



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA E ENGENHARIA DE ALIMENTOS

**POLIMORFISMO E EXPRESSÃO DE GENES
DE CELULOSE SINTASE EM *Eucalyptus***

ELAINE LIMA TRIGUEIRO

Orientador:

Prof. Alexandre Siqueira Guedes Coelho

Goiânia - GO
Brasil

Agosto - 2007

ELAINE LIMA TRIGUEIRO

**POLIMORFISMO E EXPRESSÃO DE GENES
DE CELULOSE SINTASE DE *Eucalyptus***

Dissertação apresentada ao Programa de Pós-graduação em Agronomia, da Universidade Federal de Goiás, como requisito parcial à obtenção do título de Mestre em Agronomia, área de concentração: Genética e Melhoramento de Plantas.

Orientador:

**Prof. Dr. Alexandre Siqueira Guedes
Coelho**

Goiânia, GO - Brasil
2007

“De tudo, ficaram três coisas: A certeza de que estamos sempre começando... A certeza de que precisamos continuar... A certeza de que seremos interrompidos antes de terminar... Portanto, devemos: Fazer da interrupção, um caminho novo... Da queda, um passo de dança... Do medo, uma escada... Do sonho, uma ponte... Da procura, um encontro...”

Fernando Sabino
em *O Encontro Marcado*

SUMÁRIO

LISTA DE FIGURAS.....	VII
LISTA DE TABELAS.....	X
RESUMO.....	XII
ABSTRACT.....	XIV
1 INTRODUÇÃO.....	16
2 REVISÃO DE LITERATURA.....	18
2.1 O GÊNERO <i>Eucalyptus</i>	18
2.2 CARACTERIZAÇÃO DE ALGUMAS ESPÉCIES DE <i>Eucalyptus</i>	20
2.3 A EUCALIPTOCULTURA.....	22
2.4 A CELULOSE.....	24
2.4.1 Composição e estrutura.....	24
2.4.2 Parede celular.....	29
2.4.3 Genes de Celulose Sintase.....	30
2.4.4 O gene <i>EgCesA3</i> e seus ortólogos.....	32
2.5 DIVERSIDADE NUCLEOTÍDICA.....	34
2.6 POLIMORFISMOS DE UM ÚNICO NUCLEOTÍDEO (SNPs).....	36
2.7 DESEQUILÍBRIO DE LIGAÇÃO.....	41
2.7.1 Conceito.....	41
2.7.2 Forças evolutivas que afetam o DL.....	42
2.7.3 Como medir o DL.....	43
2.7.4 Estudos de DL em plantas.....	45
2.8 ESTUDOS DE EXPRESSÃO GÊNICA <i>IN SILICO</i>	46
3 MATERIAL E MÉTODOS.....	49
3.1 MATERIAL GENÉTICO PARA O ESTUDO DE DIVERSIDADE NUCLEOTÍDICA DO GENE <i>EgCesA2</i>	49
3.2 MATERIAL GENÉTICO PARA O ESTUDO DO DESEQUILÍBRIO DE LIGAÇÃO NO GENE <i>EgCesA3</i>	50
3.3 EXTRAÇÃO DO DNA GENÔMICO.....	50
3.4 QUANTIFICAÇÃO DO DNA.....	51
3.5 <i>PRIMERS</i> UTILIZADOS.....	51
3.6 DETECÇÃO DE POLIMORFISMO NO GENE <i>EgCesA2</i> DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DO GÊNERO <i>Eucalyptus</i>	52
3.7 TRIAGEM DE UMA BIBLIOTECA DE <i>BACs</i> EM BUSCA DE CLONES CONTENDO GENES <i>EgCesA2</i>	53
3.8 DETECÇÃO DE POLIMORFISMO NO GENE <i>EgCesA3</i> PARA O ESTUDO DO DESEQUILÍBRIO DE LIGAÇÃO EM <i>E. urophylla</i>	55
3.9 ANÁLISE DAS SEQÜÊNCIAS OBTIDAS E DETECÇÃO DO DESEQUILÍBRIO DE LIGAÇÃO.....	57

3.10 ANÁLISE DOS NÍVEIS DE EXPRESSÃO DOS GENES DE CELULOSE SINTASE EM <i>Eucalyptus</i>	57
4 RESULTADOS E DISCUSSÃO	60
4.1 AMPLIFICAÇÃO DOS FRAGMENTOS DO GENE <i>EgCesA2</i> EM DIFERENTES ESPÉCIES DE <i>Eucalyptus</i>	60
4.2 SEQÜENCIAMENTO PARCIAL DO GENE <i>EgCesA2</i> DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DO GÊNERO <i>Eucalyptus</i>	60
4.3 DETECÇÃO DE POLIMORFISMO NO GENE <i>EgCesA2</i> EM DIFERENTES ESPÉCIES DO GÊNERO <i>Eucalyptus</i>	63
4.4 ANÁLISE DE DESEQUILÍBRIO DE LIGAÇÃO NO GENE <i>EgCesA2</i>	67
4.5 TRIAGEM DA BIBLIOTECA DE BACs	68
4.6 AMPLIFICAÇÃO DE FRAGMENTOS DO GENE <i>EgCesA3</i> PARA DETECÇÃO DE DESEQUILÍBRIO DE LIGAÇÃO EM <i>E. urophylla</i>	69
4.7 SEQÜENCIAMENTO PARCIAL DO GENE <i>EgCesA3</i> DE CELULOSE SINTASE EM UMA POPULAÇÃO DE <i>E. urophylla</i>	70
4.8 ESTIMATIVA DO DESEQUILÍBRIO DE LIGAÇÃO DENTRO DO GENE <i>EgCesA3</i> EM UMA POPULAÇÃO DE <i>E. urophylla</i>	74
4.9 ANÁLISE DOS NÍVEIS DE EXPRESSÃO DOS GENES DE CELULOSE SINTASE DE <i>Eucalyptus</i>	79
5 CONCLUSÃO	88
6 REFERÊNCIAS	90
7 ANEXOS	100

AGRADECIMENTOS

Ao Prof. Dr. Alexandre Siqueira Guedes Coelho, pela orientação e confiança.

À Prof. Dra. Rosana Brondani pela colaboração e apoio na obtenção dos clones da biblioteca de BACs.

Aos colegas do laboratório de Genética e Genômica de Plantas, da Universidade Federal de Goiás, pela ajuda em vários momentos da composição deste trabalho, em especial à Ludmila e Marcela, pela colaboração e amizade.

Aos colegas do laboratório de Biologia Molecular, da Embrapa Arroz-Feijão, pela colaboração, em especial à Ana Rosa e Tereza.

Aos colegas de turma pela boa convivência e amizade.

Aos professores do Programa de Pós-Graduação em Agronomia, pelo aprimoramento de minha formação acadêmica.

Ao CNPq pela concessão da bolsa durante o primeiro ano de mestrado.

À minha família e ao Thiago, pelo apoio e carinho.

A todos que contribuíram para a realização deste trabalho.

LISTA DE FIGURAS

- Figura 1.** Uma compilação de hipóteses de como a celulose seria sintetizada. **A:** vista longitudinal da roseta composta de seis partículas, durante a alongação das microfibrilas de celulose. **B:** prováveis estruturas de uma partícula elementar da roseta. (I) a partícula contém seis polipeptídeos de alongação (CesAe), um polipeptídeo de iniciação (CesAi) e uma cópia de Korrigan celulase. (II) a partícula contém três cópias de dois tipos de polipeptídios que promovem o crescimento da cadeia e uma cópia de Korrigan celulase. (III) a partícula contém três cópias de dois tipos de polipeptídios que iniciam a síntese e promovem o crescimento da cadeia e uma cópia de Korrigan celulase. **C:** início da síntese de celulose. UDP-glicosil-transferase (UGT) transfere um resíduo de glicose para uma molécula de sitosterol, na face citoplasmática da membrana, formando sitosterol- β -glicosídeos (SG). A pequena cadeia de celulose é estendida com UDP-glicose pela subunidade CesAi, para formar SCD, com uma porção voltada para a face externa da membrana. A cadeia de celodextrina é então clivada pela Korrigan celulase, possibilitando a ligação com CesAe. A extensão da cadeia é providenciada pela sucrose sintase (Fonte: Read & Bacic, 2002).....27
- Figura 2.** Características estruturais de uma proteína de celulose sintase (adaptado de Ranik & Myburg, 2006).....28
- Figura 3.** Secção transversal de hastes de *Arabidopsis* coradas com azul de toluidina. (A) tipo selvagem; (B) mutante *irx3*; xe: xilema, ph: floema (Fonte: Taylor et al., 1999).....33
- Figura 4.** Descoberta de SNP em EST. (a) Uma seqüência na extremidade 3' de um EST ou um gene conhecido é usado para desenhar um par de iniciadores; (b,c) são usados iniciadores T3/T7 para facilitar o seqüenciamento direto. Depois da amplificação, os produtos da PCR são purificados; (d) as seqüências resultantes são alinhadas e são identificados os SNPs e polimorfismos de inserção e deleção (e) três haplótipos podem ser distinguidos (Fonte: Rafalski, 2002).....39
- Figura 5.** Polimorfismos de um único nucleotídeo foram identificados em 263 segmentos de um gene de ACP-esterol desaturase, em milho. As seqüências horizontais correspondem a 32 indivíduos seqüenciados. As colunas verticais identificam nove sítios polimórficos, incluindo um indel (I/D), que é um polimorfismo de inserção/deleção. São mostrados quatro haplótipos distintos. Estes quatro haplótipos podem ser identificados, sem ambigüidade, usando-se somente três SNPs (marcados com *) (Fonte: Rafalski, 2002).....41
- Figura 6.** Dois locos mostrando SNPs em oito indivíduos. A) DL completo entre os locos; B e D) Recombinação entre os sítios; C) freqüências de associações desiguais entre os locos, não implicando necessariamente em recombinação. C pode ser uma nova mutação associada a T (Fonte: Gaut & Long, 2003, com adaptações).....43

- Figura 7.** Formação dos pools e superpools para a triagem da biblioteca de BAC em busca de insertos que contenham os genes *CesA* ortólogos ao gene *EgCesA2* (Fonte: Vaz, 2005).....54
- Figura 8.** Confirmação da amplificação do fragmento do gene *CesA2* em cinco indivíduos de diferentes espécies de *Eucalyptus*, utilizando-se os iniciadores *CesA4.3* e *CesA4.4*.....60
- Figura 9.** Alinhamento obtido pelo programa EST2GENOME evidenciando a localização dos gaps e de nucleotídeos divergentes.....61
- Figura 10.** Alinhamento obtido pelo programa Clustal X, evidenciando a sequência da proteína de celulose sintase, com seus motivos conservados: dedo de zinco; Região classe específica I (CSR I); Região conservada I; Região classe específica II; Região conservada II; e domínios transmembrânicos (TM1-TM5).....62
- Figura 11.** Localização dos três SNPs encontrados pelo programa Staden Package na região flanqueada pelos primers *CesA4.4* e *CesA4.3*.....64
- Figura 12.** Eletroferogramas representativos da região que contém o SNP da posição 69 do *contig*. **A:** Relaciona-se às comparações entre a substituição de uma citosina (1) para uma timina(2). **B:** Trata-se da substituição de uma citosina (4) para uma adenina (5). É possível notar também que o SNP em destaque no quadro (5) corresponde a uma região heterozigota do gene. Os quadros (3) e (6) evidenciam o polimorfismo detectado pelo módulo *Trace difference* do programa Staden Package.....65
- Figura 13.** Eletroferogramas representativos da região que contém o SNP da posição 207 do *contig*.66
- Figura 14.** Eletroferogramas representativos da região que contém o SNP da posição 222 do *contig*.66
- Figura 15.** Confirmação da amplificação do fragmento do gene *EgCesA2* no *pool 9*. “+” indica a amplificação do fragmento no indivíduo de *E. grandis* (controle positivo).....68
- Figura 16.** Reação de amplificação do fragmento do gene *EgCesA2*, dentro do *pool 9*, revelando que a placa que contém o gene de interesse é a EUGR BC 002 087. “+” indica a amplificação do fragmento no indivíduo de *E. grandis* (controle positivo). L (marcador molecular ladder 1Kb Plus).....69
- Figura 17.** Reação de amplificação do fragmento do gene *EgCesA2* revelando que o clone E4 da placa EUGR BC 002 087 contém o gene de interesse. L (marcador molecular ladder 1Kb Plus).....69
- Figura 18.** Reação de amplificação do fragmento do gene *CesA3*, gerou de fragmentos de 800pb e 1,6kb, utilizando-se, respectivamente, os pares de *primers* *Euc2.3*

- (*forward*) e Euc1.4 (*reverse*), ou CesaA72.3 (*forward*) e Euc2.4 (*reverse*). L (marcador molecular ladder 1Kb Plus).....70
- Figura 19.** Representação do alinhamento feito pelo programa EST2GENOME entre o *contig* e a seqüência de *E. grandis* depositada no Genbank.....71
- Figura 20.** Posição dos SNPs dentro do *contig* formado por fragmentos de seqüências do gene *CesaA3*.....72
- Figura 21.** Eletroferograma mostrando o SNPs 371 do *contig*, dentro de uma região exônica do gene *CesaA3*.....72
- Figura 22.** Eletroferograma mostrando o SNPs 614 do *contig*, dentro de uma região exônica do gene *CesaA3*.....72
- Figura 23.** Diversidade nucleotídica em diferentes regiões do gene *CesaA3*.....74
- Figura 24.** Desequilíbrio de ligação (r^2) entre os marcadores SNPs.....77
- Figura 25.** Decaimento do desequilíbrio de ligação (r^2) em função da distância física entre os locos no gene *EgCesaA3*, em *E. urophylla*.....77
- Figura 26.** Comparação da porcentagem dos genes *CesaA* expressos no xilema de diferentes espécies de *Eucalyptus*.....80
- Figura 27.** Distribuição do número de ESTs dos genes *EgCesaA* observados em diferentes tecidos de *Eucalyptus*: XL (xilema); FL (floema); SE (plântulas); ML (folhas maduras); PU (folhas infectadas com *Puccinia*); TS (plantas submetidas a diversas condições de estresse).....82
- Figura 28.** Perfil de expressão dos genes de celulose sintase a partir de dados de *microarrays*. Cada linha representa um gene, e as colunas representam os diferentes tratamentos; GL= *E. globulus*; GR= *E. grandis*; XY= xilema; ML= folhas maduras. O nível de expressão gênica é positivamente correlacionado à intensidade de sinal, que é crescente a partir da fluorescência de cor verde, até uma fluorescência de cor vermelha.....83
- Figura 29.** Nível de expressão gênica de dados de *microarrays* a partir de mRNA de xilema e folhas de *E. grandis*.....84
- Figura 30.** Nível de expressão gênica de dados de *microarrays* a partir de mRNA de xilema de *E. globulus* e de *E. grandis*.....85
- Figura 31.** Nível de expressão gênica de dados de *microarrays* a partir de mRNA de xilema de *E. globulus* e de *E. grandis*.....86

LISTA DE TABELAS

Tabela 1. Relação dos indivíduos utilizados no estudo de diversidade nucleotídica do gene <i>EgCesA2</i>	49
Tabela 2. Genes <i>CesA</i> de <i>Arabidopsis thaliana</i> utilizados como referência.....	52
Tabela 3. Iniciadores desenhados para amplificação e/ou seqüenciamento dos genes de celulose sintase de <i>Eucalyptus</i>	52
Tabela 4. Genes <i>EgCesA</i> de <i>Eucalyptus</i> usados como referência para comparações com o banco de ESTs do Projeto Genolyptus.....	58
Tabela 5. Relação das bibliotecas de cDNA utilizadas.....	58
Tabela 6. Caracterização do polimorfismo detectado em 427pb do gene <i>EgCesA2</i> , em diferentes espécies de <i>Eucalyptus</i>	63
Tabela 7. Frequência dos haplótipos encontrados na população de <i>Eucalyptus</i>	67
Tabela 8. Significância do desequilíbrio ao nível de 0,05.....	67
Tabela 9. Valores do DL entre os SNPs 69 e 207 estimados por D, D' e r ²	68
Tabela 10. Resultado da triagem da biblioteca de BAC para o gene <i>EgCesA2</i>	68
Tabela 11. Trinca de nucleotídeos evidenciando os SNPs encontrados e os respectivos aminoácidos codificados.....	73
Tabela 12. Estimativa dos índices de diversidade a partir de 64 seqüências nucleotídicas com 770pb de comprimento, levando-se em conta a informação de vários sítios polimórficos em heterozigose.....	73
Tabela 13. Estimativa da associação entre pares de alelos (p-valor, nível de significância de 0,05).....	75
Tabela 14. Valores do DL entre pares de locos polimórficos, intervalo de confiança para D' e distância entre os locos.....	76
Tabela 15. Distribuição de <i>reads</i> com elevado grau de similaridade às seqüências codantes completas (cds) de genes <i>EgCesA</i> de <i>Eucalyptus</i>	79
Tabela 16. Significância das comparações entre os níveis expressão dos genes de celulose sintase de parede celular secundária nas diferentes espécies de <i>Eucalyptus</i>	80
Tabela 17. Significância das diferenças nos níveis de expressão dos genes <i>EgCesA</i> no xilema e em tecidos foliares de diferentes espécies de <i>Eucalyptus</i>	82

Tabela 18. Significância das diferenças nos níveis de expressão dos genes *EgCesA* no xilema e em tecidos foliares de *E. grandis* a partir de dados gerados com *microarrays*.....84

Tabela 19. Significância das diferenças nos níveis de expressão dos genes *EgCesA* no xilema de *E. grandis* e *E. globulus* a partir de dados gerados por *microarrays*.....85

RESUMO

TRIGUEIRO, E. L. **Polimorfismo e expressão de genes de celulose sintase em *Eucalyptus***. 2007. Dissertação (Mestrado em Agronomia: Genética e Melhoramento de Plantas)-Escola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás, Goiânia, 2007.¹

A celulose é um dos biopolímeros mais importantes do planeta, sendo também o mais abundante, e sem dúvida uma característica chave na história evolutiva das plantas. Contudo, sua biossíntese e regulação ainda não são bem compreendidas, embora avanços importantes tenham ocorrido nos últimos anos, sobretudo na identificação de genes e grupos de genes co-expressos para a formação de celulose na parede celular primária e secundária de vegetais. Além de sua relevância biológica, a celulose possui uma grande importância econômica no Brasil e no mundo, com destaque para a produção de celulose e papel a partir de *Eucalyptus*. Para garantir os elevados níveis de produtividade e competitividade no mercado internacional, grandes investimentos têm sido realizados pelo setor florestal, destacando-se o Projeto Genolyptus – Rede Brasileira de Pesquisa do Genoma de *Eucalyptus*, fruto de um esforço de empresas do setor de produção de celulose e papel, e de diversas instituições públicas de pesquisa, que têm, dentre outros objetivos, o intuito de identificar e caracterizar genes envolvidos na formação da madeira, para, no futuro, usar essa informação no melhoramento genético do *Eucalyptus*. Nesse contexto, este trabalho foi desenvolvido com o objetivo de realizar uma caracterização preliminar de um gene de celulose sintase em *Eucalyptus*, que está relacionado à síntese da parede celular secundária, sendo ortólogo ao gene *EgCesA2*, de *E. grandis*, bem como estudar o desequilíbrio de ligação em outro gene de celulose sintase, ortólogo ao gene *EgCesA3*, em uma amostra de uma população natural de *E. urophylla*. Em relação ao gene *CesA2*, foi seqüenciada uma região exônica do gene, formada por 427pb, a partir de amostras de DNA de 12 indivíduos de diferentes espécies e regiões geográficas. Procedeu-se a uma análise de detecção de polimorfismo, e estimou-se a ocorrência de três SNPs sinônimos ao longo do *contig*. Foi estimado um índice de diversidade $\pi = 0,00212$. Foi feita também uma triagem em uma biblioteca de BAC, gerada no âmbito do Projeto Genolyptus, e foi identificado o clone que contém o gene *CesA2*, o que permitirá o desenvolvimento futuro de caracterização minuciosa da estrutura deste gene em *Eucalyptus*. Em relação ao gene *CesA3*, a partir do seqüenciamento de 32 indivíduos, formou-se um *contig* de 770pb, e foi encontrado um índice de diversidade $\pi = 0,00185$. Foi possível a detecção de nove locos polimórficos distribuídos em regiões intrônicas, exônicas, e de 3'-UTR do gene. A análise de extensão do desequilíbrio de ligação dentro do gene *CesA3* sugere que os SNPs tendem a se encontrar em forte desequilíbrio a uma distância de aproximadamente 600pb. O conhecimento da posição dos SNPs nos genes *CesA2* e *CesA3* viabiliza a utilização destas marcas em futuros estudos de mapeamento genético. A existência de SNPs sinônimos nas regiões exônicas seqüenciadas corrobora com a idéia de que a celulose é um polímero muito importante à sobrevivência da planta, e, portanto, sua maquinaria de síntese apresenta características bastante conservadas, de modo que mutações em regiões efetivamente transcritas tenderiam a ser deletérias e não seriam fixadas. Além disso, foi feita uma análise da expressão gênica dos genes *CesA* em diferentes espécies de *Eucalyptus*, a partir de duas abordagens: a “*Digital Differential Display*”, a partir de diferente bibliotecas de ESTs e os *microarrays*, otimizados no âmbito do projeto Genolyptus. A análise com dados de *microarrays* revelou-se menos sensível na detecção da expressão diferencial, provavelmente devido às chamadas “relações cruzadas”.

Palavras-chave: celulose, SNP, *contig*, BAC, desequilíbrio de ligação.

¹Orientador: Prof. Dr. Alexandre Siqueira Guedes Coelho. EA-UFG.

ABSTRACT

TRIGUEIRO, E. L. **Polymorphism and expression of cellulose synthase genes in *Eucalyptus***. 2007. Dissertation (Master in Agronomy: Genetics and Plant Breeding)-Escola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás, Goiânia, 2007.¹

Cellulose is one of the most important and the most abundant biopolymer on the planet, playing a key role on the evolutionary history of plants. Important advances have been made in recent years, in particular on the identification of genes and co-expressed genes for the formation of cellulose in the primary and secondary cellular walls of plants. In addition to its biological relevance, cellulose has a great economic importance, not only in Brazil but in the world, especially due to the production of cellulose and paper from *Eucalyptus*. The high levels of production and competition in the international market are guaranteed by great investments, which are carried through by the forestal sector, in particular by the Genolyptus Project – Brazilian Network for Research on *Eucalyptus* Genome. This project is the result of a collective effort of companies involved on the production of cellulose and paper and various public research institutions. Their main goal is to identify and characterize genes involved in wood formation with the intent to genetically improve *Eucalyptus*. Based on this goal, this work was developed with two objectives. The first is doing a preliminary characterization of the cellulose synthase gene in *Eucalyptus*, which is associated with the synthesis of the secondary cellular wall and is orthologous to the gene *EgCeA2*, of *E. grandis*. The second objective is to study the linkage disequilibrium in another gene of cellulose synthase, orthologous to the *EgCesA3* gene, sampled from a wild population of *E. urophylla*. Regarding the *CesA2* gene, an exonic region with 427bp was sequenced from DNA samples of 12 individuals from different species and geographic regions. The next step was to proceed with an analysis to detect polymorphism which gave an estimate of three SNPs synonymous along the contig, with an estimated $\pi = 0.00212$ diversity index. A clone containing the *CesA2* gene was identified through a selection from a BAC library generated in the scope of the Genolyptus Project. This clone gives the prospect for the development of a minute characterization of this gene structure in *Eucalyptus*. Additionally, concerning the *CesA3* gene, the sequencing of 32 individuals allowed for the formation of a 770bp contig with a $\pi = 0.00185$ diversity index and detection of nine polymorphic loci distributed in intron and exon regions and at the 3'-UTR of the gene. The analysis of the extension of linkage disequilibrium in the *CesA3* gene suggests that SNPs tend to be in strong linkage disequilibrium at a distance of approximately 600bp. The knowledge of the position of the SNPs in the genes *CesA2* and *CesA3* makes possible the use of these markers in future studies of genetic mapping. The lack of non-synonymous SNPs in exon regions ensures that cellulose is in fact a very important polymer for plant survival. Hence its synthesis machinery presents highly conserved characteristics and so mutations in regions with effective transcription tend mostly to be deleterious and therefore would not be fixed. Moreover, the analysis of *CesA* gene expression in different species of *Eucalyptus*, was made from two boardings: "Digital Differential Display", from different libraries of ESTs and microarrays, optimized in the scope of the Genolyptus project. The analysis with data of microarrays showed less sensible in the detention of the distinguishing expression, probably had to the calls "crossed relations".

