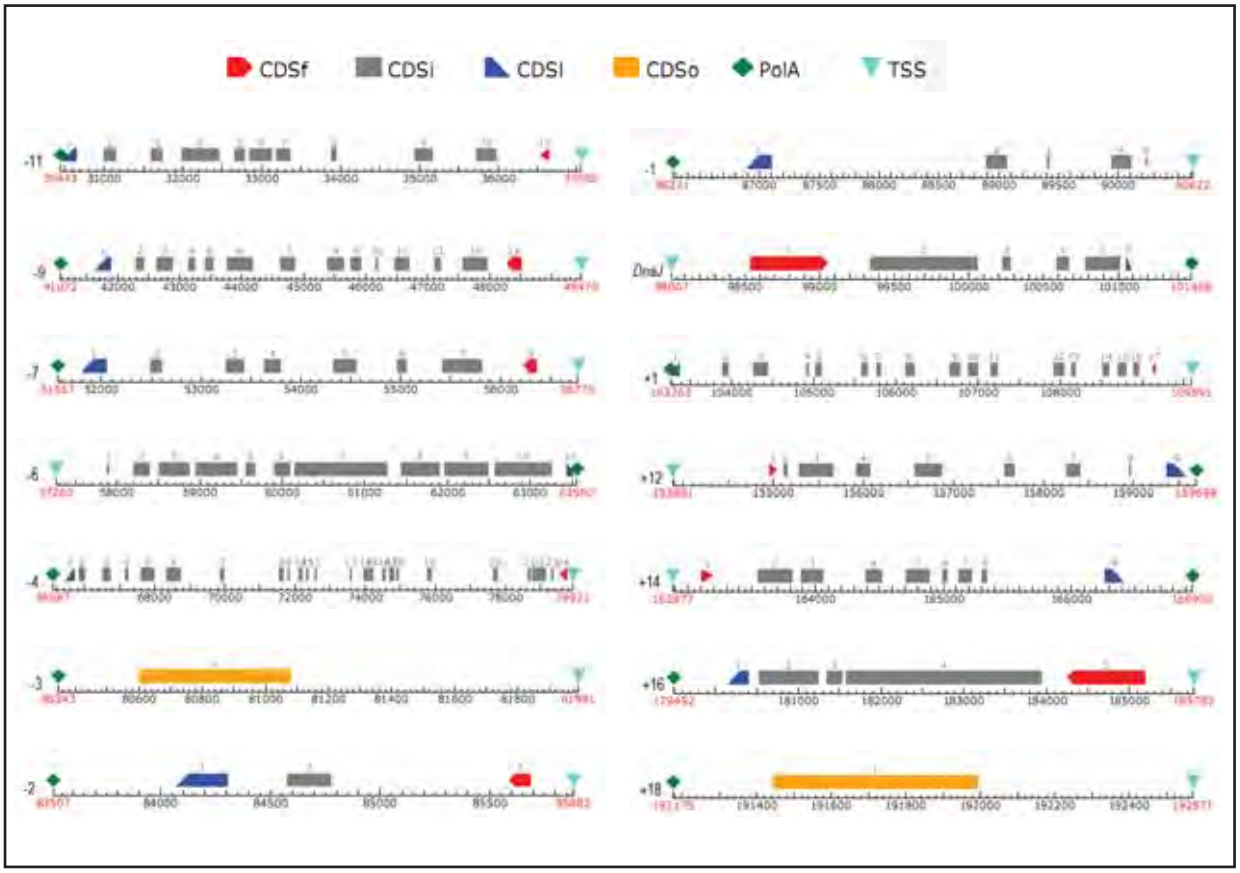

Figura 25: Genes candidatos do banco de dados de *Eucalyptus grandis* (*EucalyptusDB*) da vizinhança do *ELig*, usando uma distância de 100 Kb para cada sentido (5'e 3'). Legenda: CDSf – (First Coding Segment) Posição do primeiro exon; CDSi – (Internal Coding Segment) Posição dos exons internos; CDSl – (Last Coding Segment) Posição do último exon; CDSo – (Only Coding Segment) Posição do exon único; PolA – Posição da cauda poliA; TSS – Posição de início da transcrição.