Key words: cellulose, SNP, contig, BAC, linkage disequilibrium.

¹Adviser: Prof. Dr. Alexandre Siqueira Guedes Coelho. EA-UFG.

1 INTRODUÇÃO

A celulose é um dos biopolímeros mais importantes do planeta, sendo também o mais abundante. O desenvolvimento de mecanismos que levaram à sua produção é considerado um marco de grande importância na história evolutiva das plantas. Apesar disso, a biossíntese e os processos que regulam a sua produção ainda não são bem compreendidos, embora avanços importantes tenham ocorrido nos últimos anos, sobretudo na identificação de genes e grupos de genes que são co-expressos para a formação de celulose da parede celular primária e secundária em vegetais superiores.

Além de sua relevância biológica, a celulose possui uma grande importância econômica no Brasil e no mundo. Em 2006, a produção brasileira de celulose e papel alcançou, respectivamente, cerca de 11,1 e 8,8 milhões de toneladas, gerando um *superavit* de US\$ 2,9 bilhões, em volume de exportações. Cabe ressaltar que 75% da área de plantio florestal destinado à produção de fibras de celulose e papel é composta por *Eucalyptus* (Bracelpa, 2006).

Para garantir esse patamar de produtividade, grandes investimentos têm sido feitos pelo setor florestal, destacando-se o Projeto Genolyptus – Rede Brasileira de Pesquisa do Genoma de *Eucalyptus*, que é fruto de um esforço coletivo de empresas do setor de produção de celulose e papel, e de diversas instituições de pesquisa, que têm, dentre outros objetivos, o intuito de identificar e caracterizar genes envolvidos na formação da madeira, para, no futuro, usar essa informação no efetivo melhoramento do *Eucalyptus*.

Nesse contexto, esse trabalho foi desenvolvido com o objetivo de realizar uma caracterização preliminar do grau de polimorfismo presente em dois genes de celulose sintase já descritos em *Eucalyptus*, que estão relacionados à síntese da parede celular secundária, denominados *EgCesA2* e *EgCesA3*. Em relação ao gene *EgCesA2*, foi feita uma análise de detecção de polimorfismo mais abrangente, usando-se para isso 12 indivíduos de diferentes espécies do gênero *Eucalyptus* provenientes de diferentes regiões geográficas. Para o gene *EgCesA3*, essa caracterização consistiu ainda, além da detecção de polimorfismo e da avaliação da magnitude dessa diversidade, de análises de desequilíbrio de ligação em uma amostra de uma população natural de *Eucalyptus*

urophylla, dando continuidade à pesquisa desenvolvida por Salazar (2006). Buscou-se ainda, através das estratégias de *Digital Differential Display* e pela análise específica dos resultados obtidos para os genes *EgCesA* em um *microarray* desenvolvido no âmbito do projeto Genolyptus, investigar o padrão de expressão dos seis genes de celulose sintase em diferentes espécies e tecidos de *Eucalyptus*.

2 REVISÃO DE LITERATURA

2.1 O GÊNERO *Eucalyptus*

O gênero *Eucalyptus* é originário da Oceania, com maior riqueza de espécies na Austrália, embora existam representantes em algumas ilhas ao norte, como *Eucalyptus urophylla*, que é endêmico do Timor e de outras pequenas ilhas da Indonésia, e *E. deglupta*, que é encontrado nas ilhas de Sulawesi, Ceram e ao norte de Papua Nova Guiné.

O gênero *Eucalyptus* evoluiu em diferentes condições ambientais, desde pântanos até solos arenosos, em ambientes com características variáveis em relação à precipitação e temperatura. Fósseis dos primeiros *Eucalyptus* foram descobertos no sudeste de Queensland e datam do Paleoceno ou Eoceno. Durante o Paleoceno, 65 milhões de anos atrás, a costa sul da Austrália estava a 60° de latitude Sul, ainda conectada a Antártida. O continente lentamente moveu-se para o norte para uma posição de 35° S. Na Austrália, uma barreira se formou, a planície Nullarbor, isolando a espécie do sudoeste e do sudeste pouco antes do Mioceno. O gênero se adaptou a cada região, sendo que as espécies da porção oriental se adaptaram aos verões úmidos, com solos pobres em nutrientes comparados aos do ocidente, e as espécies ocidentais se adaptaram aos invernos chuvosos (Williams & Woinarski, 1997).

O *Eucalyptus* pertence à família *Myrtaceae* e possui mais de 600 denominações diferentes, incluindo espécies, variedades e híbridos (Pryor, 1976). A copa de um *Eucalyptus* geralmente é rala e alongada, e conforme a espécie, o tronco apresenta-se retilíneo e cilíndrico, com casca muito lisa ou muito áspera. As folhas são geralmente coriáceas, cutinizadas e possuem a característica, comum entre a maioria das espécies, de serem aromáticas e dispostas nos ramos de maneira oposta na parte inferior e alternada na parte superior. Os tipos de *Eucalyptus* mais plantados no país são os híbridos, como o *E. grandis* x *E. urophylla* (com aproximadamente 47% da área destinada à indústria de celulose), *E. grandis* (26%), *E. saligna* (4%), *E. dunnii* (3%), (Bracelipa, 2004).

Dentro da família *Myrtaceae*, o gênero *Eucalyptus* tem uma produção freqüente de grande número de flores e frutos, com óvulos e sementes numerosas. A inflorescência é uma panícula ou umbela, com botões florais sem cálice ou corola

reconhecíveis, havendo somente uma estrutura semelhante a uma tampa (opérculo) que recobre totalmente as peças reprodutivas (estames e pistilo). As flores de *Eucalyptus*, sempre pequenas e geralmente brancas, podem também ser amarelas, cor-de-rosa ou, como em *E. ficifolia*, de um vermelho intenso. Os frutos são do tipo cápsula deiscente por três a seis valvas apicais, denominados diplotégia. Todas as espécies têm flores hermafroditas, com múltiplos estames vistosos, que atraem os insetos, sobretudo Hymenópteros, Dípteros, Lepidópteros, Coleópteros e Hemípteros. Na Austrália, pequenos marsupiais e alguns pássaros também figuram como polinizadores (House, 1997).

O *Eucalyptus* possui um conjunto haplóide de 11 cromossomos e um tamanho de genoma estimado em aproximadamente 600 milhões de pares de bases (Mb) (Grattapaglia & Bradshaw Jr., 1994), que permitiu uma grande flexibilidade evolutiva (House, 1997). Segundo Pryor (1976), o *Eucalyptus* desenvolveu mecanismos para evitar a autopolinização, como a protandria, que varia em intensidade, dependendo da espécie e grupos de espécie. Além disso, uma importante barreira de isolamento pré-zigótico entre espécies diferentes de *Eucalyptus* é a incompatibilidade estrutural e fisiológica do grão de pólen para o crescimento do tubo polínico. Assim, Pryor define o gênero, além de preferencialmente alógamo (podendo ocorrer até 30% de autogamia), como tendo suas populações compostas por indivíduos heterozigotos e que apresentam uma depressão geral no vigor com a autofecundação.

Várias pesquisas indicam também a existência de mecanismos de isolamento pós-zigóticos em cruzamentos interespecíficos do gênero *Eucalyptus*. Myburg et al. (2004) investigaram barreiras à introgressão interespecífica em cruzamentos induzidos entre *E. grandis* e *E. globulus*, a partir de híbridos F₁. Neste estudo, duas famílias de retrocruzamento foram usadas e um total de 1.354 locos de marcadores AFLP foram avaliados. Os resultados indicaram que a habilidade de fertilização dos gametas da F₁ e a viabilidade dos zigotos da F₂ foram muito afetadas pela incompatibilidade genética entre os dois genomas.

Já a poliploidia em *Eucalyptus* foi conseguida artificialmente pela primeira vez por Kappor & Sharma (1985). Através de indução por colchicina, produziram tetraplóides de *E. tereticornis* x *E. grandis* (4n=44) que apresentaram um crescimento mais acelerado em relação aos híbridos diplóides. O *Eucalyptus* possui cromossomos extremamente pequenos (2 a 6µm) o que dificulta sua observação, não existindo registros de poliploidia natural no gênero (Poots & Wiltshire, 1997).

2.2 CARACTERIZAÇÃO DE ALGUMAS ESPÉCIES DE *Eucalyptus*

Ao se analisar o desenvolvimento e as características tecnológicas da madeira de um grupo de espécies de *Eucalyptus* é fácil perceber diferenças entre elas. Devida a grande variabilidade do gênero, a escolha da melhor espécie a ser plantada deve levar em consideração o objetivo da produção (uso da madeira) e as condições edafoclimáticas da região.

Uma espécie muito utilizada no Brasil é *Eucalyptus urophylla* S. T. Blake, algumas vezes referida erroneamente como *Eucalyptus alba*, uma espécie muito semelhante. É endêmico do sudoeste da Indonésia, disperso nas ilhas Timor, Flores, Adonara, Lomblem, Plantar, Alor e Wetar, situadas ao norte da Austrália, entre 7° e 10° de latitude Sul. É encontrado naturalmente em regiões de até 3000m de altitude e em áreas com precipitação média anual de 2500 a 7000mm e temperatura média variando de 24 a 28°C. É tolerante a solos pobres desde que permaneçam relativamente úmidos durante a estação seca (Eldridge, 1994).

No Brasil, *E. urophylla* apresenta grande estabilidade genética em relação a todas as áreas em que foi testada (Kise, 1977; Moura et al., 1980; Pires & Parente, 1986; Drumond et al., 1998; citados por Moura, 2004), o que contribuiu para que a espécie fosse plantada em mais de 500 mil hectares.

Outra espécie utilizada em programas de melhoramento é *E. pellita* F. Muell, com tamanho médio superando os 40m de altura e 1m de DAP (densidade medida à altura do peito). Em solos pobres alcança de 15 a 20m e em terrenos arenosos, ao nível do mar, a altura é de 5 a 10m. A temperatura máxima do mês mais quente está entre 24 e 30°C e a mínima do mês mais frio está entre 4 e 19°C (IPEF – chave de identificação, 2006). A partir de testes realizados é possível considerar a espécie como altamente potencial para as regiões onde não ocorram geadas severas.

Uma das maiores áreas de plantio no país é composta pela espécie *Eucalyptus grandis* Hill ex Maiden, que apresenta árvores muito altas (40 a 45m), com um tronco de até 2m de diâmetro e casca de textura lisa (IPEF- chave de identificação, 2006). É uma espécie nativa do norte de Nova Gales do Sul e da costa sul de Queensland, na Austrália. Está distribuída entre as latitudes 26 a 32° Sul e altitudes de até 300m, em regiões com precipitações entre 1000 e 1750mm, com temperaturas médias entre 29 e 32°C (Eldridge,

1994). Possui uma madeira medianamente leve, sendo considerada de baixa estabilidade dimensional e elevada permeabilidade (Alzate, 2004).

Já *Eucalyptus globulus* Labill é encontrada naturalmente em regiões de clima temperado na Tasmânia e no extremo sul do continente australiano, apresentando um porte médio de 15 a 25m (Eldridge, 1994). *E. globulus* e *E. grandis* têm características contrastantes em relação à taxa de crescimento e qualidade da madeira. *E. globulus* apresenta alta densidade da madeira e lento crescimento volumétrico, enquanto *E. grandis* é superior no crescimento e apresenta desempenho inferior na medida da densidade volumétrica (Kirst, 2005).

A densidade básica é uma das propriedades físicas mais importantes na caracterização tecnológica da madeira para diversos fins, pois sua variação afeta a resistência mecânica e a variação dimensional da madeira. Em termos químicos, reflete a porcentagem de celulose, hemicelulose e lignina. Como regra geral, quanto maior a densidade básica, maiores serão os valores das propriedades mecânicas de resistência da madeira.

Segundo Foelkel (1992), a densidade básica para madeiras de *Eucalyptus*, pode variar de 0,300 a 0,800g/cm³. Especificamente para a produção de polpa celulósica é recomendado o uso de madeiras com densidade básica de 0,450 a 0,550g/cm³. Matérias-primas com densidade básicas acima de 0,650g/cm³ dificultam o processo de picagem, provocando aumento do consumo de reagentes químicos e conseqüente elevação no teor de rejeitos.

Em relação à qualidade da madeira, *E. urophylla* apresenta uma densidade básica considerada alta na avaliação feita por Scanava Júnior (2001). Já o híbrido *E. urophylla* x *E. grandis* apresentou valores médios de densidade básica da madeira superiores em relação a *E. grandis* e *E. saligna* (Alzate, 2004). Numa avaliação feita por Duarte (2007), em termos de qualidade da madeira e desempenho frente ao processo *kraft* de polpação, *E. globulus* apresentou características superiores, com uma densidade básica de 0,590g/cm³ e um teor de holocelulose de 75,56%. Já o híbrido *E. grandis* x *E. urophylla* apresentou valores de 0,442g/cm³ e 65,96%, respectivamente.

Em um estudo feito por Ribeiro & Zani Filho (1993), em árvores com 80 meses de idade, provenientes da mesma região, a densidade básica da madeira de populações de *E. pellita* apresentou maior valor (0,611g/cm³), seguido por *E. urophylla* (0,559g/cm³) e *E. grandis* (0,479g/cm³). Nesse estudo, as procedências de *E. pellita* e *E. urophylla* que foram

avaliadas estão fora da faixa ótima de utilização para a produção de celulose sugerida pelo autor. Cabe ressaltar que não é simples determinar qual a melhor espécie para a indústria de papel e celulose. A avaliação deve relacionar a área a ser plantada, com as características da madeira, juntamente com o processo industrial a ser utilizado, bem como a qualidade do papel a ser obtido.

2.3 A EUCALIPTOCULTURA

O *Eucalyptus* é um gênero de grande importância econômica, sendo muito utilizado na indústria de produção de papel e celulose, devido à qualidade, crescimento e rendimento de suas fibras, como combustível (carvão vegetal), além de a madeira ser empregada na fabricação de móveis. Também é utilizado na produção de alimentos, fármacos, materiais de limpeza, óleos e na composição de tecidos sintéticos. Possui um crescimento rápido e boa adaptabilidade a diversas áreas geográficas. No Brasil, 75% da área de plantio florestal é composta por *Eucalyptus* (Bracelipa, 2006).

As plantações florestais de *Eucalyptus* podem aliviar a pressão sobre as áreas de vegetação natural nos mais diversos ecossistemas, suprimindo a demanda por madeira, devido ao seu ciclo médio de corte de cerca de sete anos. Além disso, podem contribuir significativamente aos programas de recuperação de áreas degradadas, fornecendo proteção adequada na estabilização de dunas e no controle de erosões, devido ao formato de suas raízes (Lima, 1998). O gênero possui um sistema radicular formado por uma raiz central e uma extensa rede de raízes secundárias, que produzem efeitos benéficos sobre o solo, deixando-o mais estruturado, aumentando a capacidade de armazenamento de água, de drenagem e de aeração. Outro benefício ambiental proporcionado pelo gênero se refere à sua excelente capacidade de fixação de carbono. Árvores de *Eucalyptus* fixam carbono a uma taxa média de até 10,17 milhões de toneladas/ano em plantações localizadas no Brasil (Marcolin et al., 2002).

Os primeiros exemplares de *Eucalyptus* foram introduzidos no país no ano de 1825 no Jardim Botânico do Rio de Janeiro, como planta ornamental. As primeiras florestas foram plantadas por volta do ano de 1868 no Rio Grande do Sul, para prover lenha. Contudo, a expansão em larga escala dessa cultura é devida ao trabalho de Navarro de Andrade, em 1903, na Companhia Paulista de Estrada de Ferro, com o uso da madeira de *Eucalyptus* para a fabricação de dormentes e como combustível de locomotivas. Em

1930, o *Eucalyptus* já era cultivado para fornecer combustível às indústrias e como lenha para os fogões domésticos, já que o fogão a gás só se popularizaria após o fim da II Guerra Mundial. Em 1966, o governo federal estabeleceu um programa de incentivos fiscais, a fim de atender a crescente demanda por madeira de indústrias das regiões sul e sudeste, que nesta época já estavam praticamente destituídas de cobertura vegetal nativa. Em 1967, a área plantada com *Eucalyptus* era de 14.000 hectares, em 1973, segundo a FAO, já era de aproximadamente um milhão de hectares. E em 2000 a área plantada atingiu 2.965.880 hectares (SBS-Sociedade Brasileira de Silvicultura, 2001).

As indústrias que produzem papel e celulose a partir de fibra de *Eucalyptus* compõem um dos setores de melhor desempenho brasileiro em volume de exportação (chegou à cifra de US\$ 7,5 bilhões, em 2005), sendo que essas investem valores consideráveis em pesquisa para manter um alto patamar de produtividade. O setor também contribui socialmente ao crescimento do país ao gerar 45 mil empregos diretos, e só no ano de 2005 foram criados mais de 110 mil novos postos de trabalho, entre mão-de-obra direta e indireta (Bracelpa, 2005).

Além disso, é crescente a pressão, inclusive internacional, para que os interesses econômicos sejam alinhados a sustentabilidade ambiental, e dessa forma, estudos que visem o aumento da produtividade em qualidade, ao invés do aumento da quantidade de área plantada são cada vez mais necessários. Nos últimos 30 anos, as empresas e os centros de pesquisa acumularam um amplo conhecimento científico sobre as variadas espécies de *Eucalyptus*, promoveram sua melhoria, aperfeiçoaram técnicas de produção de mudas, de preparação e conservação dos solos, de plantio e monitoramento das florestas plantadas, além de técnicas e práticas conservacionistas para preservar e recuperar áreas degradadas de mananciais e de vegetação nativa. Alguns dos objetivos das pesquisas patrocinadas pelas indústrias produtoras de papel, em conjunto com recursos governamentais, são aumentar a produção de celulose, e diminuir o teor de lignina no *Eucalyptus* para que o processo de extração da fibra utilize menos solvente, diminuindo custos e danos ambientais.

Sendo assim, para se otimizar lucratividade e competitividade, a partir da década de 70, diversos projetos de melhoramento da espécie foram desenvolvidos no país, obtendo-se clones elites com atributos importantes, sobretudo às indústrias de celulose e papel. Desde então, estudos de variabilidade e distância genética, análises de identidade, certificação e proteção de clones, transgenia (González, 2002), e em escala experimental

seleção assistida por marcadores moleculares, têm sido feitos para aumentar a produtividade das florestas. Além disso, estudos envolvendo a caracterização de seqüências de genes envolvidos na formação da madeira também foram realizados, como genes envolvidos na síntese de lignina (Santos, 2005; Póvoa, 2005) e celulose (Salazar, 2006).

2.4 A CELULOSE

2.4.1 Composição e estrutura

A celulose é um carboidrato sempre referido como o polímero mais abundante da Terra, sendo produzido principalmente pelas plantas vasculares. Entretanto sua síntese também é encontrada em muitos grupos de algas, bactérias, alguns protozoários, fungos e no reino animal, em tunicados. A celulose é um polissacarídeo extracelular essencial às plantas vasculares, sendo um componente importante da parede celular, promovendo a resistência necessária à pressão de turgor. A orientação da deposição das microfibrilas de celulose também determina a direção da expansão celular. Se uma parede é reforçada por microfibrilas distribuídas de modo aleatório, o crescimento ocorre em todas as direções, e a célula adquire um formato esférico. Quando as microfibrilas são depositadas de modo a ter uma mesma orientação, a célula se expande em ângulo reto em relação a essas, de modo que, se a orientação das microfibrilas for transversal, o crescimento celular será longitudinal, e assim, a célula possuirá uma forma mais retangular e alongada (Taiz & Zaiger, 1998).

Além disso, a celulose parece ser fundamental para as plantas como uma resposta a condições ambientais adversas. Chen et al. (2005), por exemplo, caracterizando dois alelos mutantes em *Arabidopsis*, *lew-2-1* e *lew-2-2*, sugerem que o controle do mecanismo de biossíntese de celulose representa uma das principais adaptações das plantas a condições de estresse hídrico.

Paradoxalmente, se conhece pouco a respeito do mecanismo de síntese desse carboidrato de tão grande importância biológica e econômica. Sem dúvida houve avanços, sobretudo nos últimos anos, mas ainda existem muitas dúvidas sobre como o processo de síntese é controlado.

A celulose é formada por cadeias lineares de resíduos de glicose na conformação β -1,4, que se dispõem em microfibrilas. Essas microfibrilas se ligam umas as

outras através de pontes de hidrogênio. Devido à configuração espacial alternativa das ligações glicosídicas, que une os resíduos de glicose adjacentes, as unidades repetidas de celulose formam um dissacarídeo denominado celobiose. A estrutura molecular precisa das microfibrilas de celulose é ainda incerta. Segundo Kimura et al. (1999), alguns modelos de organização sugerem que a celulose possa existir em estado cristalino e não cristalino. O cristal de celulose pode se apresentar na natureza em duas aloformas, celulose I e celulose II, ressaltando-se que a celulose I é predominante, pois nenhuma célula eucariótica sintetiza de forma abundante a celulose II *in vivo*.

A extensão das ligações covalentes entre os resíduos de glicose adjacentes proporciona às microfibrilas de celulose propriedades notáveis, como uma grande força tensora, equivalente ao aço. A celulose é também quimicamente estável, e relativamente imune ao ataque químico e enzimático. Essas propriedades fazem da celulose um material excelente para a construção de uma estrutura rígida como a parede celular (Taiz & Zeiger, 2002).

Algumas pesquisas propõem que a glicose usada na síntese de celulose possa ser obtida através da sacarose (um dissacarídeo composto por frutose e glicose). De acordo com Amor et al. (1995), a enzima Sacarose Sintase atua como um canal metabólico para transferir glicose da sacarose, via UDP-glicose, para o crescimento da cadeia de celulose. Segundo Ohmiya et al. (2000), a sacarose, além de servir como substrato para a síntese de celulose, também induz a expressão de duas celulases (β -1,4-endoglucanases) em *Populus*; ou seja, a sacarose promoveria a síntese de celulose bem como a sua degradação.

Mueller & Brown (1980), através de observações do meristema de raízes de *Zea mays* por microscopia eletrônica, sugeriram que microfibrilas de celulose são sintetizadas por um complexo ordenado, conhecido como roseta ou complexo terminal (TC), localizado na membrana plasmática. Acredita-se que essa estrutura contenha muitas unidades de celulose sintase, enzimas que sintetizam D-glicose. Em plantas, o complexo terminal possui morfologia bastante conservada, com as subunidades arranjadas em seis dobras simétricas de onde emergem cristais de celulose I. Contudo, em muitos grupos de algas, essa estrutura apresenta-se variando desde uma forma linear até uma forma composta de um octâmero de subunidades. Um dos aspectos mais interessantes da biossíntese de celulose é justamente essa grande diversidade de estruturas em que o complexo terminal se apresenta (Brown & Saxena, 2000).

Desde a descoberta do complexo terminal em plantas vasculares, dados de estudos bioquímicos e ultraestruturais promoveram informações para que o modelo em roseta pudesse ser revisado. Saxena & Brown (2005) propõem dois níveis de agregação das celulosas sintases. Primeiramente, a junção e o processamento de três diferentes domínios protéicos (cada um composto de uma única celulose sintase), ocorreria para formar um arranjo linear de seis partículas. Posteriormente, a conformação linear daria lugar à disposição em roseta com seis dobras simétricas. Os mecanismos de junção e processamento para formação do complexo terminal provavelmente ocorrem no retículo endoplasmático e no aparelho de Golgi. A roseta é então transportada para a membrana plasmática para ativar a síntese de microfibrilas de celulose. Embora ainda não completamente entendido este modelo é muito atraente à hipótese de que o complexo terminal seja formado por várias subunidades protéicas codificadas por muitos genes de celulose sintase, denominados de genes *CesA*. Na ausência de um produto específico de um gene *CesA*, a formação da roseta pode ser afetada, como se confirma em muitas análises feitas com mutantes.

Além disso, outros estudos metabólicos, revisados por Read & Bacic (2002), propõem que a síntese de celulose possa ser iniciada a partir da transferência de um resíduo de glicose de um substrato solúvel de UDP-glicose, para uma molécula de sitosterol, formando sitosterol- β -glicosídeo (SG), na face interna da membrana plasmática (Figura 1). A molécula de SG pode atuar iniciando a polimerização de cadeias de glicose, catalisadas por proteínas CESA do complexo de celulose sintase. O resultado é a formação de lipídeos ligados a polissacarídeos denominados sitosterol celodextrinas (SCDs). A formação de SG e SCDs pode ser bloqueada pela ação de herbicidas, como o 2,6-diclorobenzonitrilo (DCB). As celodextrinas podem ser clivadas do sitosterol pela ação de da β -1,4-endo-glucanase (celulase de Korrigan), em que o sítio ativo é possivelmente localizado na face externa da membrana plasmática (Nicol, et al., 1998).