Tabela 2: Análise das proximidades da sequência com identidade ao *ELig* no genoma completo de *Eucalyptus grandis*.

	Proteína	Organismo	Score	Id.	E-value	Acesso	Dist. *ELig*	Domínios
-11	Hipotética	*Ricinus communis*	332	32%	8e-^{52}	XP_002532981.1	61.957 nt	na
-9	Hipotética	*Vitis vinifera*	571	93%	1e-^{45}	XP_002269140.1	48.537 nt	(pfam02362) *B3 DNA Binding*
-7	Hipotética	*Vitis vinifera*	294	70%	1e-^{47}	XP_002269140.1	41.232 nt	(pfam02362) *B3 DNA Binding*
-6	Desconhecida	*Arabidopsis thaliana*	513	64%	8e-^{143}	AAG50806.1	34.447 nt	(pfam03372) Família Fosfatase Endonuclease Exonuclease (cd01650) Retrotransposon não-LTR
-4	Desconhecida	*Vitis vinifera*	495	31%	3e-^{77}	CBI30946.3	18.086 nt	(cl00446) Superfamília Metalo-beta-lactamase (pfam02362) *B3 DNA Binding*
-3	Hipotética	*Trichophyton verrucosum*	37.4	30%	0.65	XP_003018995.1	16.016 nt	na
-2	Hipotética	*Leifsonia xyli subsp. Xyli*	37.4	22%	0.72	YP_062541.1	12.125 nt	na
-1	Proteína família carboxipeptidase zinco	*Tetrahymena thermophila*	38.1	33%	0.64	XP_001022081.1	7.385 nt	na
0	*DnaJ – heat shock protein*	*Arabidopsis lyrata*	259	75%	9e-31	XP_002893610.1	-	(cd06257) Domínio *DnaJ*
+1	Hipotética	*Vitis vinifera*	710	75%	0.0	XP_002279222.1	1.775 nt	(cd00180) Domínio Catalítico Proteína quinase (cl04277) Domínio NAF
+12	Predita	*Populus trichocarpa*	523	70%	2e-^{31}	XP_002331306.1	52.393 nt	(cl09925) Domínio catalítico PK
+14	Hipotética	*Vitis vinifera*	190	86%	3e-^{46}	XP_002273291.1	61.389 nt	na
+16	Retroelemento não LTR transcriptase reversa putativa	*Arabidopsis thaliana*	750	82%	0.0	AAD20714.1	78.004 nt	(cl00490) Família Fosfatase Endonuclease Exonuclease (cl02808) RT_like: Transcriptase Reversa (RT, RNA-dependent DNA polymerase) _like family.
+18	Hipotética	*Vitis vinifera*	43.1	25%	0.027	XP_002275498.1	89.687 nt	na

Legenda: Id. – Identidade entre a proteína predita e a sequência *DnaJ*; Dist. *ELig* – distância da proteína predita e a sequência *DnaJ*; na – não apresenta identidade a domínios de proteínas; LTR (long terminal repeat); RT – reverse transcriptase; PKs – Proteína quinase.;

Dos genes preditos, seis apresentaram similaridade a proteínas sem domínios conservados, são eles: -11 (*Ricinus communis*); -3 (*Tricophyton verrucosum*); -2 (*Leifsonia xyli*); -1 (*Tetrahymena thermophila*); +14 (*Vitis vinifera*) e +18 (*Vitis vinifera*). As demais apresentaram similaridade a proteínas com domínios conservados, são elas: -9 (*Arabidopsis thaliana*); -7 (*Vitis vinífera*); -6 (*Arabdopisi thaliana*); -4 (*Vitis vinifera*); +1 (*Vitis vinifera*); +12 (*Populus trichocarpa*); +16 (*Arabidopsis thaliana*).

A proteína predita -9 além da identidade com a proteína hipotética de *Vitis vinifera*, ela também apresentou similaridade (76%, região de 399 resíduos e E-value de $7e^{-27}$) com uma proteína repressora de vernalização de *Arabidopsis thaliana*. Seis regiões dessa proteína apresentaram identidade ao domínio de proteínas *B3 DNA Binding* (pfam020362), esse domínio pertence a uma família de fatores de transcrição de plantas, e é importante no desenvolvimento das plantas, algumas proteínas dessa família também apresentam um domínio de ligação ao DNA.

As proteínas que apresentam o domínio *B3* são sensíveis aos fitormônios ácido abscísico e auxina, que desempenham um papel fundamental nos processos de desenvolvimento como o crescimento das plantas e a maturação das sementes (McCarty *et al*., 1989; Ulmasov *et al*., 1997). Três principais classes de fatores de transcrição contendo domínios *B3* foram identificados, incluindo os factores *ABI3* e *VP1* (*ABI3/VP1- like factors*), as proteínas similares à proteína de *Arabidopsis*, *RAV1* (família *RAV-like*), e factores de resposta auxina (*ARFs*) (Riechmann *et al*., 2000, Waltner *et al.*, 2005). Os fatores de transcrição *RAV1* e *RAV2* estão envolvidos na adaptação da planta a uma variedade de estímulos ambientais (Kagaya e Hattori, 2009), estudos tem mostrado que a expressão de *RAV1* e *RAV2* é estimulada por vários estímulos ambientais, como por exemplo: ataque de patógenos, baixas temperaturas,

estresse hídrico e salino, falta de luz e ferimentos na planta (Fowler and Thomashow, 2002; Lee *et al*., 2005; Sohn *et al*., 2006; Kagaya e Hattori, 2009).

Assim como a proteína predita -9, a proteína -7 também apresentou identidade ao domínio de proteínas *B3 DNA Binding* (pfam020362).

A proteína -6 apresentou identidade a dois domínios: (pfam03372) Família Fosfatase Endonuclease Exonuclease e (cd01650) *RT_nLTR*: *Non-LTR retrotransposon* e *non-LTR retrovirus RT*. A Família Fosfatase Endonuclease Exonuclease inclui endonucleases dependentes de magnésio e um grande número de fosfatases envolvidas na sinalização intracelular, essa família inclui proteínas endonuclease AP, proteínas DNAse I, Sinaptojanina (uma fosfatase inositol-1,4,5-trisfosfato), Esfingoielinase e *Nocturnin* (Lahm e Suck, 1991; Mol *et al*., 1995; Dlakić, 2000). A subfamília *RT_nLTR* (Retrotransposon não-*LTR – Long Terminal Repeat*) contém retrotransposons sem LTR e RTs retrovírus sem LTR. Esta classe de retrotransposons é abundante em eucariontes. Esses elementos não têm repetições terminais em tandem, em vez disso, frequentemente terminam com uma cauda poli-A no sentido 3', enquanto seus terminais 5', muitas vezes contem deleções (Eickbush, e Jamburuthugoda, 2008).

A proteína predita -4 apresentou cinco regiões com identidade a domínios de proteínas duas regiões tiveram identidade ao domínio da (cl00446) superfamília das metalo-beta-lactamases (MBLs) e três regiões com o domínio *B3 DNA Binding*. Proteínas da superfamília das MBLs são responsáveis pela resistência de algumas bactérias a antibióticos betalactâmicos, através da hidrólise dos betalactâmicos disponíveis comercialmente (Bush, 1998, Toleman *et al*., 2002).