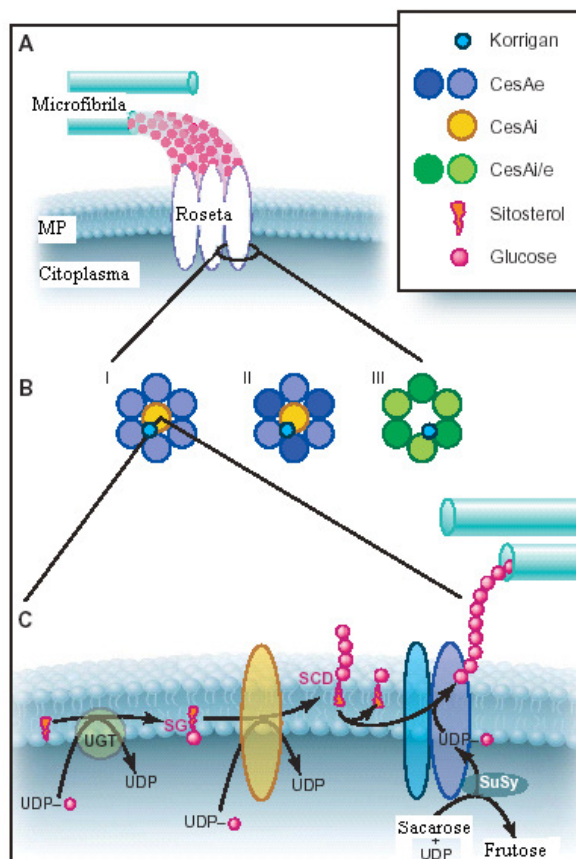


Figura 1. Uma compilação de hipóteses de como a celulose seria sintetizada. **A:** vista longitudinal da roseta composta de seis partículas, durante a elongação das microfibrilas de celulose. **B:** prováveis estruturas de uma partícula elementar da roseta. (I) a partícula contém seis polipeptídeos de elongação (CesAe), um polipeptídeo de iniciação (CesAi) e uma cópia de Korrgan celulase. (II) a partícula contém três cópias de dois tipos de polipeptídios que promovem o crescimento da cadeia e uma cópia de Korrgan celulase. (III) a partícula contém três cópias de dois tipos de polipeptídios que iniciam a síntese e promovem o crescimento da cadeia e uma cópia de Korrgan celulase. **C:** início da síntese de celulose. UDP-glicosil-transferase (UGT) transfere um resíduo de glicose para uma molécula de sitosterol, na face citoplasmática da membrana, formando sitosterol- β -glicosídeos (SG). A pequena cadeia de celulose é estendida com UDP-glicose pela subunidade CesAi, para formar SCD, com uma porção voltada para a face externa da membrana. A cadeia de celodextrina é então clivada pela Korrgan celulase, possibilitando a ligação com CesAe. A extensão da cadeia é providenciada pela sucrose sintase (Fonte: Read & Bacic, 2002).

Pouco se sabe a respeito de como o processo de cristalização da celulose é controlado, exceto que a direção da deposição das microfibrilas pode ser guiada pelos microtúbulos adjacentes à membrana celular. Paredes et al. (2006) documentaram o processo de deposição da celulose em células vivas de espécimes de *Arabidopsis* transgênicas, através da observação de fusos de fibras de celulose sintase fluorescentes. Utilizando microscopia confocal, os autores observaram que os complexos de *CesA* na

membrana plasmática moveram-se a taxas constantes em trilhas lineares, coincidentes com a orientação dos microtúbulos corticais. Dentro de cada trilha linear observada, o movimento do complexo era bidirecional. A inibição da polimerização dos microtúbulos mudou a distribuição e o padrão de movimentação dos complexos terminais na membrana, indicando um mecanismo relativamente direto para a orientação da deposição da celulose pelo citoesqueleto celular.

Todas as proteínas de celulose sintase que foram descritas possuem características estruturais conservadas (Figura 2). Em *Eucalyptus*, segundo Ranik & Myburg (2006), na extremidade amino terminal da estrutura primária destas proteínas, há um motivo similar a um domínio presente em fatores de transcrição, conhecido como “dedo de zinco”. Trata-se de uma seqüência contendo oito resíduos “CxxC” que fornecem os sítios de ligação para dois íons Zn^{2+} . Acredita-se que esse domínio tenha um papel importante nas interações proteína-proteína do complexo terminal. Na seqüência da proteína CESA, após este domínio, há duas regiões que mostram conservação limitada entre diferentes membros da família CESA de uma mesma espécie, embora sejam muito similares entre CESA ortólogas encontradas em espécies diferentes. Estas regiões são denominadas de regiões classe específica I (CSRI) e II (CSRII). Essas regiões são intercaladas por cerca de 300 a 400 resíduos de aminoácidos conservados, que apresentam dois domínios transmembrânicos (TM1-TM2), e um motivo formado por dois resíduos espaçados de ácido aspártico (D). Abaixo da CSRII, existe ainda um outro fragmento protéico conservado composto por um terceiro resíduo de ácido aspártico e o motivo QxxRW, que fariam parte do sítio catalítico, além de seis domínios transmembrânicos adicionais (TM3-TM8), já na extremidade carboxi-terminal da proteína.

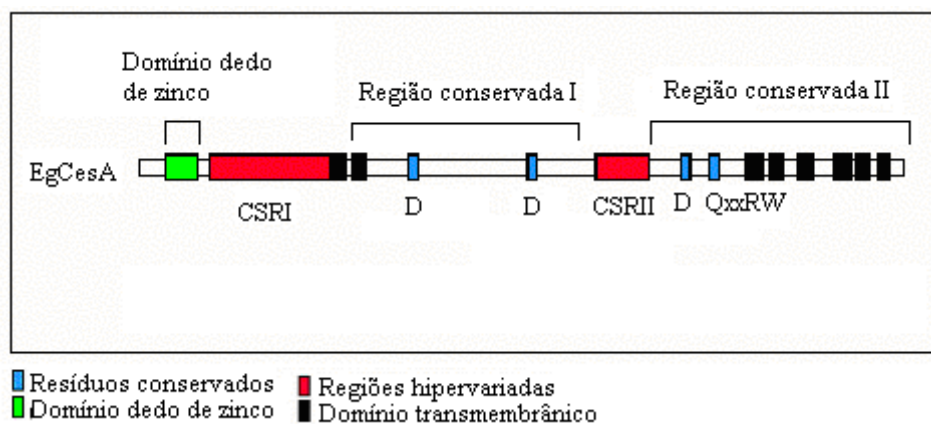


Figura 2. Características estruturais de uma proteína de celulose sintase (adaptado de Ranik & Myburg, 2006).

2.4.2 Parede Celular

Quando uma microfibrila de celulose é sintetizada, também é sintetizada uma parede que contém uma alta concentração de outros polissacarídeos que são capazes de interagir entre si e, provavelmente, modificar o crescimento dessa microfibrila. A composição dessa parede varia conforme o tipo celular que é por ela envolvido. Essas variações são importantíssimas para a organização dos diferentes tecidos vegetais. Paredes celulares como as do parênquima cortical, que são geralmente delgadas e têm poucas características distinguíveis, são conhecidas como paredes primárias e se depositam em células ainda em crescimento. Em contraste, paredes celulares de algumas células especializadas, como as células da epiderme, do colênquima, de fibras do floema, de elementos traqueais do xilema e de outras formas de esclerênquima, são densas e com múltiplas camadas, sendo denominadas de paredes secundárias, depositadas depois que o crescimento celular cessa (Raven et al., 2002).

As diferenças em relação às propriedades físicas das paredes celulares primárias e secundárias são devidas a variações no número de cadeias individuais de celulose na composição das microfibrilas. As microfibrilas de paredes primárias possuem aproximadamente 36 cadeias individuais de celulose (Delmer & Amor, 1995). Entretanto, as paredes secundárias de algumas algas contêm mais de 12.000 resíduos de celulose por microfibrila (Brown et al., 1996).

Nas paredes celulares primárias, as microfibrilas de celulose são embebidas em uma matriz amorfa, altamente hidratada. Essa matriz consiste de dois grupos de polissacarídeos, chamados de hemicelulose e pectinas, além de algumas proteínas estruturais. As hemiceluloses são polissacarídeos flexíveis que se ligam à superfície das microfibrilas de celulose. Vários tipos de hemiceluloses são encontrados em paredes de células vegetais, variando em composição conforme diferentes espécies e tecidos. O tipo mais abundante de hemicelulose em dicotiledôneas é o xiloglucano, porém, dependendo do estágio de desenvolvimento e da espécie, a fração de hemicelulose da parede também contém outros polissacarídeos, como xilanos e glucomanos. Já as pectinas formam uma fase gelatinosa em que a rede celulose-hemicelulose é envolvida e atuam como um suporte hidrofílico, determinando a porosidade da parede às macromoléculas. O papel das proteínas estruturais ainda é incerto, porém sabe-se que elas promovem uma força mecânica à parede (Taiz & Zeiger, 2002).

Depois do término do alongamento celular, algumas células continuam a sintetizar uma parede, a chamada parede secundária. As paredes secundárias são sempre densas, como em traqueídes, fibras e em outras células que formam um suporte mecânico às plantas. Esse tipo de parede possui múltiplas camadas que diferem em estrutura e composição das paredes primárias. Além disso, as paredes secundárias são sempre impregnadas de lignina, um composto fenólico, ligado a subunidades de álcoois aromáticos (coumaril, coniferil e sinapil), o que aumenta significativamente a sua densidade e resistência e reduz sua susceptibilidade ao ataque de patógenos (Taiz & Zeiger, 2002).

2.4.3 Genes de Celulose Sintase

Em 1983, num estudo desenvolvido por Aloni e colaboradores, ocorreu um avanço importante para o entendimento do processo celular de síntese de celulose. Eles obtiveram a síntese de celulose *in vitro* com a solubilização de proteínas de membrana da bactéria Gram-positiva *Acetobacter xylinum*. Já na década de 90, dois esforços independentes conduziram à identificação dos genes envolvidos diretamente na síntese da celulose. Wong et al. (1990) isolaram um operon, codificado por quatro proteínas, requerido para a biossíntese de celulose em *A. xylinum*. Saxena et al. (1990) isolaram um gene que codifica uma subunidade catalítica de celulose sintase, também em *A. xylinum*, através de microseqüenciamento de polipeptídeos purificados, que foram identificados por foto-afinidade com análogos de UDP-glicose. Desde então, genes de celulose sintase foram caracterizados em outras bactérias, como em *Agrobacterium tumefaciens* (Matthysse et al., 1995) e em plantas, como em *Gossypium hirsutum* (Pear et al., 1996), *Arabidopsis thaliana* (Arioli et al., 1998), *Populus tremuloides* (Wu et al., 2000; Samuga & Joshi, 2002), arroz (Tanaka et al., 2003), cevada (Burton et al., 2004), *Pinus taeda* (Nairn & Haselkorn, 2005) e *Eucalyptus grandis* (Ranik & Myburg, 2006).

A análise de seqüências do genoma de *Arabidopsis*, inteiramente seqüenciado, revelou que existem 10 genes *CesA* nesta espécie (nomeados *AtCesA1-AtCesA10*), considerados genes de celulose sintase verdadeiros, que formam uma subfamília de genes (<http://www.cellwall.stanford.edu/cesa>). Ensaios com alelos mutantes em *Arabidopsis* permitiram a inferência da função de várias subunidades de genes *CesA* e permitiram avanços importantes na investigação destes genes em outras espécies.

O mutante *rsw1*, em *Arabidopsis*, é causado por uma substituição de uma citosina para uma timina no gene *AtCesA1*, que conseqüentemente leva a alteração de uma

valina para uma alanina na posição 549 da subunidade de celulose sintase. Plantas portadoras do alelo mutante quando crescem em temperaturas pouco permissivas, apresentam uma redução na síntese de celulose, e há acúmulo de glicose não cristalina, causando anomalias morfológicas. Além disso, mutantes *rsw1* exibem um número reduzido de rosetas na membrana plasmática (Arioli et al., 1998). Scheible et al. (2001) identificaram dois alelos *ixr1-1* e *ixr1-2*, que surgem de mutações pontuais no gene *AtCesA3*, que conferem resistência ao inibidor de síntese de celulose isoxaben. Outra mutação que também confere resistência a isoxaben ocorre no gene *AtCesA6*, dando origem ao alelo *ixr2* (Desprez et al., 2002). Esses estudos permitiram a formação de um grupo de genes composto por *AtCesA1*, *AtCesA3* e *AtCesA6* que são importantes para a deposição da parede celular primária.

Já os mutantes que apresentam xilema irregular (*irx*) em *Arabidopsis* possuem uma severa redução na síntese de celulose da parede secundária. Três mutantes *irx* com paredes secundárias deficientes em celulose são formados devido a mutações em três diferentes membros da família de genes *CesA*. Uma mutação pontual no gene *AtCesA7*, dá origem ao alelo *irx3* (xilema irregular 3), que se expressa em *Arabidopsis* como um defeito na formação da parede secundária do xilema. Como resultado, os elementos traqueais no mutante *irx3* possuem paredes frágeis e colapsam facilmente (Taylor et al., 1999). Outra mutação no gene *AtCesA8* promove a formação do alelo *irx1*, provocando efeitos semelhantes aos do alelo *irx3* (Turner & Somerville, 1997). Taylor et al. (2003) identificaram o alelo mutante *irx5*, causado por uma alteração na seqüência do gene *AtCesA4*, produzindo um fenótipo também alterado em relação à deposição da parede secundária.

Os alelos *irx5*, *irx3* e *irx1* podem ser coexpressos nas mesmas células. Análises feitas por Taylor et al. (2003) sugerem que os três genes são requeridos para a formação da parede celular secundária, pois o conteúdo de celulose é muito menor nas hastes dos mutantes do que em outras partes da planta. Além disso, ensaios de imunoprecipitação usando anticorpos específicos para cada proteína foram realizados para investigar a interação das três subunidades em diferentes mutantes. No tipo selvagem cada um dos anticorpos é capaz de precipitar todas as três proteínas, corroborando com a idéia de que todas interagem em um mesmo complexo. Em plantas *irx5-1*, entretanto, não há produção da proteína IRX5, o anticorpo IRX1 não precipita IRX3 e o anticorpo IRX3 não precipita IRX1. Isto sugere que a presença das três proteínas é um pré-requisito para a correta junção

do complexo de celulose sintase. Entretanto, não é necessário que as três proteínas estejam ativas para que haja a formação do complexo. O alelo *irx1-1* é causado por uma mutação de ponto que resulta em um produto sem atividade catalítica, embora exista quantidade de proteína mutante comparável ao tipo selvagem.

Análises filogenéticas realizadas por Ranik & Myburg (2006) permitiram estabelecer a relação de ortologia, que é o compartilhamento de uma mesma origem evolutiva anterior a eventos de especiação (Russo, 2001), entre os genes de celulose sintase de *Eucalyptus* e *Arabidopsis*. Os genes *EgCesA1*, *EgCesA2* e *EgCesA3*, expressos durante a formação da parede celular secundária, são ortólogos aos genes *AtCesA8*, *AtCesA4* e *AtCesA7*, respectivamente. Já as subunidades de celulose sintase codificadas pelos genes *EgCesA5*, *EgCesA6*, e *EgCesA4*, relacionados à deposição da parede primária, são ortólogos aos genes *AtCesA1*, *AtCesA2*, e *AtCesA3*, nessa ordem.

2.4.4 O gene *EgCesA3* e seus ortólogos

O gene *EgCesA3* de *E. grandis* possui uma sequência de cDNA de um tamanho total de 3452 pb, incluindo além da ORF, as regiões 5' e 3'-UTR, correspondendo a uma proteína com 1040 resíduos de aminoácidos. Investigações sobre a expressão deste gene feitas por Ranik & Myburg (2006) revelaram que *EgCesA3* têm uma expressão 50 vezes maior em tecidos xilemáticos do que no floema e 20 vezes maior no xilema do que na casca. Os genes *EgCesA1*, *EgCesA2* e *EgCesA3* são co-expressos para a formação de celulose da parede secundária, como já foi observado em vários outros estudos com genes ortólogos a estes genes.

Em *Arabidopsis*, uma substituição nucleotídica de uma guanina para uma adenina no gene ortólogo ao *EgCesA3*, resulta em um códon de terminação, ao invés de codificar o aminoácido triptofano na posição 859, gerando um produto proteico de apenas 168 aminoácidos (Taylor, 1999). Como já foi dito, os portadores dessa alteração genética são denominados mutantes *irx3* e são caracterizados por uma severa deficiência na deposição de celulose da parede celular secundária, pois apresentam somente de 20 a 30% da quantidade de celulose encontrada em indivíduos normais (Turner & Somerville, 1997). Conseqüentemente, os vasos xilemáticos apresentam-se colapsados devido à falta de resistência à pressão negativa que é exercida pelo transporte de água, deixando o caule flexível (Figura 3). Não foi observada nenhuma compensação na regulação de lignina para manter a integridade estrutural das hastes nestes mutantes.

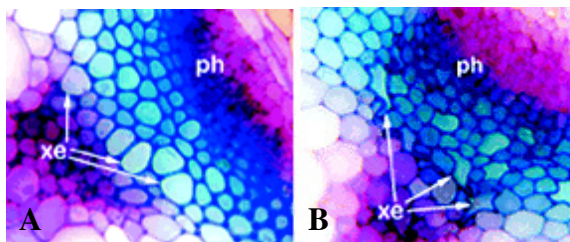


Figura 3. Secção transversal de hastes de *Arabidopsis* coradas com azul de toluidina. (A) tipo selvagem; (B) mutante *irx3*; xe: xilema, ph: floema (Fonte: Taylor et al., 1999).

Samuga & Joshi (2002) isolaram um cDNA que codifica uma subunidade de celulose sintase em *Populus tremuloides*, denominado *PtrCesA2*. O produto protéico mostrou um alto grau de similaridade e identidade, 93% e 87%, respectivamente, com o produto do gene *EgCesA3*, e um grau de identidade e similaridade muito menor, 64% e 74%, respectivamente, em relação ao gene *PtCesA1*, anteriormente caracterizado por Wu et al. (2000), sugerindo que *PtrCesA2* representa uma subunidade distinta de celulose sintase em *Populus*.

Burton et al. (2004) estudaram a família de genes *CesA* em *Hordeum vulgare* a partir do seqüenciamento de clones de cDNAs e análises de seqüências de fragmentos de DNA efetivamente expressos (*Expressed Sequence Tags* - ESTs). Os autores identificaram oito membros dessa família ao longo do genoma. De modo análogo ao que foi observado em *Arabidopsis thaliana*, *Eucalyptus* e outras espécies de plantas, foram estabelecidos dois grupos de genes que são transcritos de maneira coordenada, o primeiro formado pelos genes *HvCesA1*, *HvCesA2* e *HvCesA6*, e o outro formado por *HvCesA4*, *HvCesA7* e *HvCesA8*. Análises filogenéticas permitiram inferir que os dois grupos de genes são ortólogos aos genes de celulose sintase em *Arabidopsis*, e implicam, respectivamente, na formação da parede celular primária e secundária. Os outros dois membros de genes *CesA*, *HvCesA3* e *HvCesA5*, não apresentaram um padrão de expressão que permitisse inserí-los em nenhum dos grupos, e parecem ser regulados de maneira independente.

Tanaka et al. (2003) isolaram e caracterizaram mutantes de três genes de celulose sintase em arroz induzidos pela inserção do retrotransposon *Tos17*. O caule do arroz é composto por um colmo principal e um número variável de colmos primários e secundários (perfilhos). Nos mutantes os colmos são frágeis e quebradiços. Os três genes foram expressos em sementes, colmos, panículas e raízes, mas não foram encontrados indícios de expressão significativos em folhas maduras e, por isso, foi sugerido que

estariam relacionados à deposição da parede celular secundária. O conteúdo de celulose foi bastante reduzido nos mutantes (de 8,9% a 25,5% do que é encontrado no tipo selvagem). Uma análise filogenética feita por Ranik & Myburg (2006), indicou que os genes *EgCesA1*, *EgCesA2* e *EgCesA3* são ortólogos aos genes *OsCesA4*, *OsCesA7* e *OsCesA9*, respectivamente.

2.5 DIVERSIDADE NUCLEOTÍDICA

As espécies de *Eucalyptus* apresentam elevados níveis de variação fenotípica, dentre vários outros aspectos, quanto ao conteúdo e comprimento das fibras de celulose. De modo geral, o componente herdável desta variação tem sido amplamente aproveitado por programas de melhoramento genético da espécie, em andamento em várias instituições no país, sobretudo, ligados às indústrias produtoras de papel e celulose.

É bem conhecido o fato de que as fontes primárias da variabilidade genética são as mutações gênicas, que aumentam o número de alelos disponíveis em um loco, incrementando o número de combinações alélicas possíveis em uma população. As mutações gênicas originam-se de alterações na seqüência de bases nitrogenadas de um fragmento de DNA, mantidas durante a replicação da molécula. Essa alteração pode ser devida à substituição de nucleotídeos, ou mesmo a *indels* (inserções ou deleções de bases), originando um novo alelo capaz de codificar, por exemplo, uma nova proteína não funcional. Embora ocorram espontaneamente, as mutações também podem ser induzidas por agentes mutagênicos, como radiações e certas substâncias químicas.

As mutações ocorrem ao acaso, de modo que não é possível prever qual gene sofrerá mutação, nem relacionar a existência dessa mutação com a adaptabilidade às condições ambientais. As substituições de base que ocorrem em regiões codantes ou regulatórias dos genes e alteram o produto gênico, ou mesmo seu nível de transcrição, são denominadas mutações não-silenciosas. Por outro lado, as inserções e deleções alteram o quadro de leitura durante a tradução, e via de regra, fazem com que os aminoácidos codificados após a ocorrência de *indels* sejam muito diferentes do que seria observado na proteína do tipo selvagem, levando a formação de produtos não funcionais.

Já as mutações que ocorrem em regiões não-codantes dos genes são ditas silenciosas, pois, em geral, não resultam em alterações perceptíveis no produto gênico. Os eventos de substituição nucleotídica decorrentes dessas mutações são muito mais

frequentes, por não estarem, em geral, sob seleção direta. Já as mutações que ocorrem em sítios codantes são controladas pela seleção natural e só irão se disseminar em uma população se aumentarem o valor adaptativo de seus portadores ou se estiverem ligadas a algum outro alelo que esteja sob seleção favorável e não causarem nenhuma desvantagem à adaptação (nesse caso, serão neutras).

Na década de 1960, com a descoberta do polimorfismo de isoenzimas, foi possível se observar que a diversidade genética ao nível molecular era muito maior do que a esperada pela teoria da seleção natural. Motoo Kimura então propôs que a maioria das mutações detectáveis ao nível molecular seria de natureza neutra. As mutações neutras seriam então as responsáveis pelo aumento da diversidade de alelos em uma população, diversidade esta que seria limitada pelos efeitos da deriva genética, determinados em magnitude pelo tamanho efetivo da população.

Para estimar a diversidade genética a partir de dados de seqüências de DNA, uma medida apropriada é Π , que representa o número médio de diferenças de nucleotídeos entre duas seqüências amostradas aleatoriamente de uma população:

$$\Pi = \frac{\sum_{i < j} \Pi_{ij}}{[n(n-1)/2]}$$

em que Π_{ij} é o número de diferenças entre as seqüências i e j , e $n(n-1)/2$ é o número de pares de seqüências. Watterson (1975) demonstrou que a esperança matemática de Π é dada por:

$$E[\Pi] = \theta$$

em que $\theta=4Ne\mu$, sendo Ne o tamanho efetivo de uma população diplóide e μ a taxa de mutação por sítio, por geração.

Segundo Li (1997), o valor de Π é influenciado pelo comprimento das seqüências, assumindo valores crescentes com o aumento do tamanho dessas, pois $v=L\mu$, em que L é o número de nucleotídeos numa seqüência e v é a taxa de mutação por seqüência. Por isso, este autor sugere que não convêm usar Π para comparar variações de seqüências de DNA entre locos diferentes. Nei & Li (1979) standardizaram Π por L :

$$\pi = \Pi / L$$

sendo π o número médio de diferenças nucleotídicas por sítio entre duas seqüências aleatoriamente amostradas em uma população.