A proteína predita +1 foi a proteína mais próxima do gene com identidade ao *ELig* e domínio *DnaJ*. Essa proteína apresentou duas regiões da sua sequência com identidade a domínios de proteínas: domínio catalítico proteína quinase (PK) e domínio NAF. As proteínas quinases são enzimas que catalisam a fosforilação de outras proteínas por meio da transferência de um grupo fosfato de ATP ou GTP (em raros casos), para resíduos de tirosina (Tyr), treonina (Thr) e serina (Ser) (Nishizuka, 1986). As proteínas quinases compõem a maior família de proteínas nos eucariontes e é um componente fundamental da cascata de "comunicação" que ocorre no controle intracelular, na regulação e transdução de sinais (Mellor e Parker, 1998). A fosforilação de proteínas é um mecanismo pelo qual uma grande variedade de proteínas celulares, tais como enzimas e canais de membrana, são reversivelmente reguladas em resposta a certos estímulos. A regulação pode ocorrer através do controle transcricional, ou através de alterações químicas e estruturais das proteínas, dentre outros. As PKs regulam muitos processos celulares, incluindo proliferação, divisão, diferenciação, mobilidade, sobrevivência, metabolismo, a progressão do ciclo celular, rearranjo do citoesqueleto, a imunidade, e as funções neuronais (Steinberg, 2008).

A proteína predita +1 também apresentou identidade ao domínio NAF. O domínio NAF é um domínio de 24 aminoácidos. Este domínio possui quinases heterólogas e é encontrado em um subgrupo de proteínas quinases serina-treonina (CIPKs) específicas de plantas, que interagem com proteínas calcineurina, cálcio B-sensor (CBLs), envolvidas em diferentes processos de sinalização, incluindo luz, hormonios e resposta a estresse de diversas espécies de plantas. (Albrecht, 2001).

A proteína predita +12 também apresentou identidade ao domínio catalítico PK.

A proteína predita +16 apresentou duas regiões com identidade a domínios de proteínas, uma com a família fosfatase Endonuclease Exonuclease e a outra com Transcriptase Reversa (RT, RNA-dependente de DNA polimerase).

Dentre os genes identificados localizados, o gene +1 é um forte candidato para análises de expressão, uma vez que as PKs estão envolvidas em uma diversidade de processos regulatórios e se encontra mais próxima da sequência *ELig*. Os demais genes preditos também são considerados candidatos e podem ser utilizados em experimentos de expressão gênica, o que permitira investigar se ocorre participação de algum destes genes na formação do lignotuber em eucalipto. Alem disso, outros candidatos relacionados a estresse podem ser identificados na literatura ou em banco de dados, e a partir de técnicas de bioinformática esses genes podem ser localizados no genoma do *Eucalyptus* através do genoma completo do eucalipto e podem ser utilizados nesses experimentos.

A variação genética dentro de uma espécie pode ser devida a variações alélicas de uma determinada região genômica, ou variações no nível de expressão destas regiões. Os dados obtidos nesse trabalho identificaram regiões candidatas à formação do órgão lignotuber em eucalipto, bem como tolerância a estresse abiótico. Experimentos com maior controle das condições ambientais deverão ser conduzidos para estabelecer essas correlações. Da mesma forma análise das variações alélicas e de expressão dessas regiões genômicas candidatas devem ser realizadas, novas metodologias de análise molecular em larga escala estão disponíveis e poderão ser utilizadas nesse estudo.

5. CONCLUSÕES

As conclusões obtidas com o trabalho foram as seguintes:

- O marcador *ELig* é eficiente em identificar indivíduos de *E. grandis* sem lignotuber e pode ser utilizado em seleção assistida por marcadores.

- A região *ELig* apresenta identidade a um *contig* de EST de *Eucalyptus* com o domínio de proteínas *DnaJ*.

- A região *ELig* está presente em cópia única no genoma de *Eucalyptus grandis* e *E urophylla*.

- O fragmento gerado pela amplificação dos *primers ELig* inclui parte da região transcrita, a região 3' UTR, e uma região após a 3' UTR de um gene com domínio *DnaJ*.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ALBRECHT, V.; RITZ, O.; LINDER, S.; HARTER, K.; KUDLA, J. The NAF domain defines a novel protein–protein interaction module conserved in Ca^{2+}-regulated kinases **EMBO J.**, v.20, n.5, p1051–1063, 2001.

ALTSCHUL, S.F.; MADDEN, T.L.; SCHÄFFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, v.25, n.17, p.3389-3402, 1997.