Uma outra estimativa comumente utilizada de diversidade nucleotídica é o valor θ_w de Watterson, que pode ser obtida por:

$$\theta_w = \frac{K}{a}$$

em que K é o número de sítios segregantes, e $a = 1 + \frac{1}{2} + \dots + \frac{1}{n-1}$, e n é o número de seqüências amostradas.

2.6 POLIMORFISMOS DE UM ÚNICO NUCLEOTÍDEO (SNPs)

Cada vez mais, genomas são completamente seqüenciados (como o genoma humano e o de *Arabidopsis*), e o interesse dessas pesquisas está voltado para a descoberta e análise de diferentes seqüências intraespecíficas. Por outro lado, ainda são limitadas as informações sobre freqüência, natureza e distribuições de SNPs, polimorfismos de um único nucleotídeo, em plantas. Embora existam vários estudos em espécies cultivadas, como soja (Zhu et al., 2003), milho (Remington et al., 2001; Ching et al., 2002), arroz (Nasu et al., 2002) e em vegetais-modelos como *Arabidopsis* (Cho et al., 1999; Nordborg et al., 2002) e até mesmo em espécies arbóreas como *Pinus* (Pot, et al., 2005), a catalogação e a avaliação de níveis de associação envolvendo SNPs em humanos estão consideravelmente mais avançados. Estas informações, já há alguns anos, têm sido utilizadas na caracterização de fenótipos como susceptibilidade e resistência a doenças (Collins et al., 1997) e respostas a drogas (Kleyn & Vessel, 1998). Para Jorde (2000), esses polimorfismos podem ser usados como simples marcadores genéticos, mas podem ser identificados entre as extremidades de qualquer gene, e têm um uso potencial na detecção de associações entre formas alélicas de genes e fenótipos de interesse.

Em uma definição mais conservadora, os SNP são tidos como porções do DNA genômico com diferentes seqüências alternativas (alelos) em posições de uma única base nucleotídica, com uma freqüência alélica populacional de pelo menos 1% (Brookes, 1999). Na prática, o termo SNP é usado mais livremente em relação ao que foi definido acima. Variações de uma única base em cDNAs (cSNPs) são usualmente classificadas como SNPs, desde que possam ser verificados no DNA genômico. Isso, entretanto, ignora a possibilidade de que eles possam ser resultantes da edição do RNA mensageiro. Além disso, *indels*, por vezes são catalogados como SNPs e vários trabalhos tem relatado SNPs com freqüências menores do que 1%.

Os SNPs ocorrem tanto em regiões codificadoras como em regiões não codificadoras dos genomas de diferentes espécies. Em regiões codificadoras, podem alterar a seqüência de aminoácidos na cadeia protéica, sendo denominados, neste caso, de SNPs não sinônimos, podendo a substituição ser conservativa ou não conservativa, em função das características dos aminoácidos envolvidos na troca. Nesses casos, pode haver modificações estruturais e funcionais na proteína.

Os SNPs podem ser sinônimos devido à propriedade de o código genético ser redundante, ou seja, a mudança em um nucleotídeo da molécula, pode não alterar a seqüência protéica. Porém, mesmo assim, vários estudos têm relatado que SNPs sinônimos podem modificar a estrutura e a estabilidade do RNA mensageiro, e, conseqüentemente afetar a quantidade de proteína produzida, através de algum mecanismo ainda desconhecido. Genes contendo um ou mais SNPs sinônimos podem originar duas ou mais formas de mRNAs. Essas variantes de mRNA podem promover diferentes funções biológicas como resultado de diferenças estruturais que afetam a sua interação com outros componentes celulares. Conforme relatado por Duan et al. (2003), existem evidências de que um polimorfismo na posição 957 do gene que codifica o receptor para dopamina afete a expressão *in vitro*. O efeito do polimorfismo foi associado a mudanças na estrutura secundária do mRNA, diminuindo sua estabilidade, apresentando um papel importante na regulação da expressão gênica. Shen et al. (1999) também observaram diferenças na estrutura secundária de mRNA associadas a SNPs em regiões codificantes de dois mRNAs em humanos: alanila-tRNA-sintase e mRNA de 70-kDa da proteína A.

Um SNP sinônimo pode ainda promover *splicing* alternativo, mudanças nos códons de iniciação de tradução, além de geração ou supressão de códons de terminação ou poliadenilação na molécula de RNA mensageiro (Guimarães & Costa, 2002).

Para os geneticistas de plantas o investimento em novas tecnologias como os SNPs têm merecido atenção especial. O alto nível de polimorfismo de muitas espécies vegetais, como o milho, por exemplo, facilitam a identificação de SNPs e demonstram a aplicabilidade destas técnicas em programas de seleção assistida. Rafalski (2002) salienta que as linhas puras, comumente obtidas em várias espécies vegetais cultivadas, quando avaliadas, permitem leitura direta dos haplótipos, que são combinações de alelos encontrados em locos próximos, ao longo de um cromossomo, herdados em conjunto (Rafalski, 2002). Este fato permite a determinação precisa da fase de ligação entre os SNPs, sem que dados de mais de uma geração tenham que ser obtidos.

A identificação de SNPs responsáveis por variações em caracteres quantitativos, como o conteúdo de celulose em espécies de *Eucalyptus*, ou em outras características intrínsecas à formação da madeira, pode ser uma ferramenta importante para o melhoramento genético no futuro. Tal estratégia permitiria, pelo estabelecimento da associação entre o polimorfismo genético e as diferenças no conteúdo e qualidade de fibras, a identificação de combinações genotípicas mais satisfatórias para a produção de papel, por exemplo, pela utilização de procedimentos de genotipagem automatizada.

Várias rotas para o descobrimento de SNPs podem ser traçadas. Isto inclui o re-seqüenciamento de produtos de PCR (reação da polimerase em cadeia) com ou sem pré-*screening*; a descoberta de SNP eletrônicos (eSNP) em bibliotecas genômicas de *shotgun*, e de eSNP em bibliotecas de EST (etiquetas de seqüências expressas). Segundo Rafalski (2002), o seqüenciamento de segmentos de DNA (amplificados por PCR) é o caminho mais direto para a identificação de SNP. Iniciadores para PCR são desenhados para amplificar segmentos de DNA de 400 a 700 pb, geralmente a partir de EST do gene de interesse (Figura 4). A seleção de fragmentos amplificados de regiões não-codantes, como íntrons, ou regiões 3' não traduzidas (3'- UTR), geralmente aumentam a freqüência do encontro de polimorfismos. A PCR é otimizada com a localização de regiões de alta divergência em indivíduos que representem à diversidade da população de interesse. As seqüências resultantes são alinhadas, e o polimorfismo identificado, sendo possível à distinção entre o polimorfismo verdadeiro e o erro de seqüenciamento, através da utilização de programas como o PHRED (Ewing et al., 1998), que interpreta a informação obtida a partir dos eletroferogramas associando uma medida de qualidade para cada base nucleotídica.

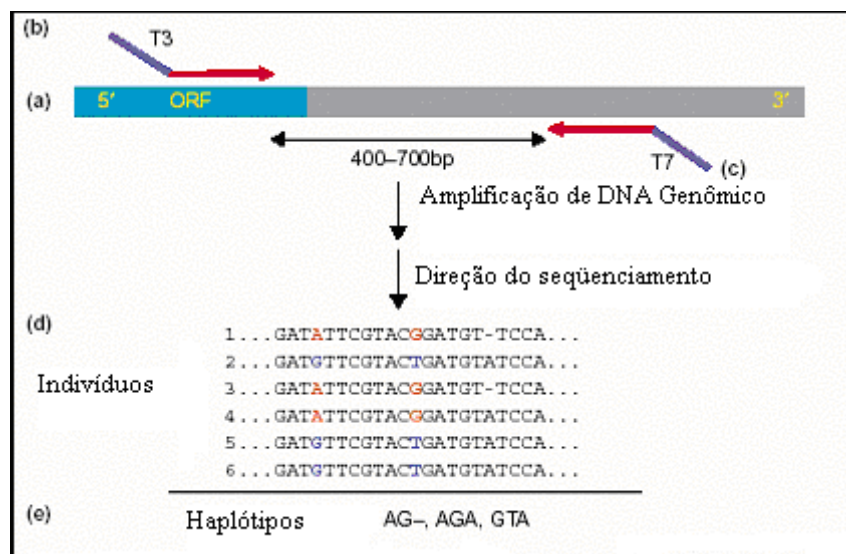


Figura 4. Descoberta de SNP em EST. (a) Uma seqüência na extremidade 3' de um EST ou um gene conhecido é usado para desenhar um par de iniciadores; (b,c) são usados iniciadores T3/T7 para facilitar o seqüenciamento direto. Depois da amplificação, os produtos da PCR são purificados; (d) as seqüências resultantes são alinhadas e são identificados os SNPs e polimorfismos de inserção e deleção (e) três haplótipos podem ser distinguidos (Fonte: Rafalski, 2002).

A freqüência de ocorrência de SNPs varia entre regiões do genoma, entre seqüências codificantes e não codificantes, e entre espécies. A magnitude da diversidade nucleotídica varia de 0,0003 até 0,0050 em humanos, ou seja, existem de 3 até 50 SNPs por 10 kb quando dois cromossomos são comparados (Chakravarti, 1999).

Em um estudo realizado por Bhatramakki et al. (2002), com base em oito linhagens elite de milho, foram analisados 502 locos derivados de ESTs, sendo re-seqüenciados 400-500pb por loco. Desse total, em 86% dos locos foram encontrados polimorfismos. Mais da metade dos locus (52%) foram polimórficos em duas linhagens comuns, B73 e Mo17. Os autores identificaram 215 locos (43% do total analisado) que continham polimorfismos de inserção e deleção do tamanho de 1pb.

Zhu et al. (2003), em um estudo sobre a diversidade de seqüências de genótipos de soja, encontraram uma freqüência de 1,64 SNPs por kb em regiões codificantes e 4,85 SNPs por kb em regiões não codificantes. Um terço dos terminais 3' dos genes testados (com um comprimento médio de 430 pb) contém um SNP.

Uma dificuldade surge em análises utilizando SNP se não são avaliadas linhas puras, ou indivíduos homocigotos (Rafalski, 2002). Neste caso, *softwares* como o *PolyPhred*, facilitam a identificação de posições heterocigóticas (Nicherson et al., 1997).

Devido ao desconhecimento da fase de ligação de SNPs adjacentes a alelos em indivíduos heterozigotos, um haplótipo pode não ser derivado imediatamente do seqüenciamento de produtos de PCR. Ou seja, se um loco tem um polimorfismo A/G e um loco adjacente tem o polimorfismo C/T, isto não indica se os dois haplótipos são A (no loco 1)-C(no loco 2) e G(loco 1)-T(loco2), ou A (loco 1)-T(loco2) e G(loco 1)-C(loco 2). Por isso, métodos computacionais de derivação de haplótipos corretos têm sido desenvolvidos.

Vários métodos para genotipagem de SNP podem ser usados e esse assunto tem sido amplamente discutido (Gut, 2001; Kwok, 2001). A escolha do método particular depende de muitos fatores, incluindo custos, equipamentos necessários, dificuldade no desenvolvimento do ensaio e potencial de multiplexagem.

Numerosos tipos de marcadores de DNA baseados na detecção indireta do polimorfismo a partir de seqüências de DNA foram desenvolvidos. Frequentemente, marcadores muito informativos, como os microsátélites (SSR), são preferidos. Entretanto, marcadores microsátélites são pouco adequados para estudos de associação devido à ocorrência de homoplasia, isto é, a ocorrência de alelos SSR de tamanho idêntico, mas diferentes origens evolutivas (Rafalski, 2002). SNPs possuem a vantagem de serem muito mais abundantes no genoma do que os SSR. Por outro lado, SNPs são, em geral, bialélicos e a heterozigosidade esperada é baixa. Taramino & Tingey (1996) encontraram, por exemplo, uma heterozigosidade esperada de 0,263 em um conjunto de linhagens elite de germoplasma de milho, comparado com 0,77 para os microssatélites. A informação provida pelos SNPs é muito útil quando vários intervalos próximos ao polimorfismo definem um haplótipo na região examinada. Na presença de um pequeno desequilíbrio de ligação, alguns dos SNPs redundantes podem ser suficientes para definir um haplótipo. No exemplo mostrado na Figura 5, os três SNPs indicados por asteriscos definem completamente quatro haplótipos. Todo haplótipo definido por um subconjunto dos SNPs existentes é denominado de “*tag*” (Jonhson et al., 2001).

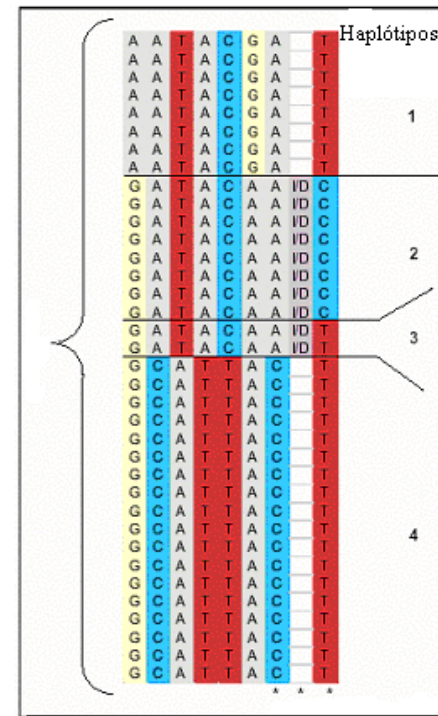


Figura 5. Polimorfismos de um único nucleotídeo foram identificados em 263 segmentos de um gene de ACP-esterol desaturase, em milho. As seqüências horizontais correspondem a 32 indivíduos seqüenciados. As colunas verticais identificam nove sítios polimórficos, incluindo um indel (I/D), que é um polimorfismo de inserção/deleção. São mostrados quatro haplótipos distintos. Estes quatro haplótipos podem ser identificados, sem ambigüidade, usando-se somente três SNPs (marcados com *) (Fonte: Rafalski, 2002).

2.7 DESEQUILÍBRIO DE LIGAÇÃO

2.7.1 Conceito

O desequilíbrio de ligação (DL) é também conhecido como desequilíbrio de fase gamética, desequilíbrio gamético ou associação alélica. Num sentido mais restrito, o DL seria uma associação não aleatória de alelos de locos diferentes. Isto é, o desequilíbrio de ligação é uma medida do grau de correlação entre polimorfismos (SNP, por exemplo), que é causada por eventos compartilhados de mutação e recombinação. Numa população em que os cruzamentos ocorrem aleatoriamente, com a segregação de locos acontecendo independentemente, e na ausência de seleção, mutação, ou migração, os locos polimórficos podem estar em equilíbrio de ligação. Em contraste, ligação, seleção e cruzamentos preferenciais tendem a aumentar o DL.

Os termos ligação e desequilíbrio de ligação são muitas vezes confundidos. Embora o DL e a ligação gênica sejam relacionados, são conceitualmente distintos. A ligação se refere à herança correlacionada de locos através de uma conexão física no cromossomo, enquanto o DL refere-se a uma correlação entre alelos de uma população. A confusão ocorre porque uma forte ligação pode resultar em altos níveis de desequilíbrio gamético. Por exemplo, se duas mutações ocorrem a poucos pares de bases uma da outra, elas estarão sob as mesmas pressões de seleção e deriva genética ao longo do tempo. Devido à recombinação entre as bases vizinhas ser rara, a presença desses SNPs é altamente correlacionada a uma forte ligação, resultando em um alto desequilíbrio de ligação. Em contraste, SNPs em cromossomos separados experimentam diferentes pressões seletivas, e segregam independentemente, então, estes SNPs têm uma baixa correlação ou nível de DL (Flint-Garcia et al., 2003).

2.7.2 Forças evolutivas que afetam o DL

Muitos dos processos observados em genética de populações são refletidos no padrão do desequilíbrio de ligação entre locos, como a frequência alélica e a recombinação. Mutações promovem a matéria-prima para a produção de polimorfismos que estão em desequilíbrio de ligação. A recombinação é um fenômeno que enfraquece o DL intracromossomal, enquanto o desequilíbrio de ligação intercromossomal é quebrado pela segregação independente durante a meiose.

O tipo de sistema reprodutivo também influencia fortemente a magnitude de desequilíbrio alélico. Em geral, o DL decai mais rapidamente em espécies alógamas em comparação a espécies autógamas. Por outro lado, as espécies autógamas podem sofrer um aumento na taxa de recombinação por meiose. Estima-se que a taxa de recombinação em *Arabidopsis* seja de duas a seis vezes maior do que é verificado em *Drosophila* e milho, respectivamente. Entretanto, a autofecundação aumenta a homozigose e, por isso, limita o número de duplos heterozigotos que podem ser embaralhados pela recombinação. Isso resulta em uma taxa de recombinação efetiva menor em espécies que apresentam autofecundação, fazendo com que os polimorfismos genéticos tendam a permanecer correlacionados. Sendo assim, espera-se que o DL seja mantido em distâncias físicas maiores (Gaut & Long, 2003).

O DL pode surgir em populações que tenham passado por uma recente redução do tamanho populacional acompanhado de intensa deriva genética. Durante o afinilamento

genético, somente poucas combinações de alelos serão passadas às futuras gerações, e isto pode resultar em um forte desequilíbrio gamético. Em estudos de genética humana, populações que foram submetidas a severos gargalos genéticos têm sido utilizadas para mapeamento do DL em vários genes que são responsáveis por doenças.

A seleção natural, quando produz afinilamento loco-específico, também causa um desequilíbrio de ligação entre o alelo selecionado e os locos ligados a esse. Entretanto, a seleção quando favorável a um fenótipo que é controlado por dois locos não ligados (epistasia) pode resultar em DL, apesar do fato de que os locos não estejam conectados fisicamente.

2.7.3 Como medir o DL

Sendo o desequilíbrio de ligação relacionado à não independência de alelos, a mensuração desse desequilíbrio é uma medida quantitativa de associação alélica. Na Figura 6A é possível observar dois sítios polimórficos que estão em completo desequilíbrio de ligação, porque o alelo **A** aparece sempre associado a **G**, e o alelo **T** está sempre associado ao alelo **C**. Contudo, no exemplo B, os alelos encontrados no primeiro sítio, **A** e **T**, se associam de modo aleatório aos alelos do segundo, pois estão em equilíbrio de ligação.

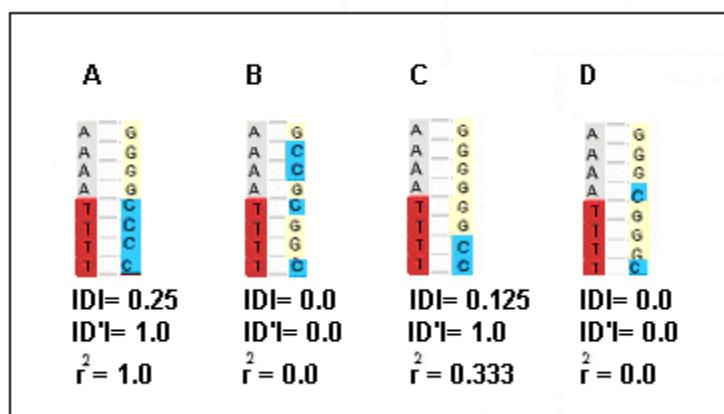


Figura 6. Dois locos mostrando SNPs em oito indivíduos. A) DL completo entre os locos; B e D) Recombinação entre os sítios; C) frequências de associações desiguais entre os locos, não implicando necessariamente em recombinação. C pode ser uma nova mutação associada a T (Fonte: Gaut & Long, 2003, com adaptações).

Para calcular o desequilíbrio de ligação entre os sítios 1 e 2, em relação à Figura 6, considera-se $D_{12} = p_{1A2G} - (p_{1A})(p_{2G})$, de modo que p_{1A2G} é a frequência de

seqüências que contém A no sítio 1 e G no sítio 2, p_{1A} é a freqüência absoluta do alelo A no sítio 1, e p_{2G} é a freqüência do alelo G no sítio 2. Em relação à Figura 6A, $D = 0.5 - (0.5)(0.5) = 0.25$, e de modo análogo, para a Figura 6B, $D = 0.0$. Ressalta-se que o valor absoluto de D é simétrico, não importando quais alelos são medidos e associados, ou seja, para se calcular D pode-se medir a freqüência de A ou T.

Segundo Weir (1996), levando em conta uma população de gametas de tamanho $2n$, pode-se obter o seguinte estimador para o desequilíbrio de ligação:

$$E(\hat{D}_{12}) = \frac{2n-1}{2n} D_{12}$$

cuja variância é dada por:

$$Var(\hat{D}_{12}) = \frac{1}{2n} [p_1(1-p_1)p_2(1-p_2) + (1-2p_1)(1-2p_2)D_{12} - D_{12}^2]$$

Para testar a hipótese da não existência do desequilíbrio, $H_0: D_{12} = 0$, pode-se construir um teste de chi-quadrado a partir do esquadramento de um variável normal assintótica z :

$$z = \frac{(\hat{D}_{12}) - E(\hat{D}_{12})}{Var(\hat{D}_{12})}$$

$$X_{12}^2 = z^2 \quad X_{12}^2 = \frac{2n(\hat{D}_{12}^2)}{p_1(1-p_1)p_2(1-p_2)}$$

A medida D incorpora informação sobre associação e freqüências alélicas. Por exemplo, imagine que somente uma entre as oito seqüências contenha um A no loco 1 e um G no loco 2, de modo que as seqüências remanescentes contêm um T e um C. Embora os dois locos ainda estejam em completo desequilíbrio, a mudança na freqüência dos alelos A e G gera uma medida de $D = 0,125 - (0,125)(0,125) = 0,109$. Devido a essa dependência de freqüência, valores de D podem variar muito, mesmo quando os SNPs estão em completo desequilíbrio (Gaut & Long, 2003).

Existem duas possibilidades para se contornar a dependência do valor de D em relação à freqüência absoluta dos alelos. A primeira é ignorar as variantes com freqüências baixas. Muitos estudos de DL ignoram alelos com freqüência menores que 5 ou 10%. A

outra, é usar medidas de associação que sejam normalizadas com a respectiva frequência dos alelos.

A medida normalizada de D mais comum é D' (Lewontin, 1964). D' tem a propriedade de que o valor absoluto seja 1 se SNPs estão em completo desequilíbrio gamético, ou se existem somente 3 tipos gaméticos na amostra (Figura 6C). Um inconveniente de D' é que seu significado fica mais difícil de ser compreendido quando $|D'| < 1.0$.

Outra estimativa de DL pode ser obtida pelo cálculo de r^2 , que é igual a D^2 dividido pelo produto das frequências alélicas nos dois locos:

$$r^2 = \frac{(D^2)}{(p_{1A})(p_{2G})(p_{1T})(p_{2C})}$$

Hill & Robertson (1968) deduziram que em pequenas populações, quando a deriva genética e a recombinação estão em equilíbrio, o desequilíbrio de ligação pode ser estimado por:

$$r^2 = \frac{1}{(1 + 4Nc)}$$

sendo N o tamanho efetivo populacional e c a fração de recombinação entre os locos.

De modo geral, a significância estatística (p-valor) para o valor estimado do desequilíbrio de ligação de uma população, em relação à hipótese de nulidade, dada pela inexistência de desequilíbrio, pode ser calculada utilizando-se o Teste Exato de Fisher, quando são comparados dois sítios com dois alelos cada, ou podem ser feitas análises de permutações multifatoriais para comparar sítios com mais de dois alelos.

2.7.4 Estudos de DL em plantas

Geralmente, o desequilíbrio de ligação esperado ao longo do genoma varia muito devido ao acaso. Entretanto, é possível traçar alguns padrões gerais de desequilíbrio. O DL entre dois locos tende a decair com o aumento da distância em centiMorgans e a taxa de decaimento depende muito da história demográfica da população, bem como de vários outros parâmetros. Espécies que se reproduzem através da autofecundação apresentam altos níveis de homozigose e a extensão do desequilíbrio de ligação tende a ser muito maior do que em espécies de fecundação cruzada.

Arabidopsis thaliana têm um índice de autogamia estimado em 99% e um padrão de distribuição de suas populações altamente condensado, resultando em uma diversidade nucleotídica estimada de 0,0014 (Bergelson et al., 1998). Nordborg et al. (2002) seqüenciaram 13 pequenos fragmentos que cercam o alelo FRI, que regula o tempo de florescimento em *Arabidopsis*, a partir de 20 acessos de uma amostra global e encontraram que o DL decaiu rapidamente com a distância de 1cM, correspondente a 250Kb.

Já em espécies alógamas, os padrões de DL variam substancialmente com o tipo de população escolhida. Tenaillon et al. (2001) investigaram diversas seqüências em 21 locos do cromossomo 1 de milho e o desequilíbrio de ligação r^2 diminuiu para menos de 0,25, numa extensão de 200 pb em média. Remington et al. (2001) examinaram o DL entre SNPs intragênicos, em um grupo composto de 102 linhagens melhoradas, uma amostra representativa da variabilidade genética global do milho. Observando seis genes candidatos, o DL intragênico decaiu rapidamente ($r^2 < 0,1$ com 1500bp). Entretanto, alguns locos apresentaram desvios para esta pequena extensão do DL. Um gene com um DL baixo em relação ao loco *sul* teve seu valor diminuído ainda mais após 12Kb. Rafalski (2002) sugere que a extensão do LD é maior que 100Kb entre os locos *adh1* e *y1*, em populações elite de milho, indicando que este germoplasma teria uma base genética mais estreita em relação ao que foi estudado por Tenaillon et al. (2001). Existem várias hipóteses que poderiam explicar a variação dos padrões de DL em milho observada em diferentes amostras. Vários haplótipos que foram identificados foram gerados, provavelmente, antes da domesticação de seu ancestral, e as diferentes taxas de decaimento do DL refletiriam níveis diferentes de gargalos demográficos populacionais sofridos pela espécie.

2.8 ESTUDOS DE EXPRESSÃO GÊNICA *IN SILICO*

Os estudos de expressão gênica em vários organismos, desde procariotos a humanos, tiveram consideráveis avanços nas últimas décadas. Sendo que, neste contexto, o seqüenciamento de fragmentos de ESTs a partir de bibliotecas de cDNA contribuiu muito para o entendimento de como vários processos biológicos são coordenados ao nível molecular.

Bancos de ESTs têm sido uma fonte importante de informação sobre a função de genes candidatos. Através de comparações entre uma seqüência de interesse e um banco

de dados é possível fazer inferências sobre sua função química e biológica (Carneiro et al., 2000).

Além disso, as ferramentas de análise de expressão gênica têm permitido a realização de inferências sobre quais genes são expressos em determinadas fases de desenvolvimento, além da comparação destes níveis de expressão com aqueles obtidos em diferentes tecidos. Essa informação é de grande utilidade em estudos destinados à identificação dos genes envolvidos na determinação de caracteres de grande interesse econômico, bem como dos determinantes da variação observada nestes caracteres em diferentes ambientes.

Existem diferentes abordagens para se mensurar a expressão gênica. Figueiredo et al. (2001), visando identificar genes relacionados especificamente ao processo de florescimento de cana-de-açúcar, usaram a abordagem denominada “*Digital Differential Display*”, a partir da contagem direta dos ESTs de 32 bibliotecas de cDNA não-normalizadas. Geralmente, a frequência de um determinado EST está correlacionada ao acúmulo de transcritos nos tecidos dos quais as bibliotecas de cDNA foram construídas e neste caso é possível se comparar os transcritos de diferentes tecidos, pela análise computacional de bancos de dados de ESTs, além de testes estatísticos robustos, como os preconizados por Audic & Claverie (1997) e Romualdi et al. (2001).

Uma outra tecnologia de utilização crescente em estudos de expressão gênica é a análise de *microarrays*, que permite a comparação do nível de expressão de milhares de genes ao mesmo tempo. Os *arrays* são arranjos de um grande número de genes sobre um suporte, como por exemplo uma lâmina de vidro. Por esta técnica, milhares de fragmentos de cDNA são marcados com fluorescência e hibridizados com oligonucleotídeos (sondas) imobilizados na lâmina de vidro. As sondas são formadas por oligonucleotídeos de seqüência pré-determinada de modo a representar os milhares de genes de interesse. Um *array* de 1,6cm² pode abrigar mais de um milhão de sondas diferentes com um espaçamento de 1µm (Primrose, 2003).

O desenho das sondas é uma etapa essencial no processo de construção de um *array*. Em organismos eucariotos, a partir de uma seqüência de mRNA de referência, são sintetizados pares de sondas selecionados preferencialmente a partir da região 3' do gene, ou de transcritos que possuam uma cauda poli-A, para reduzir problemas que possam surgir com o uso de mRNA degradados (Lipshutz et al., 1999). Os resultados são produzidos sob a forma de diferentes intensidades de fluorescência, que são captadas por

microscopia de fluorescência a laser, em função de diferentes níveis de expressão de cada gene, sendo toda informação processada por computadores e programas específicos. Para Carneiro (2000), análises feitas a partir de *microarrays* podem complementar metodologias como a “*Digital Differential Display*” por gerar informações quantitativas, capazes de detectar mudanças súbitas na expressão dos genes avaliados.

Em *Eucalyptus*, Kirst et al. (2004) investigaram a expressão coordenada de genes envolvidos na biossíntese de lignina, a partir da análise conjunta de dados de *microarrays* e de medidas fenotípicas do crescimento secundário, obtendo o mapeamento dos QTLs relacionados ao nível de expressão de diferentes genes (eQTLs). A madeira, composta por xilema secundário, um tecido especializado de condução e suporte produzido pelo crescimento e diferenciação do câmbio vascular meristemático, apresenta características físicas e químicas dependentes do nível de transcrição de diversos genes expressos durante o processo de xilogênese (Fukuda et al., 1996). Neste estudo, os autores analisaram a expressão de 2.608 genes transcritos durante a diferenciação do xilema em 91 indivíduos originados de uma progênie de um cruzamento de *E. grandis* x *E. globulus*, retrocruzada com um indivíduo não-relacionado de *E. grandis*. Eles relacionaram os resultados obtidos com QTLs para o crescimento em diâmetro, anteriormente mapeados por Myburg (2001). Cada gene foi testado para associação entre o nível de transcrição e a variação do crescimento em diâmetro na progênie através de uma análise de correlação. A análise revelou uma redução coordenada dos transcritos de genes que codificam enzimas envolvidas na biossíntese de lignina, na progênie que exibiu um crescimento superior. A correlação mais significativa foi observada entre a expressão de um gene que codifica a enzima coniferaldeído 5-hidroxilase (Cald5H), envolvida na via biossintética de siringil, um dos álcoois aromáticos que compõem a molécula de lignina. Indivíduos que apresentaram um nível de expressão maior de Cald5H tinham um diâmetro de caule menor.

3 MATERIAL E MÉTODOS

3.1 MATERIAL GENÉTICO PARA O ESTUDO DE DIVERSIDADE NUCLEOTÍDICA DO GENE *EgCesA2*

Para o estudo de avaliação da magnitude de diversidade nucleotídica presente no gene *EgCesA2*, ortólogo ao gene *AtCesA4* de *Arabidopsis thaliana*, foram amostradas folhas de 12 indivíduos de diferentes espécies de *Eucalyptus* (*E. grandis*, *E. urophylla*, *E. globulus*, *E. dunnii* e *E. camaldulensis*) e de dois híbridos interespecíficos: *E. grandis* x *E. dunnii* e *E. urophylla* x *E. globulus*, utilizados em programas de melhoramento no Brasil. Os dois indivíduos representantes de *E. grandis* são originários de diferentes regiões da Austrália, ilha de Atherton (17°16'S e 145°29'E) e região de Coffs Harbour (30°19'19''S e 153°07'01''E). Os dois indivíduos representantes da espécie *E. urophylla* são provenientes de duas ilhas da Indonésia, Timor (8°50'S e 125°55'E) e Flores (latitudes entre 8 e 11°S e longitudes entre 118° e 125°). Os genótipos das espécies *E. globulus*, *E. dunnii* e *E. camaldulensis*, representadas por dois indivíduos cada e dos híbridos, representados por um indivíduo cada, foram provenientes de diferentes plantações florestais localizadas no Brasil. O DNA foi extraído conforme descrito no item 3.3.

Tabela 1. Relação dos indivíduos utilizados no estudo de diversidade nucleotídica do gene *EgCesA2*.

Indivíduo	Espécie
E.GR	<i>E. grandis</i>
G44	<i>E. grandis</i>
E.U.1	<i>E. urophylla</i>
E.U.2	<i>E. urophylla</i>
E.C.1	<i>E. camaldulensis</i>
E.C2	<i>E. camaldulensis</i>
E.D.1	<i>E. dunnii</i>
E.D.2	<i>E. dunnii</i>
E.G.1	<i>E. globulus</i>
E.G.2	<i>E. globulus</i>
E.U.G	<i>E. urophylla</i> x <i>E. globulus</i>
E.GR.D	<i>E. grandis</i> x <i>E. dunnii</i>

3.2 MATERIAL GENÉTICO PARA O ESTUDO DO DESEQUILÍBRIO DE LIGAÇÃO NO GENE *EgCesA3*

Para o estudo do desequilíbrio de ligação no gene *EgCesA3* em *Eucalyptus*, foram amostradas folhas de 32 indivíduos de uma população base de *E. urophylla*, estabelecida no Brasil pela empresa Suzano Celulose e Papel, em um plantio florestal localizado na cidade de Suzano-SP (23°32'33''S e 48°18'39''W). Essa população base foi constituída a partir de sementes coletadas ao acaso de uma floresta natural de *Eucalyptus*, localizada no Timor (8°50'S e 125°55'E).

As folhas foram previamente estocadas a -80°C e a partir delas foi extraído o DNA genômico, conforme o protocolo a seguir.

3.3 EXTRAÇÃO DO DNA GENÔMICO

A extração do DNA genômico de *Eucalyptus* foi feita a partir do protocolo desenvolvido por Doyle & Doyle (1991), seguindo as modificações propostas por Ferreira & Grattapaglia (1998). Cada amostra foi composta por cerca de 100mg de folhas frescas congeladas, contidas em microtubos plásticos de 2mL. Ao conteúdo dos microtubos foi adicionado nitrogênio líquido para permitir a maceração dos tecidos, feita com o auxílio de um bastão de vidro. Para o rompimento das membranas celulares e para a separação de ácidos nucleicos e polissacarídeos, após a maceração, foram adicionados 700µL de tampão de extração CTAB (CTAB 2%; EDTA, 20mM; Tris-HCl, 100mM, pH 8,0; β-mercaptoetanol, 0,2%) pré-aquecido a uma temperatura de 65°C. Os tubos foram mantidos em banho-maria a uma temperatura de 60°C por um período de 30 minutos, sendo agitados a cada 10 minutos para manter o extrato em suspensão. Aos microtubos, já em temperatura ambiente, foram adicionados 600µL de clorofórmio-álcool isoamílico (24:1). Depois de homogeneizados, os microtubos foram centrifugados a 13000rpm, durante 5 minutos. A fase aquosa superior foi removida para um novo tubo e adicionou-se 50µL de CTAB 10%. Após a homogeneização, a extração com clorofórmio-álcool isoamílico foi repetida. A fase aquosa superior foi transferida para um novo tubo e acrescentou-se 400µL de isopropanol frio, seguida de agitação leve e incubação *overnight* a -20°C. Depois disso, os tubos foram centrifugados a 7500 rpm por 5 minutos, para a formação do *pellet*. O sobrenadante foi descartado e os *pellets* foram lavados duas vezes em 1mL de etanol 70%, por 10 minutos, e

em etanol absoluto, uma vez, por 3 minutos. Depois de secos, os *pellets* foram ressuspendidos em 50 μ L de tampão TE (Tris-HCl, 10mM, pH 8; EDTA, 1mM).

3.4 QUANTIFICAÇÃO DO DNA

As amostras de DNA extraídas foram analisadas através da eletroforese em gel de agarose a 1,4%, em tampão TBE (Tris-Ácido Bórico, 90mM; EDTA, 2mM) e coradas em brometo de etídio, com a finalidade de verificar a integridade e a quantidade de DNA. A eletroforese foi feita numa cuba horizontal sob voltagem constante de 120V.

A quantificação do DNA foi feita por meio da comparação entre a intensidade das bandas visualizadas nas amostras extraídas, com aquela observada nos padrões de DNA do fago λ (Invitrogen, Inc.), nas concentrações de 20, 50 e 100ng/mL. Os resultados obtidos foram digitalizados com o auxílio de um fotodocumentador KODAK EDAS 290, utilizando-se luz UV. As amostras, após a quantificação, foram diluídas de modo a fazer com que a concentração de trabalho fosse de 2,5ng/ μ L.

3.5 PRIMERS UTILIZADOS

Os iniciadores específicos destinados à amplificação e seqüenciamento dos dois diferentes genes de celulose sintase em *Eucalyptus* foram desenhados por Salazar (2006). Para isso, foi construído um dendograma a partir de seqüências publicadas no Genbank, através do site do NCBI (<http://www.ncbi.nlm.nih.gov/>), baseado nos genes da família *CesA* anteriormente caracterizados em *Arabidopsis*. O alinhamento e a subsequente reconstrução filogenética baseada nas distâncias entre as seqüências foram executados pelo programa ClustalX (Thompson, 1997). O algoritmo Neighbor-Joining (Saitou & Nei, 1987) foi utilizado na análise e os resultados foram visualizados pelo programa TreeView (Page, 1996). A partir desse dendograma, duas seqüências foram escolhidas: AB006703, AF088917, correspondentes, respectivamente, aos genes *AtCesA4*, *AtCesA7*, descritos por Holland et al. (2000), ortólogas ao gene *EgCesA2*, e *EgCesA3*, de *E. grandis*. Essas seqüências foram comparadas às do banco de EST do Projeto Genolyptus (Rede Brasileira de Pesquisa do Genoma de *Eucalyptus*), através do algoritmo BLAST (Altschul et al., 1990) para a identificação de seqüências homólogas e foram selecionadas 320 seqüências com um valor-E menor do que 10^{-10} . Por fim, após novo alinhamento, houve a formação de dois grupos de homologia, cada um agrupando seqüências de *Eucalyptus* ortólogas a um

dos genes *CesA*: *AtCesA4* ou *AtCesA7*, permitindo o desenho dos iniciadores para a reação de amplificação dos fragmentos dos genes de celulose sintase relacionados à produção de celulose da parede secundária.

Tabela 2. Genes *CesA* de *Arabidopsis thaliana* utilizados como referência.

gene	n° de acesso (<i>GenBank</i>)	Referência
<i>AtCesA4</i>	AB006703	Holland <i>et al.</i> (2000)
<i>AtCesA7</i>	AF088917	Holland <i>et al.</i> (2000)

Salazar (2006) analisou os genes de forma a posicionar os íntrons e éxons, através da comparação da seqüência consenso de cada alinhamento, com a versão genômica correspondente aos genes homólogos de *Arabidopsis*, utilizando o programa EST2GENOME (Mott, 1997). Os *primers* foram desenhados com o auxílio do programa Primer3 (Rozen & Skaletsky, 2000).

Tabela 3. Iniciadores desenhados para amplificação e/ou seqüenciamento dos genes de celulose sintase de *Eucalyptus*.

Ortologia	Nome	Seqüência (5'→3')	Temperatura de anelamento
<i>EgCesA2</i>	CesA4.3	AGGCGAGAAGTAGGGACCAA	60°C
<i>EgCesA2</i>	CesA4.4	TGAACTACGGTGGAGTGGTG	60°C
<i>EgCesA3</i>	Euc1.4	CTCAAGCACTTCTGTGCTGC	60°C
<i>EgCesA3</i>	Euc2.3	CCTTACCACTCCTTGCCTACT	60°C
<i>EgCesA3</i>	Euc2.4	CCATCAGCCCTTTCAGGAA	60°C
<i>EgCesA3</i>	CesA72.3	TAAGCATAGTGCCAATGGGG	60°C

3.6 DETECÇÃO DE POLIMORFISMO NO GENE *EgCesA2* DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DO GÊNERO *Eucalyptus*

Para obter informações sobre o polimorfismo do gene *EgCesA2* de celulose sintase em *Eucalyptus*, foram seqüenciados fragmentos desse gene em 12 indivíduos representantes de diferentes espécies. Para isso, o DNA genômico foi submetido à PCR utilizando os iniciadores *CesA4.4* (*forward*) e *CesA4.3* (*reverse*), para amplificação dos fragmentos. O protocolo das reações foi o seguinte: tampão contendo Tris-HCl (10mM, pH 8,4), KCl (50mM) e Triton X-100 (0,1%), BSA (0,25µg/µL), MgCl₂ (1,66mM), dNTPs (0,25 µM), 0,50 pmol de cada iniciador (*forward* e *reverse*); 1U da enzima Taq DNA polimerase (Phoneutria) e 6 ng de DNA molde, para um volume final de reação de 12 µL.

As reações de amplificação seguiram as seguintes etapas: desnaturação a 94°C durante 5 minutos; 45 ciclos compostos de uma etapa de 94°C por 1 minuto, temperatura de anelamento dos iniciadores (60°C) durante 1 minuto, extensão a 72°C por 1 minuto; seguidos de extensão final a 72°C por 10 minutos. Para confirmar a amplificação dos fragmentos, os produtos de PCR foram submetidos à eletroforese em gel de agarose a 1,4%, corados com brometo de etídeo e visualizados com luz UV.

Na etapa de seqüenciamento, para um volume final de 10 µL de reação, foi usado 1 µL do produto da reação de PCR, 1 picomol de *primer* (*forward* ou *reverse*), 2 µL de tampão (Tris-HCl, 200mM, pH9,0; MgCl₂, 0,5mM) e 1 µL de premix DYEnamic™ ET Terminator Cycle Sequencing (Pharmacia Biotech, EUA). Procedeu-se às reações de seqüenciamento através de 30 ciclos de 95°C por 20 segundos e 60°C por 1 minuto e 15 segundos, em termociclador GeneAmp PCR System 9700 (Applied Biosystems, Inc.).

Em seguida o produto da reação foi precipitado adicionando-se 40µL de isopropanol 65%, homogeneizado em um agitador do tipo vórtex, por 1 minuto. A solução foi submetida à incubação a temperatura ambiente, durante 20 minutos, e centrifugada a 2000 rcf por 45 minutos, sendo descartado o sobrenadante. Foram feitas duas lavagens adicionando-se 250µL etanol 60%, seguidas de centrifugação a 2000rcf por 10 minutos e novo descarte do sobrenadante. Para retirar todo o excesso de etanol, a placa foi centrifugada de modo invertido, numa velocidade de 500 rpm por 1 minuto.

O produto da reação de seqüenciamento, já precipitado e livre de traços de etanol, foi ressuspendido em 10 µL de formamida Hi-Di (Applied Biosystems, Inc.). E após desnaturação com 5 minutos a 95°C, e 3 minutos a 4°C, foi submetido à eletroforese capilar em um analisador automático de fragmentos ABI-3100 (Applied Biosystems, Inc.).

3.7 TRIAGEM DE UMA BIBLIOTECA DE BACs EM BUSCA DE CLONES CONTENDO GENES *EgCesA2*

No âmbito do Projeto Genolyptus foi construída uma biblioteca de BAC (cromossomo artificial de bactérias), a partir do genoma de um indivíduo da espécie *E. grandis* (S. Brommonschenkel, com. pess.), composta por 20.160 clones, com tamanho médio de insertos entre 120 a 150Kb, apresentando uma estimativa de cobertura genômica de quatro vezes.

Para fins de identificação de um clone, contendo a versão genômica do gene *EgCesA2*, a biblioteca de BAC foi submetida a uma triagem a partir dos iniciadores

desenhados por Salazar (2006). Para fins de redução de custos e otimização do tempo gasto nesta triagem, os clones contidos em placas de polietileno para PCR com 96 poços foram organizados em um esquema de *pools* e *superpools*. Cada *pool* foi formado por uma placa de BAC com 96 clones, purificados conjuntamente. Cada *superpool* foi constituído por um conjunto de 10 *pools*, perfazendo um total de 21 *superpools* (Figura 7). Essa hierarquização de placas foi desenvolvida por Vaz (2005) e disponibilizada para triagem pelo Laboratório de Biotecnologia da Embrapa Arroz e Feijão.

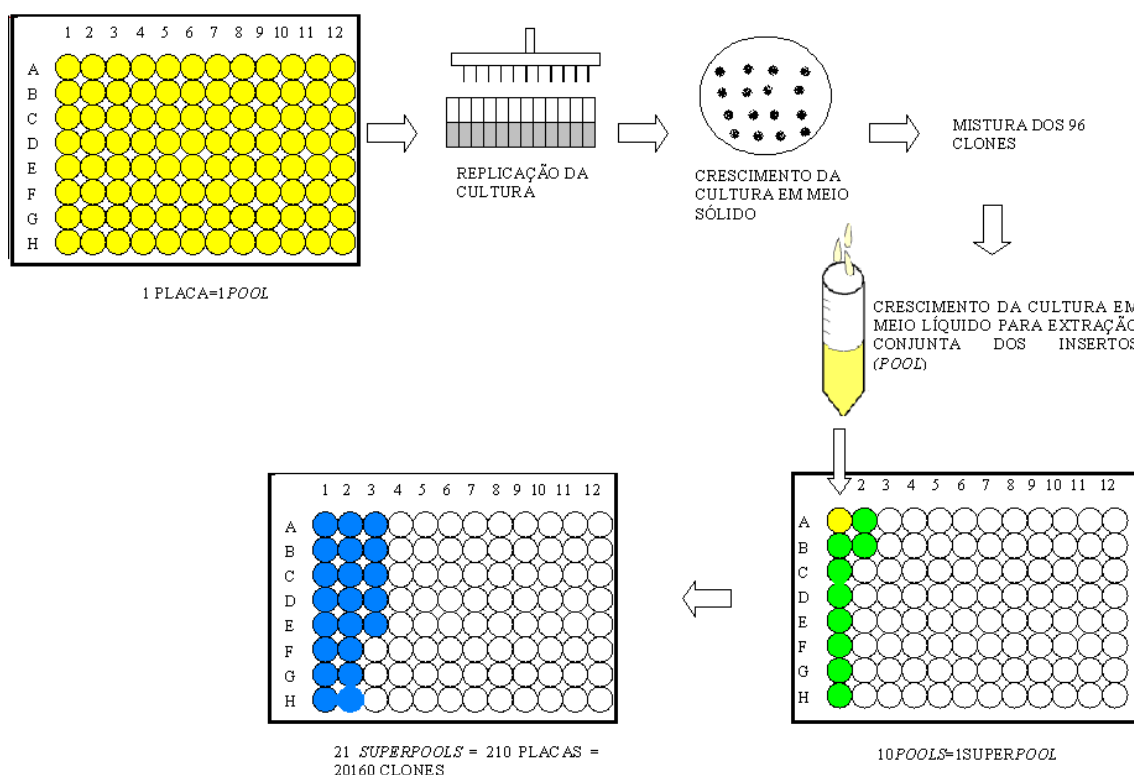


Figura 7. Formação dos *pools* e *superpools* para a triagem da biblioteca de BAC em busca de insertos que contenham os genes *CesA* ortólogos ao gene *EgCesA2* (Fonte: Vaz, 2005).

Para a triagem em busca do gene *EgCesA2*, foram usados os *primers* *CesA4.4* (*forward*) e *CesA4.3* (*reverse*). A triagem foi realizada através da realização de PCR, primeiramente nos 21 *superpools*. Foram feitas 22 reações para o par de *primers*, 21 utilizando o DNA dos *superpools* de clones BAC e uma reação controle usando o DNA genômico do indivíduo da espécie *E. grandis*, utilizado para a construção da biblioteca de BACs.

O protocolo para a triagem de *superpools*, *pools* e placas de BACs foi o seguinte: 6ng de DNA de BAC, tampão constituído por Tris-HCl (10mM, pH 8,4), KCl

(50mM), Triton X-100 (0,1%), 2µg/µL de BSA, MgCl₂ (1,33mM), dNTPs (0,16 µM); 2,50 pmoles de cada primer (*forward* e *reverse*) e 0,6U da enzima Taq DNA polimerase (Phonutria), para um volume final de 15 µL. As reações de amplificação foram feitas através dos seguintes passos: desnaturação a 94°C durante 5 minutos; 30 ciclos compostos de um passo de 94°C por 1 minuto, temperatura de anelamento dos primers (56°C) durante 1 minuto, extensão a 72°C por 1 minuto; seguidos de um passo final de extensão a 72°C por 10 minutos. A visualização dos resultados foi realizada através de uma eletroforese em gel de agarose (1,4%) e a estimativa do tamanho do inserto foi obtida através da comparação do *superpool* amplificado com o marcador molecular 1Kb Plus DNA Ladder (Invitrogen, Carlsbad, CA).

Uma vez identificados os *superpools* que continham os clones de interesse foi feita uma nova triagem dentro dos 10 *pools*, correspondentes a cada *superpool* positivo, e após a identificação dos *pools* positivos, foi feita uma triagem para identificação da posição das placas que continham o clone BAC de interesse, contendo o inserto relativo ao gene *EgCesA2*.