ALVIM, F.C. CAROLINO, S.M.B.; CASCARDO, J.M.C.;NUNES, C.C.; MARTINEZ, C.A.; OTONI, W.C.; FONTES, E.P.B. Enhanced accumulation of BiP in transgenic plants confers tolerance to water stress. **Plant Physiology**, v.126, p.1042-1054, 2004.

AOKI, K.; KRAGLER, F.; XOCONOSTLE-CA´ ZARES, B;. LUCAS, W.J. A subclass of plant heat shock cognate 70 chaperones carries a motif that facilitates trafficking through plasmodesmata **PNAS.**, v.99, n.25, p.16342-16347, 2002.

BAMBER, R.K., MULLETTE, K.J. Studies of the lignotubers of *Eucalyptus gummifera* (Gaertn e Hochr) II Anatomy. **Australian Journal of Botany**, v.26, p.15-22, 1978.

BORK, P.; SANDER, C.; VALENCIA, A.; BUKAU, B. A module of the DnaJ heat shock proteins found in malaria parasites. **Trends Biochem Sci**., v.17, n.4, p.129, 1992.

BUSH, K. Metallo-b-lactamases: a class apart. **Clin Infect Dis**., v.27, n.1, p. 48-53, 1998.

CANADELL, J.; LÓPEZ-SORIA, L. Lignotuber reserves support regrowth following clipping of two mediterranean shrubs. **Functional Ecology**, v.12, p. 31-38, 1998.

CAPLAN, J.A.; CYR, D.M.; DOUGLAS, M.G. Eukaryotic Homologues of *Escherichia coli* dnaJ: A Diverse Protein Family That Functions with HSP70 Estresse Proteins. **Molecular Biology of the Cell**, v.4:p 555-563, 1993.

CHANG, S.; PURYEAR, J.; CAIRNEY, J. A simple and efficient method for isolating RNA from pine trees. **Plant Molecular Biology Reporter**, v.11, n.2, p.113-116, 1993.

CHEETHAM, M.E.; CAPLAN, A.J. Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. **Cell Estresse e Chaperones**, v.3, n.1, p 28-36, 1998.

CHO, E.K.; HONG, C.B. Over-expression of tobacco NtHSP70-1 contributes to drought-stress tolerance in plants. **Physiology and Biochemestry**, v.25, p.349-358, 2006.

CYR, D.M.; LANGER, T.; DOUGLAS, M.G. DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. **Trends Biochem Sci**, v.19, n.4, p.176-181, 1994.

DLAKIĆ, M. Functionally unrelated signalling proteins contain a fold similar to Mg2+-dependent endonucleases. **Trends Biochem Sci.**, v.25, n.6, p.272-273, 2000.

DOS REIS, G.G., REIS, M.G.F. Fisiologia da brotação de eucalipto com ênfase nas suas relações hídricas. **Série Técnica IPEF**, v.11, n. 30, p. 9-22, 1997.

EICKBUSH, T.H.; JAMBURUTHUGODA, V.K. The diversity of retrotransposons and the properties of their reverse transcriptases. **Virus Res., v.**134, n.1-2, p.221–234, 2008.

ELDRIDGE, K. **Eucalyptus Domestication and Breeding.** Claredon Press, 1994, 288 p.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Brasília, EMBRAPA - CENARGEN, 1995, 220p.

FINN, R.D.; MISTRY, J.; TATE, J.; COGGILL, P.; HEGER, A.; POLLINGTON, J.E.; GAVIN, O.L.; GUNESEKARAN, P.; CERIC, G.; FORSLUND, K.; HOLM, L.; SONNHAMMER, E.L.L.; EDDY, S.R.; BATEMAN, A. The Pfam protein families database: **Nucleic Acids Research**, v.38, (suppl 1), p.D211-D222, 2010.

FOWLER, S.; THOMASHOW, M.F. Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. **Plant Cell**, v.14, p.1675–1690, 2002.

GRATTAPAGLIA, D.; KIRST, M. *Eucalyptus* applied genomics: from gene sequences to breeding tools. **New Phytologist**, v.179, n.4, p.911-929, 2008.