3.8 DETECÇÃO DE POLIMORFISMO NO GENE *EgCesA3* PARA O ESTUDO DO DESEQUILÍBRIO DE LIGAÇÃO EM *E. urophylla*

Para se detectar o polimorfismo dentro do gene *EgCesA3*, na população de *E. urophylla*, procedeu-se à amplificação dos fragmentos do gene, utilizando-se na reação de polimerase em cadeia o par de *primers* formado por Euc2.3 (*foward*) e Euc1.4 (*reverse*), ou o par Cesa72.3 (*foward*) e Euc2.4 (*reverse*), de acordo com o protocolo: tampão contendo Tris-HCl (10mM, pH 8,4), KCl (50mM) e Triton X-100 (0,1%), BSA (0,25µg/µL), MgCl₂ (1,66mM), dNTPs (0,25 µM), 2,5 pmol de cada iniciador (*forward* e *reverse*), 1U da enzima Taq DNA polimerase (Phonutria) e 6 ng de DNA molde, para um volume final de reação de 25 µL. Procedeu-se às reações de acordo com as seguintes etapas: desnaturação a 94°C durante 5 minutos, 45 ciclos compostos de uma etapa de 94°C por 1 minuto, temperatura de anelamento dos iniciadores (60°C) durante 1 minuto, extensão a 72°C por 1 minuto, seguidos de extensão final a 72°C por 10 minutos. Para confirmar a amplificação dos fragmentos, os produtos de PCR foram submetidos à eletroforese em gel de agarose a 1,4%, corados com brometo de etídeo e visualizados com luz UV.

Antes de prosseguir à etapa de seqüenciamento, os produtos de PCR foram purificados adicionando-se 40µL de isopropanol 65%, homogeneizando-se em um agitador do tipo vórtex, durante 1 minuto, incubando-se à temperatura ambiente, durante 20 minutos e submetendo-se a centrifugação (velocidade de 2000 rcf) por 45 minutos, sendo descartado o sobrenadante. Foram feitas duas lavagens adicionando-se 250µL etanol 60%, seguidas de centrifugação a 2000 rcf por 10 minutos, e novo descarte do sobrenadante. Para retirar todo o excesso de etanol, a placa foi centrifugada de modo invertido, numa velocidade de 500 rpm por 1 minuto e ressuspendeu-se os fragmentos amplificados em 10 µL de água ultrapura.

Na etapa de seqüenciamento, para um volume final de 10 µL de reação, foi usado 2,5 µL do produto da reação de PCR purificado, 1 picomol do *primer* (*forward* ou *reverse*), 2 µL de tampão (Tris-HCl, 200mM, pH9,0; MgCl₂, 0,5mM) e 1 µL do kit Big Dye™ Terminator v3.0 Ready Reaction Cycle Sequencing (Applied Biosystems, Inc.). As reações de seqüenciamento foram realizadas através de 30 ciclos de 95°C, por 10 segundos, e 60°C por quatro minutos e cinco segundos, em termociclador GeneAmp PCR System 9700 (Applied Biosystems, Inc.).

Após isso, o produto da reação foi precipitado adicionando-se 40 µL de isopropanol 65%, homogeneizado, sendo submetido à incubação a temperatura ambiente durante 15 minutos e centrifugado a 3700 rpm por 45 minutos, e, em seguida, descartado o sobrenadante. Foram feitas duas lavagens adicionando-se, respectivamente, 200 µL de etanol 60% e 200 µL de etanol 70%, seguidas de centrifugação a 3700 rpm por 10 minutos e novo descarte do sobrenadante. Para retirar todo o excesso de etanol, a placa foi centrifugada de modo invertido, numa velocidade de 900 rpm por 10 segundos.

O produto, já precipitado e livre de traços de etanol, foi ressuspendido em 10 µL de formamida Hi-Di (Applied Biosystems, Inc.), desnaturado a 95°C, durante 5 minutos, com o auxílio de um termociclador GeneAmp PCR System 9700 (Applied Biosystems, Inc.) e incubado no gelo, durante 2 minutos. Em seguida, foi submetido à eletroforese Capilar em um analisador automático de fragmentos ABI-3100 (Applied Biosystems, Inc.).

3.9 ANÁLISE DAS SEQUÊNCIAS OBTIDAS E DETECÇÃO DO DESEQUILÍBRIO DE LIGAÇÃO

Todas as seqüências de nucleotídeos obtidas foram submetidas à análise de qualidade pelo programa PHRED (Ewing et al., 1998), sendo estabelecido um valor mínimo igual a 15. Após isso, para formação dos *contigs*, que correspondem a um conjunto de seqüências de DNA sobrepostas, e detecção dos prováveis SNPs, as seqüências foram analisadas pelo programa *Staden Package* (Staden et al., 2000). Os sítios polimórficos foram caracterizados pela natureza da mutação (substituição ou *indel*) e região de ocorrência (íntron e éxon). Além disso, a detecção da seqüência de aminoácidos que é codificada pelo *contig* foi obtida pela análise no programa GeneRunner (disponível em <http://www.generunner.com>). A fase gamética dos haplótipos, através do algoritmo EM, e a magnitude do desequilíbrio de ligação entre os sítios polimórficos foram estimados com o auxílio do programa Arlequin versão 3.1 (Excoffier, et al., 2006). Os índices de diversidade nucleotídica π e θ_w foram estimados utilizando-se o programa DnaSP (Rozas & Rozas, 1999). Análises semelhantes foram feitas para o gene *EgCesA3*, com base nos dados de seqüenciamento obtidos para a população de *E. urophylla*. Para estimar a magnitude do desequilíbrio de ligação nesse gene foram utilizados os programas Arlequin versão 3.1 (Excoffier, et al., 2006), e Haploview ver. 3.32 (Barret et al., 2005).

3.10 ANÁLISE DOS NÍVEIS DE EXPRESSÃO DOS GENES DE CELULOSE SINTASE EM *Eucalyptus*

Os níveis de expressão dos seis genes conhecidos de celulose sintase em *Eucalyptus* foram avaliados pela utilização de duas estratégias: i. “*Digital Differential Display*”, através da análise do banco de dados de ESTs do projeto Genolyptus e ii. *microarrays*, pela análise dos resultados de um ensaio com 10 *arrays* realizado também no âmbito do projeto Genolyptus.

Para analisar a expressão diferencial dos seis genes de celulose sintase de *Eucalyptus* em diferentes tecidos e espécies, foi utilizada a abordagem denominada “*Digital Differential Display*”. Para isso, as seqüências codantes (cds) dos seis genes *EgCesA* de *Eucalyptus* depositadas no GenBank (Tabela 4), correspondentes a diferentes subunidades de genes de celulose sintase, foram comparadas às seqüências de 93.252 *reads* de ESTs depositadas no banco de dados do Projeto Genolyptus, com o uso do algoritmo

blastn. As seqüências que apresentaram um e-value $\leq 10^{-10}$ foram selecionadas. Cada uma das seqüências selecionadas foi então submetida a uma nova comparação com as seqüências dos seis genes *EgCesA*, também utilizando-se o algoritmo Blastn, para fins de classificação das mesmas entre as seis formas parálogas conhecidas, ou seja, entre os seis genes que se diferenciaram após eventos de duplicação gênica.

Tabela 4. Genes *EgCesA* de *Eucalyptus* usados como referência para comparações com os banco de ESTs do Projeto Genolyptus.

Gene	Nº de acesso (GenBank)
<i>EgCesA1</i>	gi/67003906
<i>EgCesA2</i>	gi/67003908
<i>EgCesA3</i>	gi/67003910
<i>EgCesA4</i>	gi/67003912
<i>EgCesA5</i>	gi/67003914
<i>EgCesA6</i>	gi/67003916

Com os ESTs classificados em função dos diferentes grupos de ortologia, foi feita a comparação da frequência de ocorrência de produtos de cada um dos genes *EgCesA* nas diferentes bibliotecas construídas a partir de mRNAs de tecidos de xilema, floema, folhas maduras, plântulas, em diferentes tratamentos e espécies do gênero *Eucalyptus*. A significância estatística das diferenças entre as frequências observadas de ESTs ortólogos a cada um dos genes de celulose sintase nas diferentes bibliotecas foi avaliada conforme preconizado por Audic & Claverie (1997), com o uso do programa AC-Calc (Coelho, 2006).

Tabela 5. Relação das bibliotecas de cDNA utilizadas.

Biblioteca	Espécie	Tecido	Tamanho da Biblioteca
EUGL-XY	<i>E. globulus</i>	Xilema	15703
EUGR-ML	<i>E. grandis</i>	Folhas maduras	5934
EUGR-SE	<i>E. grandis</i>	Plântula	10690
EUGR-XY	<i>E. grandis</i>	Xilema	8463
EUGR-TS	<i>E. grandis</i>	Plantas submetidas à estresse	12884
EUPE-XY	<i>E. pellita</i>	Xilema	9755
EUSP-FX	<i>E. spp.</i>	Floema	12248
EUUR-XY	<i>E. urophylla</i>	Xilema	9500
EUGR-PU	<i>E. grandis</i>	Folhas infectadas com <i>Puccinia</i>	8075

O estudo de expressão dos genes de celulose sintase em *Eucalyptus* também foi realizado utilizando-se a estratégia de *microarrays*. Para tanto, foram avaliados os resultados de um experimento conduzido no âmbito do projeto Genolyptus envolvendo o estudo da expressão de 21.442 genes hipotéticos em 10 condições diferentes.

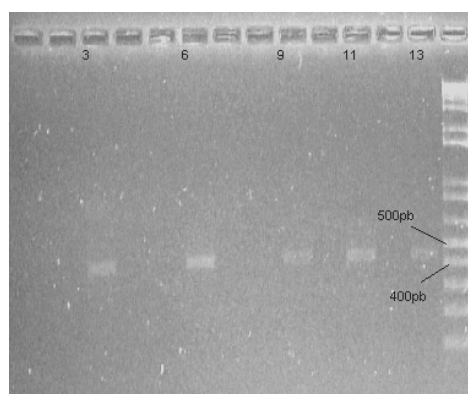
No âmbito do projeto Genolyptus, a partir do banco de ESTs, foram montados 21.442 *contigs*. Para cada um dos *contigs* foram construídas 9 sondas de 60pb cada, dispostas em duplicata no *microarray*. Foram avaliadas amostras de RNA obtidas do xilema de duas espécies de *Eucalyptus*: *E. globulus* e *E. grandis*, representadas por dois clones cada, de três anos de idade, com dois *ramets* representando cada clone. Adicionalmente, foram ainda avaliadas amostras de RNA obtidas de folhas maduras de um dos *ramets* de *E. grandis*. As reações de hibridização, bem como a digitalização das imagens e processamento inicial dos dados foram realizados pela empresa *Nimblegen*.

Na busca realizada nos resultados de anotação preliminar dos *contigs* utilizados para a construção do *array*, feita no sistema de Bioinformática do projeto Genolyptus, Laboratório de Bioinformática da Universidade Católica de Brasília-DF, foram identificados 32 *contigs* relacionados a genes de celulose sintase. A identidade das 9 sondas de cada um dos 32 *contigs* anotados como de genes de celulose sintase foi verificada pela comparação com as seqüências dos seis genes de *EgCesA*. Somente os resultados relativos aos *contigs* consistentemente identificados como correspondentes a uma das seis formas parálogas foram utilizados para fins de análise de expressão. A comparação dos níveis de expressão gênica observados entre diferentes amostras de tecidos e espécies foi realizada pela utilização do software MEV versão 4.0 (Saeed et al., 2003).

4 RESULTADOS E DISCUSSÃO

4.1 AMPLIFICAÇÃO DOS FRAGMENTOS DO GENE *EgCesA2* EM DIFERENTES ESPÉCIES DE *Eucalyptus*

Antes de proceder ao seqüenciamento dos fragmentos do gene *EgCesA2* nos 12 indivíduos de diferentes espécies de *Eucalyptus*, foi confirmada a amplificação dos produtos de PCR. O fragmento amplificado apresentou um tamanho estimado de 450pb.



(3) Híbrido *E. grandis* x *E. dunnii*; (6) Híbrido *E. urophylla* x *E. globulus*; (9) *E. grandis*;
(11) *E. camaldulensis*; (13) *E. grandis*.

Figura 8. Confirmação da amplificação do fragmento do gene *CesA2* em cinco indivíduos de diferentes espécies de *Eucalyptus*, utilizando-se os iniciadores *CesA4.3* e *CesA4.4*.

4.2 SEQÜENCIAMENTO PARCIAL DO GENE *EgCesA2* DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DO GÊNERO *Eucalyptus*

Os resultados do seqüenciamento dos fragmentos do gene *EgCesA2* em diferentes espécies de *Eucalyptus*, com os iniciadores *CesA4.4* (*foward*) e *CesA4.3* (*reverse*) permitiram a construção de um *contig* de tamanho igual a 427pb. A identidade deste *contig* foi verificada pelo uso do algoritmo Blast, como sendo de fato do gene *EgCesA2* e apresentou uma identidade de 98% em relação à seqüência depositada no

10). O *contig* obtido corresponde à região que codifica três domínios transmembrânicos, observados na extremidade carboxi-terminal da proteína *EgCESA2*.

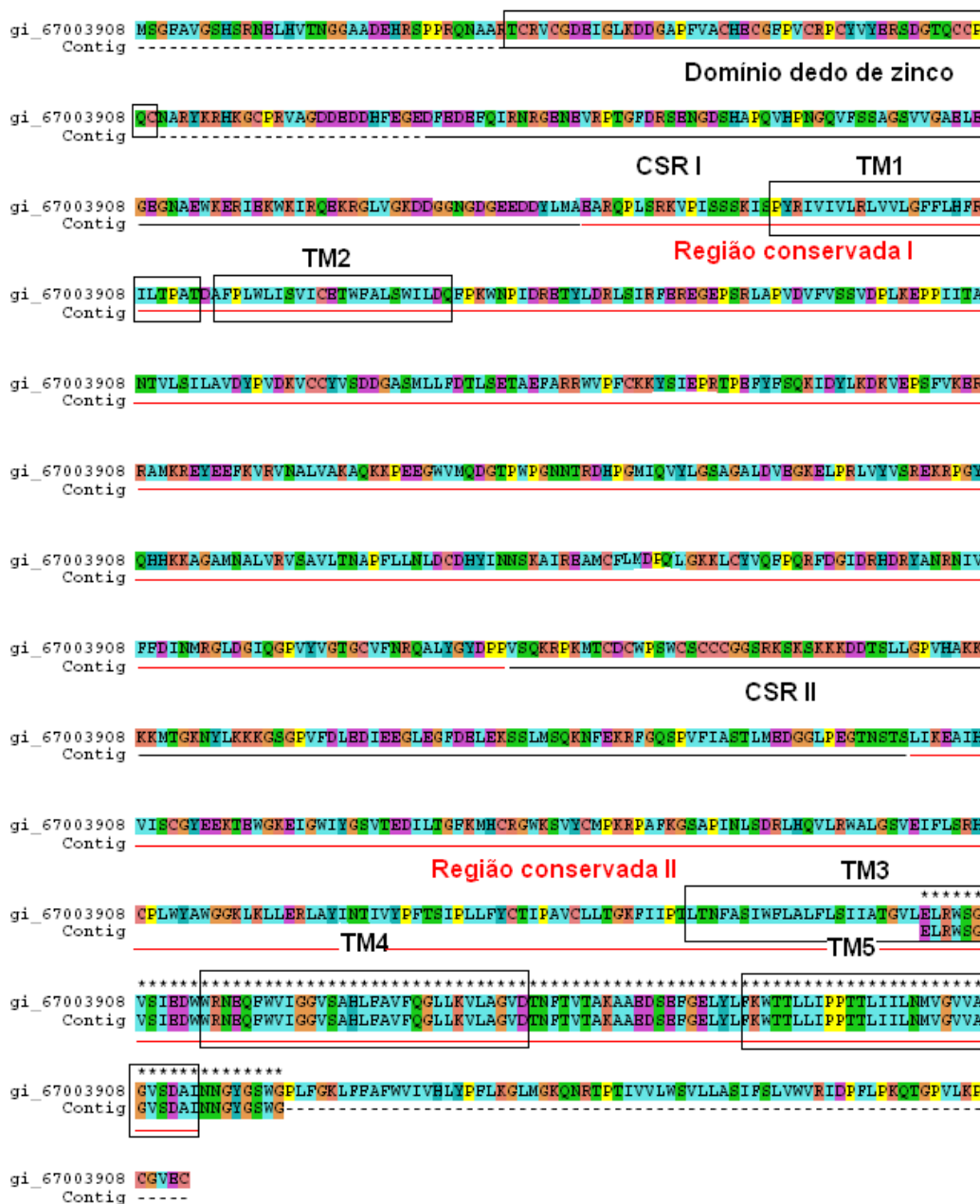


Figura 10. Alinhamento obtido pelo programa Clustal X, evidenciando a seqüência da proteína de celulose sintase, com seus motivos conservados: dedo de zinco; Região classe específica I (CSR I); Região conservada I; Região classe específica II; Região conservada II; e domínios transmembrânicos (TM1-TM5).

4.3 DETECÇÃO DE POLIMORFISMO NO GENE *EgCesA2* EM DIFERENTES ESPÉCIES DO GÊNERO *Eucalyptus*

As estimativas obtidas de diversidade nucleotídica para o fragmento seqüenciado do gene *EgCesA2* estão sumarizadas na Tabela 6. Como anteriormente relatado, o *contig* formado corresponde a um éxon com 427 pares de bases. Foram obtidas estimativas de diversidade nucleotídica através do cálculo do número de sítios polimórficos (S), dos índices π e θ_w , a partir de 38 seqüências originadas de 12 indivíduos de diferentes espécies de *Eucalyptus* (Tabela 6).

Tabela 6. Caracterização do polimorfismo detectado em 427pb do gene *EgCesA2*, em diferentes espécies de *Eucalyptus*.

Parâmetros	Análises para cada sítio
S	3
π (diversidade nucleotídica)	0,00212
θ_w (por sítio)	0,00169

Santos (2005) estimou a diversidade nucleotídica em *Eucalyptus* a partir de um segmento do gene *ccoamt*, relacionado à síntese de lignina. *E. grandis* ($\pi=0,00356$; $\theta_w=0,001666$) apresentou o dobro da diversidade nucleotídica estimada para *E. globulus* ($\pi=,00168$; $\theta_w=0,00093$), considerando-se o índice π , e cerca de 1,4 vezes maior do que o obtido para *E. urophylla* ($\pi=0,00254$; $\theta_w=0,00179$). Embora os valores obtidos tenham divergido em relação a *E. globulus*, os valores dos índices de diversidade são próximos aos obtidos para o gene *EgCesA2*, sobretudo o valor de π obtido para *E. urophylla* e o valor de θ de *E. grandis*. Entretanto não é possível tirar conclusões gerais sobre esses valores, se são significativamente concordantes, ou apenas coincidência, pois constituem-se de estimativas preliminares.

Em um estudo feito com o gene ortólogo à *EgCesA2*, em *Pinus pinaster* e *Pinus radiata*, Pot et al. (2005) encontraram um índice de diversidade, para a região codante, de $\pi = 0,00019$ e $0,00101$, respectivamente. Eles estimaram também o θ_w e estimaram um valor de $0,00058$ para *P. pinaster* e $0,00064$ para *P. radiata*. As estimativas obtidas para o gene *EgCesA2* em *Eucalyptus* (Tabela 6) sugerem a existência de uma diversidade nucleotídica bem maior neste gênero, comparado ao que foi verificado em

Pinus. Contudo, devem-se ressaltar as diferentes características das populações estudadas. As estimativas feitas com *Eucalyptus* foram obtidas a partir de uma ampla amostragem, de indivíduos de diferentes espécies, distribuídos em diferentes regiões geográficas, que experimentaram diferentes pressões evolutivas ao longo das gerações.

Os resultados obtidos pela utilização do programa *Staden Package* confirmaram a ocorrência de três SNPs, também detectados através da inspeção visual dos eletroferogramas, com o auxílio do módulo *Trace Difference* (Figuras 12 a 14). Cada base polimórfica detectada apresentou um valor de *phred* mínimo de 15. Os SNPs foram nomeados a partir da primeira base do *contig* (Figura 11). No estudo realizado por Pot et al. (2005), em relação a um ortólogo do gene *EgCesA2* em *Pinus*, coincidentemente também foram encontrados 3 SNPs na região exônica de 396pb. Todos os SNPs detectados neste estudo correspondem a substituições sinônimas.

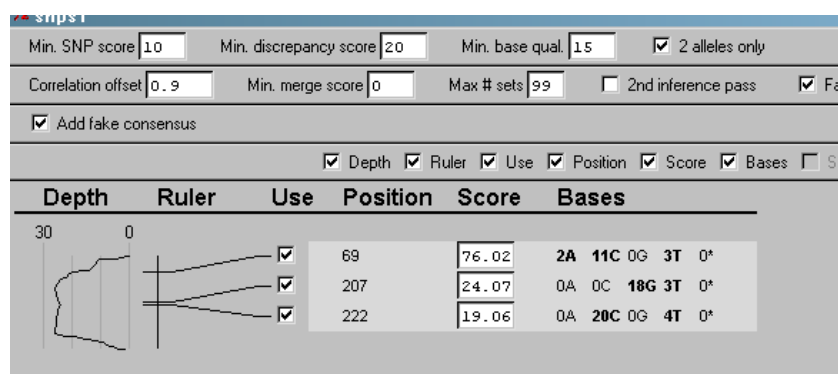
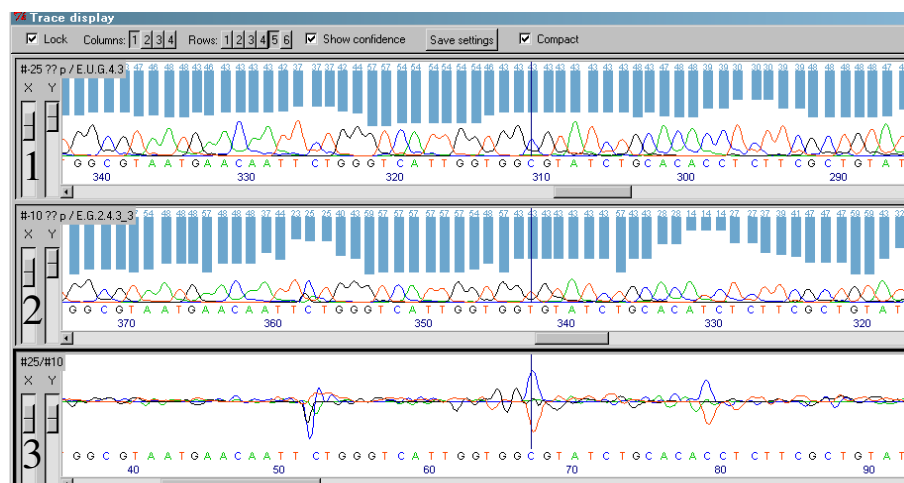


Figura 11. Localização dos três SNPs encontrados pelo programa *Staden Package* na região flanqueada pelos *primers* *CesA4.4* e *CesA4.3*.

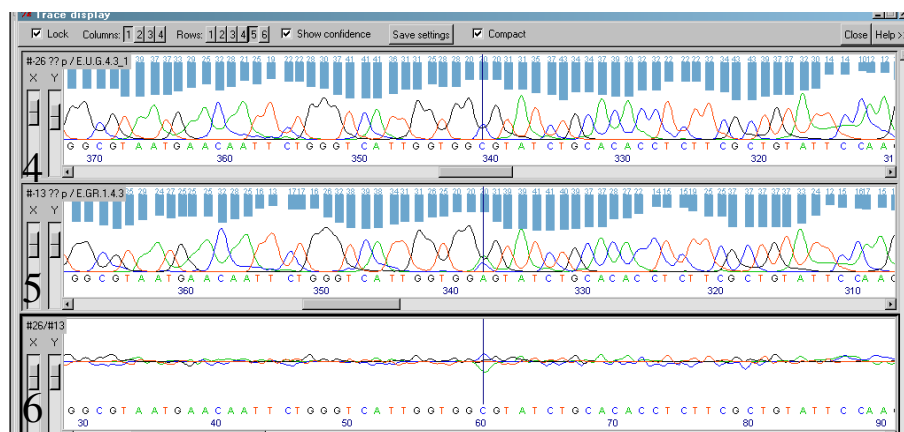
O SNP encontrado na posição 69 do *contig* é trialélico. Trata-se de uma transição de citosina para timina, e de uma transversão, no caso da substituição de uma citosina para adenina. Ressalta-se pela inspeção do eletroferograma (Figura 12), que a transversão C/A foi, na verdade, observada pelo seqüenciamento de um indivíduo heterozigoto para o gene *EgCesA2*. Tal fato permitirá a localização no mapa genético do loco correspondente ao gene *EgCesA2*, utilizando-se como parental o indivíduo heterozigoto observado. De acordo com a ORF encontrada para essa seqüência no programa Gene Runner, tanto a transição C/T, quanto a transversão C/A não alteram o aminoácido encontrado neste ponto da proteína, uma glicina, que devido à redundância do código genético é codificada pelos códons GGT, GGC e GGA.