HAMMOND, C.; HELENIUS, A. Quality control in the secretory pathway. *Current Opinion* in **Cell Biology**, v.7, p.523-529, 1995.

KAGAYA, Y.; HATTORI, T. *rabidopsis* transcription factors, RAV1 and RAV2, are regulated by touch-related stimuli in a dose-dependent and biphasic manner. **Genes Genet. Syst.,** v. 84, p.95-99, 2009.

LAHM, A.; SUCK, D. DNase I-induced DNA conformation. 2 A structure of a DNase I-octamer complex. **J Mol Biol.**, v. 222, n. 3, p.645-67, 1991.

LEE, D.; POLISENSKY, D. H.; BRAAM, J. Genome-wide identification of touch- and darkness-regulated Arabidopsis genes: a focus on calmodulin-like and XTH genes. **New Phytol.**, v.165, p.429–444, 2005.

LEE, J.H.; SCHOFFL, F. An Hsp70 antisense gene affects the expression of HSP70/HSC70, the regulation of HSF, and the acquisition of thermotolerance in transgenic *Arabidopsis thaliana*. **Mol Gen Genet** 252:11–19, 1996.

MACÁRIO, A.J.L.; MACÁRIO, E.C. de. The Archaeal Molecular Chaperone Machine: Peculiarities and Paradoxes. **Genetics**, v.152: p.1277–1283, 1999.

MCCARTY, D.R.; CARSON, C.B.; STINARD, P.S.; ROBERTSON, D.S. Molecular analysis of viviparous-1: An abscisic acid-insensitive mutant of maize. **Plant Cell,** v.1, p.523–532, 1989.

MARTINS, L. **Estudo do controle genético do lignotuber em *Eucalyptus sp e* desenvolvimento de marcador molecular relacionado a esse caráter.** Botucatu, 2006. Dissertação (Mestrado em Ciências Biológicas – Área de Concentração: Genética) – Instituto de Biociências, Universidade Estadual Paulista.

MELLOR, H.; PARKER, P.J. The extended protein kinaseC superfamily. **Biochem.J.**, v.332,281-292, 1998.

MIERNYK, J.A. The J-domain proteins of *Arabidopsis thaliana*: an unexpectedly large and diverse family of chaperones. **Cell Estresse e Chaperones**, v.6, n.3, p 209-218, 2001.

MOL, C.D.; KUO, C.F.; THAYER, M.M.; CUNNINGHAM, R.P.; TAINER, J.A. Structure and function of the multifunctional DNA-repair enzyme exonuclease III. **Nature**, v.374, n.6520, p. 381-386, 1995.

MUKHOPADHYAY, I.; NAZIR, A.; SAXENA, D.K.; KAR CHWDHURI, D. Heat Shock Response: *hsp70* in Environmental Monitoring. **Journal of Biochemical and** *Molecular Toxicology*, v.17, p.249-253, 2003.

NISHIZUKA,Y. Studies and perspectives of protein kinase **C. Science**, v. 233, n.4761, p. 305-312 ,1986.

PRYOR L. Aspectos da cultura do eucalipto no Brasil. **IPEF**, n.2/3, p.53-59, 1971.

PRYOR, L. *Eucalyptus*. In: Halevy, A. **CRC Handbook of Flowering.** 1.ed. Boca Raton: Franklin Book Company, 1985. v.2, p. 476-482.

RIECHMANN, J.L.; HEARD, J.; MARTIN, G.; REUBER, L.; JIANG, C.; KEDDIE, J.; ADAM, L.; PINEDA, O.; RATCLIFFE, O.J.; SAMAHA, R.R.; *et al.* Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. **Science,** v.290, p.2105–2110. 2000.

SAMPSON, J.F. Multiple paternity of in E. rameliana (Myrtaceae). **Heredity**, v. 81, p. 349-355, 1998.