O SNP observado na posição 207 do *contig* originou-se de uma transversão de uma guanina para uma timina. Contudo, também gera uma mutação sinônima, pois o aminoácido em questão, uma leucina, é igualmente codificado pelas trincas CTG e CTT.

O SNP na posição 207 também gera uma mutação sinônima, a partir da substituição de uma citosina para timina (transição), já que ambos os códon, ACC e ACT, codificam para treonina.



A



B

Figura 12. Eletroferogramas representativos da região que contém o SNP da posição 69 do *contig*. **A:** Relaciona-se às comparações entre a substituição de uma citosina (1) para uma timina(2). **B:** Trata-se da substituição de uma citosina (4) para uma adenina (5). É possível notar também que o SNP em destaque no quadro (5) corresponde a uma região heterozigota do gene. Os quadros (3) e (6) evidenciam o polimorfismo detectado pelo módulo *Trace difference* do programa Standen Package.

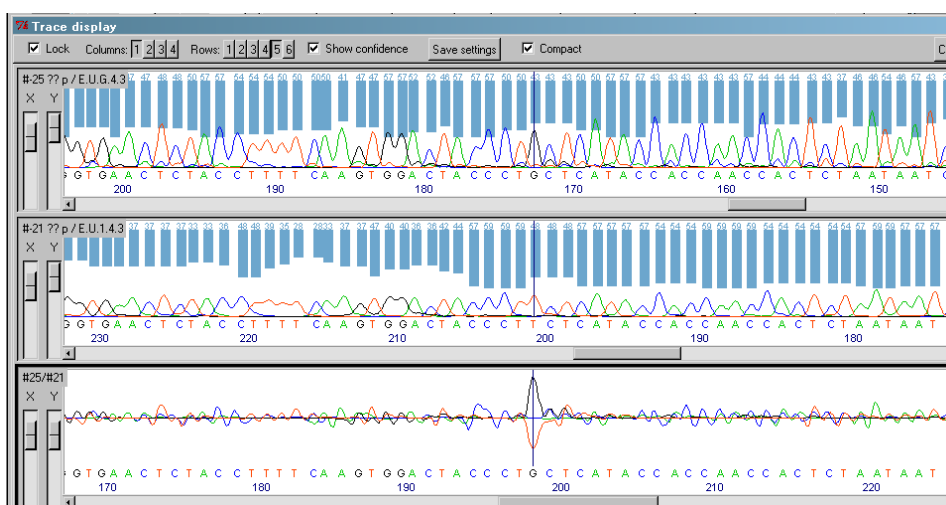


Figura 13. Eletroferogramas representativos da região que contém o SNP da posição 207 do *contig*.

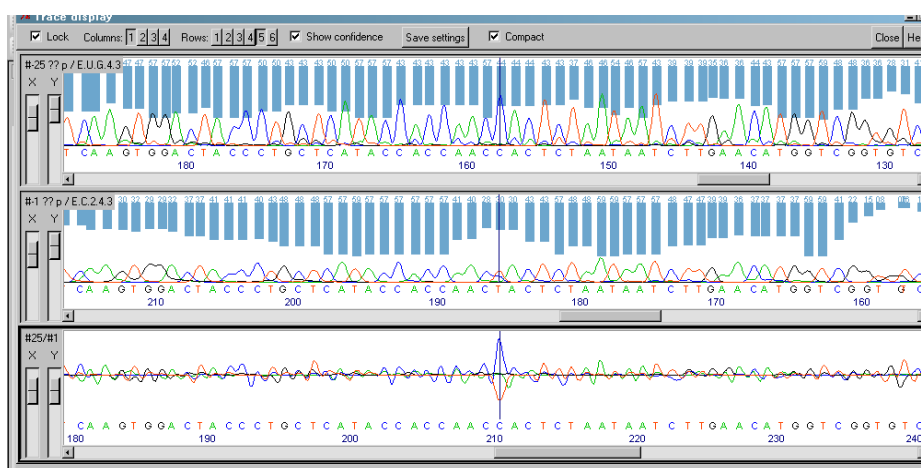


Figura 14. Eletroferogramas representativos da região que contém o SNP da posição 222 do *contig*.

Somente um percentual pequeno do genoma de eucariotos é formado por genes. A grande maioria dos SNPs são encontrados em regiões silenciosas do genoma, e têm efeitos, a maioria das vezes, imperceptíveis no fenótipo. Outros SNPs, embora estejam localizados dentro de regiões gênicas que são expressas, não causam alterações na composição de aminoácidos da cadeia protéica.

Todos os SNPs encontrados dentro do gene *EgCesA2* neste estudo são sinônimos. Porém, o conhecimento da localização desses SNPs permite que os mesmos sejam usados como marcadores em estudos de mapeamento genético. Além disso, sabe-se que SNPs sinônimos detectados em diferentes genes humanos podem de alguma forma

modificar a estrutura e a estabilidade do mRNA produzido, interferindo no nível da expressão gênica, (Shen et al., 1999; Duan et al., 2003). Sendo assim, algum mecanismo análogo poderá ser detectado futuramente em genes de *Eucalyptus* que apresentam esse tipo de polimorfismo.

Cada SNP encontrado foi privativo a uma única espécie e procedência. O SNP de posição 69 do *contig* foi único para *E. globulus* (procedência 2). O polimorfismo de posição 207 foi encontrado somente em *E. urophylla* (Timor), enquanto o SNP 222 foi privativo de *E. camaldulensis* (procedência 2).

Estudos futuros enfocando a importância de polimorfismos desse tipo dentro de genes de celulose sintase possibilitarão encontrar mais respostas ao enigma de como a síntese desse carboidrato é controlada.

4.4 ANÁLISE DE DESEQUILÍBRIO DE LIGAÇÃO NO GENE *EgCesA2*

Os haplótipos estimados pelo algoritmo EM estão apresentados na Tabela 7. Os níveis de significância das medidas de desequilíbrio de ligação obtidas por randomização (10.000 permutações) estão apresentados na Tabela 8. Ao nível de significância de 0,05, somente o par formado pelos SNPs de posição 69 e 207 do *contig* estão em desequilíbrio de ligação.

Tabela 7. Freqüência dos haplótipos encontrados na população de *Eucalyptus*.

Haplótipos			Freqüência
SNP			
69	207	222	
C	G	T	0,079
C	G	C	0,500
T	G	C	0,263
A	T	C	0,053
C	T	C	0,105

Tabela 8. Significância do desequilíbrio ao nível de 0,05.

SNPs	p-valor
69/207	0,0223
69/222	0,3020
207/222	0,5260

O resultado da análise do desequilíbrio de ligação entre os sítios 69 e 207, obtida admitindo-se como verdadeiros os haplótipos estimados (Tabela 7), está apresentado na Tabela 9. É interessante notar que o desequilíbrio se estende por 138pb. Existem poucos estudos sobre o padrão de extensão do DL em espécies como o *Eucalyptus*. Thumma et al. (2005) fizeram análises de desequilíbrio de ligação em um gene que determina o ângulo de microfibrilas (MFA), uma característica importante para a rigidez da madeira, e detectaram SNPs altamente correlacionados com a característica numa extensão média de 300pb.

Tabela 9. Valores do DL entre os SNPs 69 e 207 estimados por D, D' e r^2 .

69/207	G		T
	D	D'	r^2
C	0,003	0,026	0,0003
T	0,042	1,000	0,0670
A	0,044	1,000	0,2963

4.5 TRIAGEM DA BIBLIOTECA DE BACs

A triagem da biblioteca de BACs e as subseqüentes confirmações da amplificação via PCR dos clones contendo insertos de interesse, permitiram o isolamento de um clone BAC para o gene *EgCesA2* (Tabela 10). Dispondo-se deste clone, torna-se viável a caracterização minuciosa das regiões correspondentes ao promotor do gene, de grande importância para o conhecimento dos aspectos relacionados à regulação de sua expressão, bem como de seus íntrons.

Tabela 10. Resultado da triagem da biblioteca de BAC para o gene *EgCesA2*.

Gene	Superpool	Pool	Placa	Clone
<i>EgCesA2</i>	1	9	87	E4

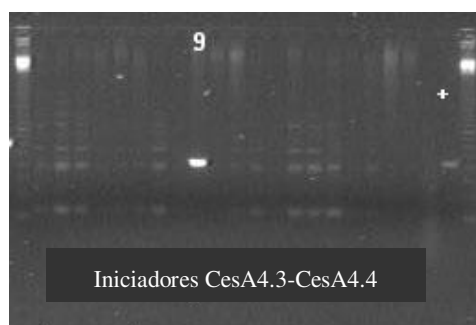


Figura 15. Confirmação da amplificação do fragmento do gene *EgCesA2* no *pool* 9. “+” indica a amplificação do fragmento no indivíduo de *E. grandis* (controle positivo).

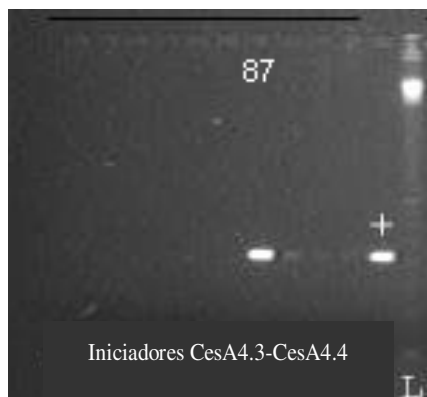


Figura 16. Reação de amplificação do fragmento do gene *EgCesA2*, dentro do *pool* 9, revelando que a placa que contém o gene de interesse é a EUGR BC 002 087. “+” indica a amplificação do fragmento no indivíduo de *E. grandis* (controle positivo). L (marcador molecular ladder 1Kb Plus).

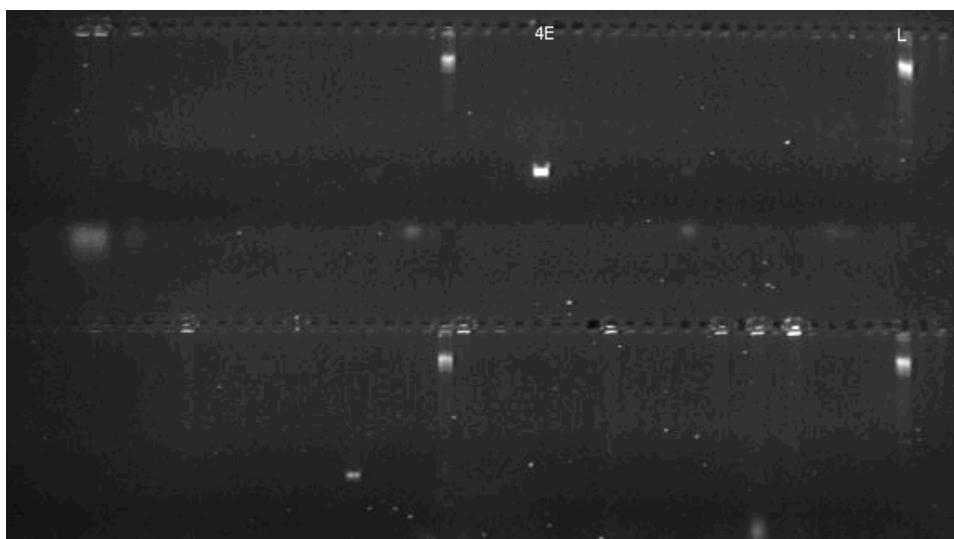


Figura 17. Reação de amplificação do fragmento do gene *EgCesA2* revelando que o clone E4 da placa EUGR BC 002 087 contém o gene de interesse. L (marcador molecular ladder 1Kb Plus).

4.6 AMPLIFICAÇÃO DE FRAGMENTOS DO GENE *EgCesA3* PARA DETECÇÃO DE DESEQUILÍBRIO DE LIGAÇÃO EM *E. urophylla*

Antes de proceder ao seqüenciamento dos fragmentos do gene *EgCesA3* nos 32 indivíduos de diferentes espécies de *Eucalyptus*, foi confirmada a amplificação dos produtos de PCR. O par de *primers* composto por Euc2.3 (*forward*) e Euc1.4 (*reverse*)

permitiu a amplificação de um fragmento de DNA estimado em 800pb, enquanto o fragmento amplificado a partir dos *primers* CesA72.3 (*forward*) e Euc 2.4 (*reverse*) apresentou um tamanho de 1,6 kb (Figura 18).

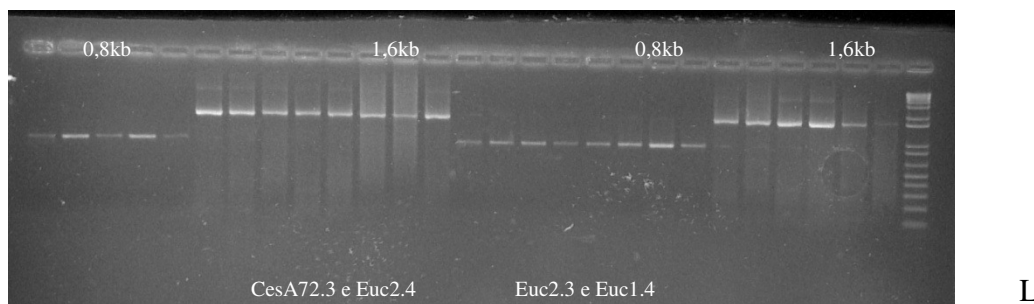


Figura 18. Reação de amplificação do fragmento do gene *CesA3*, gerou de fragmentos de 800pb e 1,6kb, utilizando-se, respectivamente, os pares de *primers* Euc2.3 (*forward*) e Euc1.4 (*reverse*), ou CesA72.3 (*forward*) e Euc2.4 (*reverse*). L (marcador molecular ladder 1Kb Plus).

4.7 SEQÜENCIAMENTO PARCIAL DO GENE *EgCesA3* DE CELULOSE SINTASE EM UMA POPULAÇÃO DE *E. urophylla*

O seqüenciamento parcial do gene *EgCesA3* nos indivíduos da população de *E. urophylla* possibilitou a formação de um *contig* com 770pb. A identidade do fragmento foi confirmada através da análise Blast como se tratando, como esperado, do gene *EgCesA3*, ortólogo ao gene *AtCesA7* de *Arabidopsis thaliana* (Ranik & Myburg, 2006). Utilizando o programa EST2GENOME (Mott, 1997) foi possível perceber que o *contig* é formado por dois éxons, um com 33pb e outro com 582pb, flanqueando um íntron de 95pb, além de uma região 3'-UTR de 60pb (Figura 19).

A análise de detecção de SNPs, realizada pelo programa *Standen Package*, identificou a presença de nove sítios polimórficos, em diferentes regiões do gene, em *E. urophylla*. A presença de cada sítio polimórfico foi confirmada pela inspeção visual dos eletroferogramas, utilizando o módulo *Trace difference*.

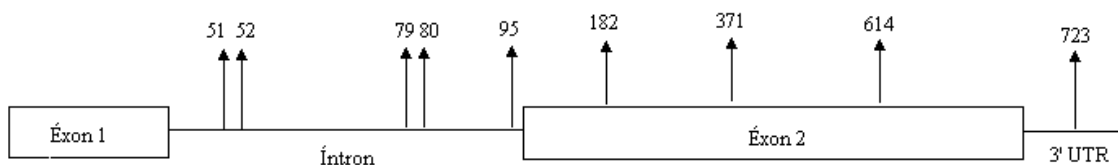


Figura 20. Posição dos SNPs dentro do *contig* formado por fragmentos de seqüências do gene *CesA3*.

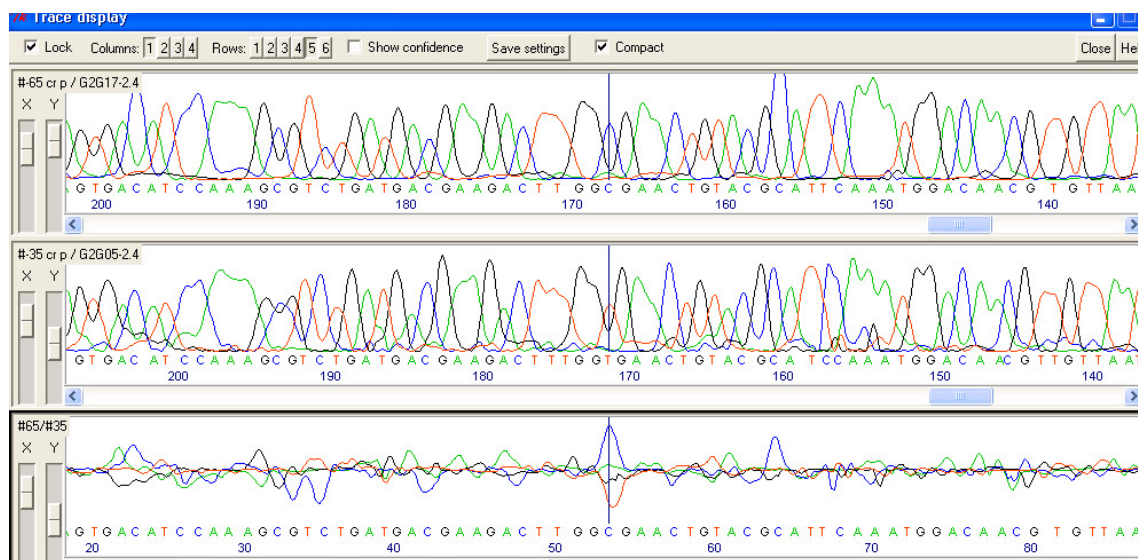


Figura 21. Eletroferograma mostrando o SNPs 371 do *contig*, dentro de uma região exônica do gene *CesA3*.

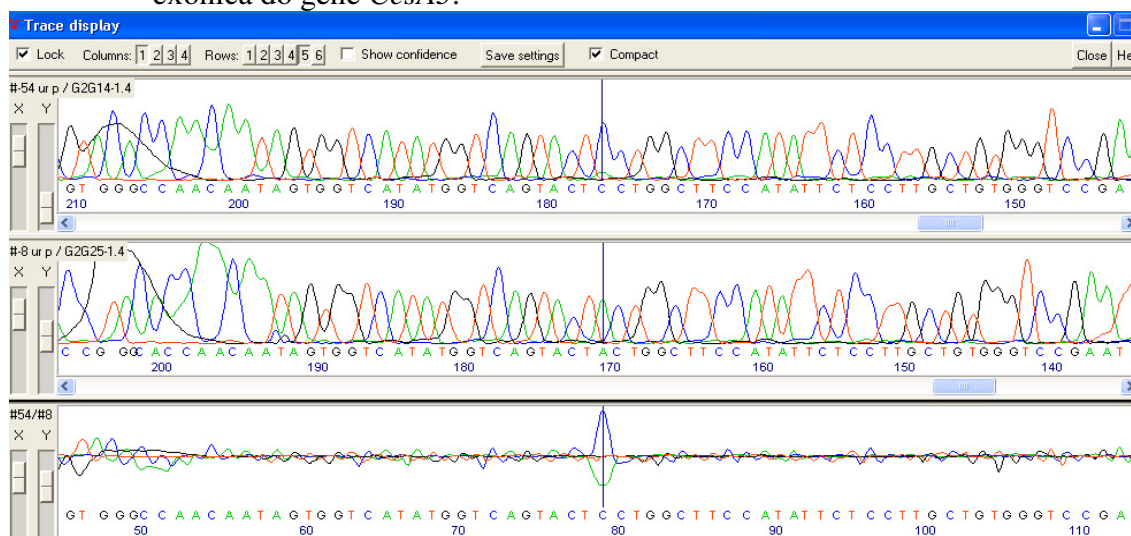


Figura 22. Eletroferograma mostrando o SNPs 614 do *contig*, dentro de uma região exônica do gene *CesA3*.

A análise dos SNPs encontrados no éxon 2, em termos de efeitos sobre a seqüência de aminoácidos codificada, evidenciou que todas as substituições encontradas são sinônimas, ou seja, codificam o mesmo aminoácido predito pela seqüência ancestral. Entretanto, é importante ressaltar que mesmo não alterando a seqüência protéica, os SNPs sinônimos podem alterar a estrutura e a estabilidade dos mRNA, e acabar afetando a quantidade de proteína produzida (Shen et al., 1999; Duan et al., 2003). Futuros estudos de expressão, utilizando-se técnicas como *Northern blot* e PCR em tempo real, poderiam investigar melhor a importância desse tipo de variação dentro do gene.

Tabela 11. Trinca de nucleotídeos evidenciando os SNPs encontrados e os respectivos aminoácidos codificados.

SNP	Trinca de nucleotídeos	Aminoácido
182	GCA/GCG	alanina
371	GGC/GGT	glicina
614	CTC/CTA	leucina

Para estimar a diversidade nucleotídica foram feitas comparações entre pares de seqüências, conforme sumarizado na Tabela 12. A diversidade nucleotídica (π) estimada para o gene *CesA3* foi menor do que a obtida por Salazar (2006), $\pi = 0,0543$, em um fragmento de 474pb, no mesmo gene de celulose sintase de *Eucalyptus*. Porém o índice de diversidade estimado neste estudo ($\pi = 0,00185$), foi maior do que o obtido por Brown et al. (2004), num estudo efetuado em um gene ortólogo ao gene *EgCesA3*, em *Pinus* ($\pi = 0,00087$).

Tabela 12. Estimativa dos índices de diversidade a partir de 64 seqüências nucleotídicas com 770pb de comprimento, levando-se em conta a informação de vários sítios polimórficos em heterozigose.

Comparações entre pares de seqüências	
S	9
π	0,00185
θ_w	0,0028

Foram feitas estimativas dos índices de diversidade nucleotídica levando em conta as diferentes regiões do *contig* (Figura 23). É esperado que em regiões não transcritas a diversidade nucleotídica seja maior do que nos éxons, porque tais locos não

sofrem altas pressões de seleção, exceto em regiões promotoras e envolvidas no processo de *splicing*. Os valores estimados para π revelam uma magnitude de diversidade de 7,5 vezes maior para o íntron, em relação ao éxon. A região 3'-UTR apresentou uma estimativa de 1,6 vezes maior em relação ao éxon.

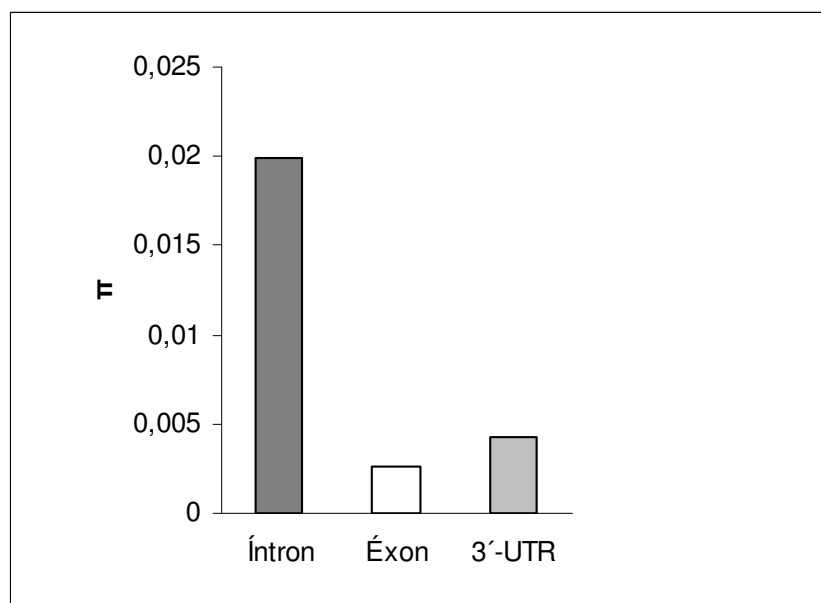


Figura 23. Diversidade nucleotídica em diferentes regiões do gene *Cesa3*.