SOHN, K.H.; LEE, S.C.; JUNG, H.W.; HONG, J.K.; HWANG, B.K. Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance. **Plant Mol. Biol.**, v.61, p.897–915, 2006.

STEINBERG S.F. Structural basis of protein kinase C isoform function. **Physiol Rev.** v.88, n.4, p.1341-1378, 2008.

STEPHEN A.F.; L. MADDEN,T.L.; SCHÄFFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W; LIPMAN, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, **Nucleic Acids Res**., 25: 3389-3402, 1997.

TOLEMAN, M.A. SIMM, A.M.; MURPHY, T.A.; GALES, A.C.; BIEDENBACH, D.J.; JONES, R.N.; WALSH, T.R. Molecular characterization of SPM-1, a novel metallo-b-lactamase isolated in Latin America: report from the SENTRY Antimicrobial Surveillance Program. **J Antimicrob Chemother**., v.50, n.5, p. 673-679, 2002.

ULMASOV, T.; HAGEN, G.; GUILFOYLE, T.J. ARF1, a transcription factor that binds to auxin response elements. **Science,** v.276, p.1865–1868, 1997.

VICENTINI, R.; SASSAKI, F. T. ; GIMENES, M. A. ; MAIA, I. G.; MENOSSI, M. In silico evaluation of the Eucalyptus transcriptome.**Genetics and Molecular Biology**, v.28, n.3 (suppl), p.487-495, 2005.

WACHTER, A.; TUNC-OZDEMIR, M.; GROVE, B.C.; GREEN, P.J.; SHINTANI, D.K.; BREAKERA R.R. Riboswitch Control of Gene Expression in Plants by Splicing and Alternative 3' End Processing of mRNAs **ThePlantCell**,v.19, p.3437-3450, 2007.

WALSH, P.; BURSAC, D.; LAW, Y.C., CYR, D.; LITHGOW, T. The J protein family: modulating protein assembly, disassembly and translocation. **European Molecular Biology Organization**, v.5 n.6 p.567-571, 2004.

WALTNER, J.K.; PETERSON, F.C.; LYTLE, B.L.; VOLKMAN, B.F. The Protein Society Structure of the B3 domain from *Arabidopsis thaliana* protein At1g16640. **Protein Sci.** v.14, n.9, p.2478–2483, 2005.

WANG, W.; VINOCUR, B.; SHOSEYOV, O.; ALTMAN, A. Role of plant heat-shock proteins and molecular chaperones in the abiotic estresse response. **TRENDS in Plant Science**, v.9, n.5, p.244-253, 2004

WEISING, K.; NYBOM, H.; WOLFF, K. **DNA fingerprinting in plants: Principles, methods and applications**. 2. ed., Boca Raton: CRC Press, 2005, 472 p.

WHITTOCK, S. P.; APIOLAZA, L. A.; KELLY, C. M.; POTTS, B. M. Genetic control of coppice and lignotuber development in *Eucalyptus globulus.* **Australian Journal of Botany**, v.51, p. 57-67, 2003.

WOUTERS, M. **Fire in Mallee Communities.** Proceedings of the National Malleefowl Forum 70-82, 2004.

YOSHIMURA, K.; MASUDA, A.; KUWANO, M.; YOKOTA, A.; AKASHI, K. Programmed Proteome Response for Drought Avoidance/Tolerance in the Root of C_3 Xerophyte (Wild Watermelon) Under Water Deficits. **Plant and Cell Physiology**, v.49, n.2, p.226-241, 2008.

- Sites:

BRACELPA, Associação Brasileira de Celulose e Papel. Relatório de Dados do Setor. Disponível em: <http://www.bRACElpa.org.br> Acesso em: 11 mar. 2011.

SBS- Sociedade Brasileira de Silvicultura. Disponível em: <http://www.sbs.org.br/estatisticas.htm> Acesso em: 16 abr. 2009.

Eucalyptus grandis Genome Project 2010. Disponível em: <http://www.phytozome.net/eucalyptus.php>

http://softberry.com/

http://blast.ncbi.nlm.nih.gov/

http://pfam.sanger.ac.uk/