4.8 ESTIMATIVA DO DESEQUILÍBRIO DE LIGAÇÃO DENTRO DO GENE *EgCesa3* EM UMA POPULAÇÃO DE *E. urophylla*

O desequilíbrio de ligação é sempre mensurado usando estatísticas que façam associação entre diferentes estados alélicos de pares de locos. Em geral, os SNPs são bialélicos, mas quando se avalia uma população natural, mais de dois alelos por loco podem ser amostrados e utilizados para se estimar o DL. Além disso, o problema de se estimar o desequilíbrio de ligação entre pares de locos, cada um com múltiplos alelos, é particularmente dificultado quando são avaliados indivíduos heterozigotos, como é o caso do *Eucalyptus*, uma espécie altamente alógama. Sob estas condições, a fase haplotípica dos SNPs é desconhecida e precisa ser estimada. Na amostra de 32 indivíduos de *E. urophylla* empregou-se o algoritmo EM para estimar as frequências haplotípicas. De posse dos haplótipos, estimou-se o valor do DL entre os locos. Segundo Weir (1996), o algoritmo EM é muito útil quando há dados incompletos. Houve um número razoável de observações perdidas, e sendo assim, não foi possível obter qual seria o nucleotídeo existente nos locos polimórficos avaliados para todos os indivíduos amostrados. Apesar do número pequeno

de observações, o p-valor estimado para cada par de locos avaliados em relação à existência de DL revelou que, num nível de significância de 0,05, a associação foi significativa na maioria dos casos. Somente três pares de SNPs não apresentaram associação consistente (Tabela 13).

Tabela 13. Estimativa da associação entre pares de alelos (p-valor, nível de significância de 0,05).

	51	52	79	80	95	182	371	614	723
51	-								
52	0,00000	-							
79	0,00002	0,00000	-						
80	0,00000	0,00000	0,00000	-					
95	0,00000	0,00000	0,00000	0,00000	-				
182	0,00000	0,00000	0,00000	0,00000	0,00000	-			
371	0,06529	0,03885	0,02676	0,01830	0,01465	0,07465	-		
614	0,01090	0,00000	0,00000	0,00000	0,00228	0,00000	0,0403	-	
723	0,01000	0,00089	0,01515	0,00723	0,00149	0,00356	0,30673	0,00000	-

Os valores estimados para D' e r^2 para todos os pares de sítios estão sumarizados na Tabela 14. Ressalta-se que o SNP 95 foi descartado da análise por ter uma frequência muito baixa na população, na verdade foi detectado somente em um indivíduo. Foram obtidos os valores de LOD para cada par de sítios analisados e foi admitido um valor de LOD mínimo de 2 para que se pudesse sugerir a existência de DL. Observa-se que os SNPs 79 e 80, por exemplo, são fortemente ligados, com um valor de $r^2 = 1$, o que denota um desequilíbrio de ligação completo (Figura 24). Neste caso, observações sobre um marcador permitem a obtenção de informações sobre o outro, de modo que os dois são ditos redundantes.

Tabela 14. Valores do DL entre pares de locos polimórficos, intervalo de confiança para D' e distância entre os locos.

L1	L2	D'	IC _{95%}		LOD	r ²	Dist	n*
snp051	snp052	1,000	0,37	1,00	2,07	0,200	1	36
snp051	snp079	0,452	0,11	0,74	0,92	0,118	28	38
snp051	snp080	0,452	0,11	0,74	0,92	0,118	29	38
snp051	snp182	0,734	0,20	0,93	1,21	0,148	131	33
snp051	snp371	0,397	0,04	0,78	0,26	0,061	320	31
snp051	snp614	1,000	0,28	1,00	1,58	0,245	563	23
snp051	snp723	1,000	0,05	0,98	0,23	0,031	672	23
snp052	snp079	1,000	0,63	1,00	4,45	0,415	27	38
snp052	snp080	1,000	0,65	1,00	4,98	0,500	28	36
snp052	snp182	0,867	0,59	0,96	5,36	0,663	130	31
snp052	snp371	0,579	0,24	0,79	1,81	0,296	319	29
snp052	snp614	1,000	0,60	1,00	4,03	0,604	562	26
snp052	snp723	1,000	0,13	0,99	0,87	0,144	671	23
snp079	snp080	1,000	0,85	1,00	11,95	1,000	1	45
snp079	snp182	1,000	0,65	1,00	4,98	0,500	103	36
snp079	snp371	0,667	0,27	0,87	1,85	0,222	292	33
snp079	snp614	1,000	0,60	1,00	4,03	0,604	535	26
snp079	snp723	1,000	0,11	0,99	0,74	0,118	644	23
snp080	snp182	1,000	0,65	1,00	4,98	0,500	102	36
snp080	snp371	0,667	0,28	0,87	1,91	0,242	291	31
snp080	snp614	1,000	0,60	1,00	4,03	0,604	534	26
snp080	snp723	1,000	0,11	0,99	0,74	0,118	643	23
snp182	snp371	0,541	0,19	0,78	1,53	0,224	189	26
snp182	snp614	1,000	0,61	1,00	4,24	0,694	432	23
snp182	snp723	1,000	0,13	0,99	0,87	0,160	541	20
snp371	snp614	0,760	0,28	0,93	1,64	0,400	243	20
snp371	snp723	1,000	0,07	0,98	0,46	0,080	352	17
snp614	snp723	1,000	0,11	0,99	0,72	0,176	109	23

*número de gametas observados

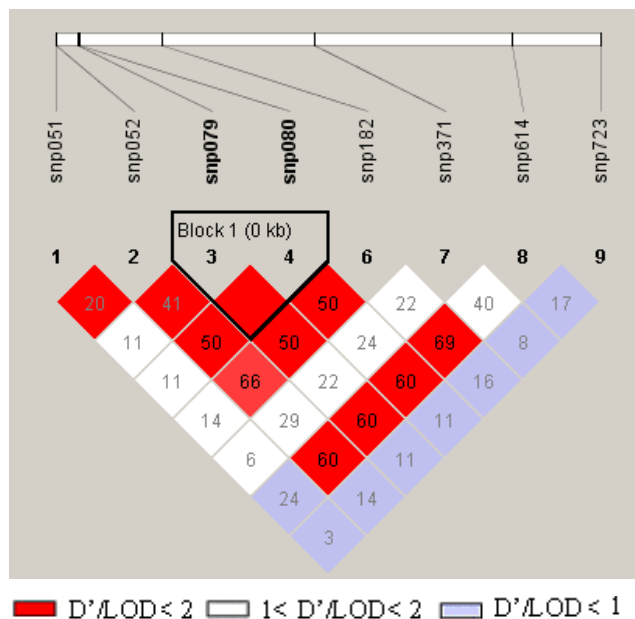


Figura 24. Desequilíbrio de ligação (r^2) entre os marcadores SNPs.

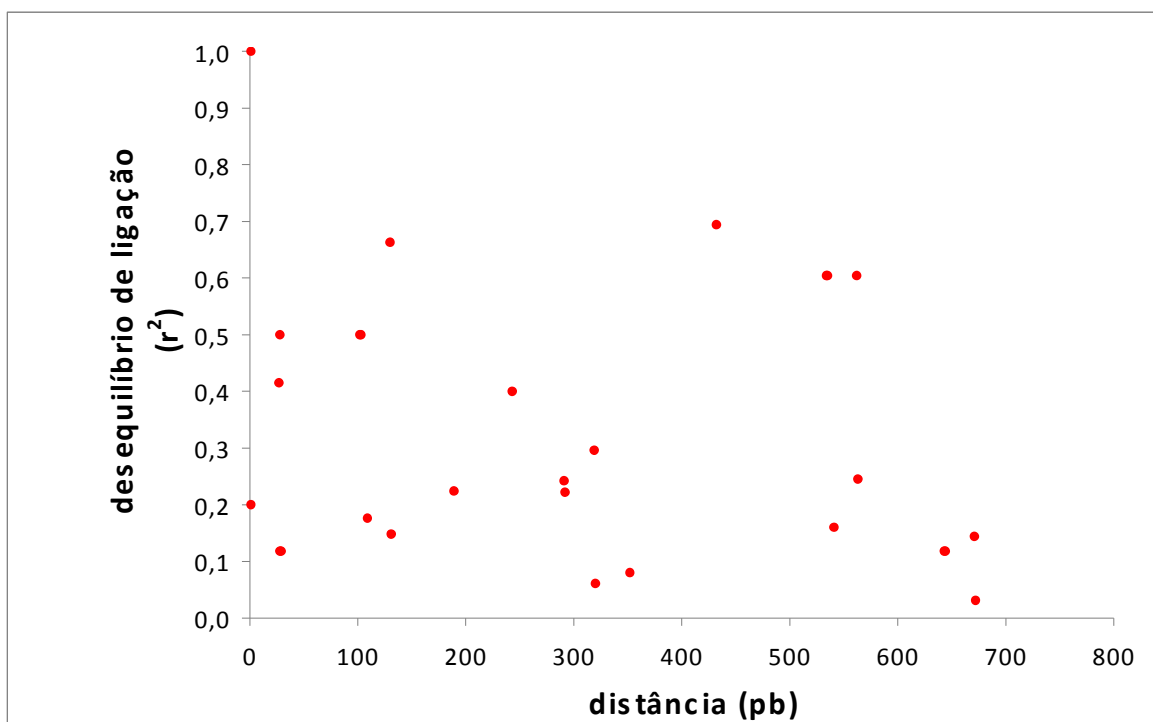


Figura 25. Decaimento do desequilíbrio de ligação (r^2) em função da distância física entre os locos no gene *EgCesA3*, em *E. urphylla*.

O desequilíbrio de ligação ao longo do *contig* é representado por um modelo não linear de decaimento do valor de r^2 em função da distância física no cromossomo (Figura 25). Considerando um valor de r^2 mínimo de 0,3 para que haja desequilíbrio, observa-se que o DL se mantém por uma extensão de aproximadamente 600pb ao longo do

fragmento avaliado do gene *EgCesA3*. O valor de DL para os locos 52 e 614, que estão a uma distância de 562pb, foi estimado em 0,6, com um valor de LOD de 4,03, o que significa que é aproximadamente 10 mil vezes mais verossímil admitir que os sítios estejam em desequilíbrio de ligação (de magnitude igual a 0,6), do que admitir que seus alelos estejam em equilíbrio. Entretanto, os dados levantados não permitem inferências sobre o comportamento do DL além da extensão do segmento estudado, muito menos para o genoma como um todo.

A extensão do desequilíbrio de ligação varia ao longo do genoma e espera-se que fenômenos como gargalo demográfico populacional e sistema de cruzamento afetem o padrão do DL. Em plantas alógamas como o milho, o desequilíbrio tende a manter-se em distâncias menores do que 1kb devido ao fenômeno da recombinação (Remington et al., 2001). Brown et al. (2004), em um estudo a partir de 19 genes analisados ao longo do genoma de *Pinus*, constatou que o DL decaiu de aproximadamente 0,5 para 0,25 com 2000pb. Já em plantas autógamas, como *Arabidopsis*, o DL se estende por distâncias superiores a 200kb (Nordborg et al., 2002).

Informações sobre o DL em *Eucalyptus* são escassas. Póvoa (2005) caracterizou o polimorfismo de uma região de 358 nucleotídeos do gene CAD2, que codifica uma enzima que catalisa uma etapa da rota de biossíntese de lignina, em 78 indivíduos de diferentes espécies de *Eucalyptus*. O autor detectou a ocorrência de 15 substituições nucleotídicas (SNPs), distribuídas de modo diverso entre as populações, que definiram, em conjunto, 11 haplótipos. Santos (2005) analisou a estrutura de DL em *E. grandis* e *E. urophylla* e, embora o estudo tenha sido limitado a um número pequeno de sítios polimórficos dentro de um fragmento de 440pb do gene *ccoamt* (também relacionado à biossíntese de lignina), os resultados sugerem que os SNPs tendem a se encontrar em DL a distâncias menores que 250pb em *E. grandis* e aproximadamente 70pb em *E. urophylla*.

Para traçar um padrão geral sobre o desequilíbrio de ligação em *Eucalyptus* são necessários mais estudos, com um número maior de seqüências analisadas, envolvendo também um maior número de locos. Informações sobre o DL em genes de interesse permitirão avaliar a eficiência da estratégia de mapeamento por associação, bem como das modernas abordagens de seleção assistida baseadas na análise do genoma completo. Essa é uma ferramenta importante, sobretudo em plantas perenes, devido à dificuldade de se gerar populações de mapeamento e de se avaliar características de expressão tardia.

4.9 ANÁLISE DOS NÍVEIS DE EXPRESSÃO DOS GENES DE CELULOSE SINTASE DE *Eucalyptus*

As seis seqüências de cDNA de *Eucalyptus* quando comparadas ao banco de ESTs do Projeto Genolyptus, permitiram a identificação de 225 *reads* homólogos aos seis diferentes genes de celulose sintase, distribuídos em diferentes bibliotecas de cDNA (Tabela 15).

Tabela 15. Distribuição de *reads* com elevado grau de similaridade às seqüências codantes completas (cds) de genes *EgCesA* de *Eucalyptus*.

Biblioteca	Tamanho da Biblioteca	Nº de reads homólogos à <i>CesA1</i>	Nº de reads homólogos à <i>CesA2</i>	Nº de reads homólogos à <i>CesA3</i>	Nº de reads homólogos à <i>CesA4</i>	Nº de reads homólogos à <i>CesA5</i>	Nº de reads homólogos à <i>CesA6</i>
EUGL-XY	15703	23	12	20	2	6	0
EUGR-ML	5934	1	1	0	0	0	1
EUGR-SE	10690	1	0	0	1	0	1
EUGR-XY	8463	7	14	7	2	3	2
EUGR-TS	12884	0	0	0	0	2	1
EUPE-XY	9755	6	11	5	5	1	3
EUSP-FX	12248	3	4	0	3	3	2
EUUR-XY	9500	17	20	19	2	6	2
EUGR-PU	8075	0	0	0	4	1	1

A Figura 26 representa a porcentagem de ESTs homólogos aos diferentes genes *EgCesA* expressos em tecidos xilemáticos de quatro espécies de *Eucalyptus*. Nota-se que os genes *EgCesA1*, *EgCesA2* e *EgCesA3* apresentam um maior nível de expressão no xilema do que os genes *EgCesA4*, *EgCesA5* e *EgCesA6*. Isso está de acordo com os resultados obtidos por Ranik & Myburg (2006), que separaram dois grupos de genes *EgCesA*: o grupo S, formado por *EgCesA1* – *EgCesA3*, de parede celular secundária, que se expressam de forma abundante no xilema, e o grupo P, formado por *EgCesA4*-*EgCesA6*, de parede celular primária, que apresentam um nível de expressão muito menor e um perfil mais uniforme no xilema.

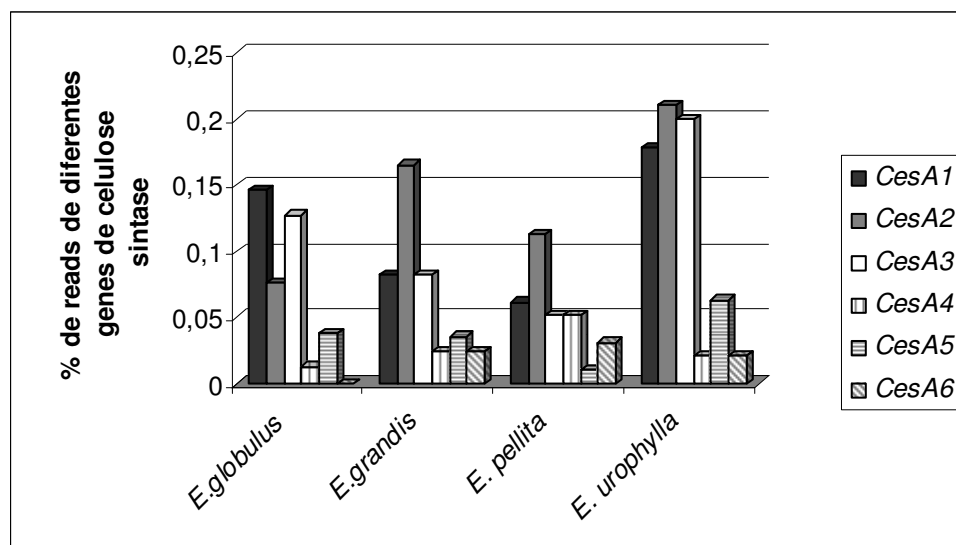


Figura 26. Comparação da porcentagem dos genes *CesA* expressos no xilema de diferentes espécies de *Eucalyptus*.

O perfil de expressão dos genes *CesA1-CesA3*, fortemente expressos em tecidos ricos em parede celular secundária, também variou entre as espécies. *E. urophylla* apresentou um perfil de expressão semelhante para os três genes, mas quando comparado ao nível de expressão destes genes nas demais espécies, foi possível perceber uma expressão significativamente superior destes três genes nesta espécie (Tabela 16). Se compararmos os perfis de expressão de *E. grandis* e *E. globulus*, é possível perceber uma expressão significativamente superior do gene *CesA2* em *E. grandis*. Da mesma forma, a expressão de *CesA2* foi superior em *E. pellita*, se compararmos ao que foi encontrado em *E. globulus*. Além disso, quando comparamos *E. grandis* e *E. pellita*, percebemos que os perfis de expressão para os três genes podem ser considerados semelhantes ($p > 0,05$).

Tabela 16. Significância das comparações entre os níveis de expressão dos genes de celulose sintase de parede celular secundária nas diferentes espécies de *Eucalyptus*.

Espécie 1	Espécie 2	p-Valor		
		<i>EgCesA1</i>	<i>EgCesA2</i>	<i>EgCesA3</i>
<i>E. urophylla</i>	<i>E. globulus</i>	0,078740	0,001953	0,035928
<i>E. urophylla</i>	<i>E. grandis</i>	0,037365	0,114535	0,019863
<i>E. urophylla</i>	<i>E. pellita</i>	0,010244	0,034441	0,002072
<i>E. globulus</i>	<i>E. grandis</i>	0,084464	0,029682	0,134286
<i>E. globulus</i>	<i>E. pellita</i>	0,027667	0,132289	0,034412
<i>E. grandis</i>	<i>E. pellita</i>	0,175427	0,093639	0,151209

A densidade básica da madeira é definida como a relação entre o peso seco absoluto e seu volume saturado, sendo de fundamental importância para o processo de obtenção da polpa celulósica (Silva Júnior, 2005). A densidade apresenta uma correlação com propriedades físicas, químicas e anatômicas da madeira e sua variação está associada ao teor de holocelulose, lignina, com a frequência de vasos e parênquimas, com a espessura da parede celular, dentre outros. Madeiras de alta densidade podem ser mais produtivas, em termos quantitativos, mas se o teor de lignina for elevado, a matéria-prima pode se tornar pouco interessante às indústrias de papel devido à elevação dos custos no processo de deslignificação. Outros estudos apontam que madeiras com densidades muito baixas também são indesejáveis para a produção de pasta de celulose, devido ao aumento do consumo volumétrico, diminuindo a capacidade de produção.

É possível sugerir uma tendência entre o perfil de expressão dos genes de celulose da parede secundária e às propriedades da madeira das diferentes espécies de celulose. As espécies *E. grandis* e *E. globulus* têm características contrastantes em relação às propriedades da madeira e da taxa de crescimento. Segundo uma avaliação feita por Duarte (2007), comparando-se o híbrido *E. grandis* x *E. urophylla* e *E. globulus*, o segundo apresenta características superiores na qualidade da madeira, como alta densidade, associado a um baixo teor de lignina, e um maior conteúdo de holocelulose, porém um crescimento lento em regiões tropicais. Já *E. grandis* apresenta um crescimento muito rápido, alta plasticidade fenotípica, mas a qualidade da madeira para a produção de celulose é inferior. *E. globulus* apresentou uma expressão maior dos genes *EgCesA1* e *EgCesA3*, em relação a *EgCesA2*, enquanto *E. grandis* e *E. pellita* apresentaram um perfil de expressão oposto: expressaram mais o gene *EgCesA2*, em relação à expressão de *EgCesA1* e *EgCesA3*. Já *E. urophylla* apresentou um maior nível de expressão de todos os genes *EgCesA*, possui uma densidade básica da madeira considerada alta, porém, um rendimento inferior em relação à produção de celulose.

Mas porque uma maior expressão dos genes *EgCesA* não significaria um aumento no rendimento da produção de celulose? A produção de celulose é afetada por vários outros aspectos e certamente é preciso analisar a quantidade e qualidade de hemicelulose, o teor de lignina, as características intrínsecas da madeira como um todo antes de tirar conclusões mais gerais sobre o rendimento de uma espécie na produção de fibras.

Quando comparamos a expressão dos genes *EgCesA1-EgCesA3* nos tecidos xilemáticos e em tecidos foliares, conforme o esperado, é possível notar uma diferença significativa na expressão desses genes no xilema (Tabela 17). Nota-se que foi encontrada expressão do gene *EgCesA3* somente no xilema (Figura 27). Esse resultado contrasta com o que foi observado por Ranik & Myburg (2006), que mediram a expressão tecido-específica desses genes em *E. grandis*, e constataram uma expressão, ainda que em pequena magnitude, do gene *EgCesA3* em tecidos como floema, casca e flores. Eles não encontraram expressão significativa desses genes em folhas jovens.

Tabela 17. Significância das diferenças nos níveis de expressão dos genes *EgCesA* no xilema e em tecidos foliares de diferentes espécies de *Eucalyptus*.

		p-valor					
		<i>EgCesA1</i>	<i>EgCesA2</i>	<i>EgCesA3</i>	<i>EgCesA4</i>	<i>EgCesA5</i>	<i>EgCesA6</i>
EUGL-XY	TECIDOS FOLIARES	0,000005	0,004035	0,000032	0,363725	0,094504	0,256914
EUGR-XY	TECIDOS FOLIARES	0,005303	0,000009	0,005303	0,202293	0,113738	0,202293
EUPE-XY	TECIDOS FOLIARES	0,021034	0,000405	0,04266	0,04266	0,325869	0,148546
EUUR-XY	TECIDOS FOLIARES	0,000002	0,000000	0,00000	0,233682	0,019071	0,233682

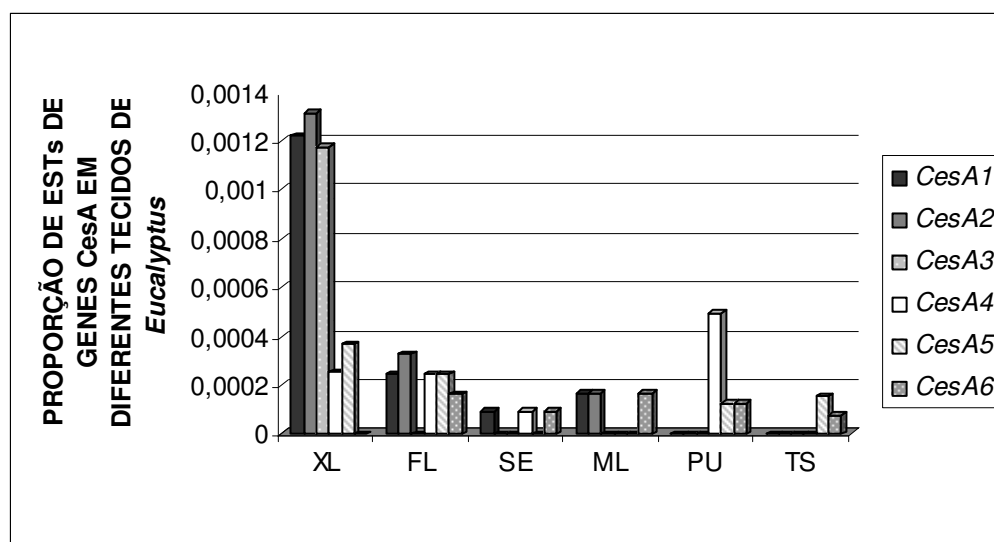


Figura 27. Distribuição do número de ESTs dos genes *EgCesA* observados em diferentes tecidos de *Eucalyptus*: XL (xilema); FL (floema); SE (plântulas); ML (folhas maduras); PU (folhas infectadas com *Puccinia*); TS (plantas submetidas a diversas condições de estresse).

A análise da expressão gênica, a partir de dados de *microarrays* também apresentou o perfil esperado para os genes *EgCesA1*, *EgCesA2* e *EgCesA3*, com maior expressão em tecidos relacionados à deposição de parede celular secundária. Já os genes

EgCesA4, *EgCesA5* e *EgCesA6*, apresentaram um perfil de expressão uniforme em tecidos com crescimento primário e secundário (Figura 28), em conformidade com os achados de Ranik & Myburg (2006).

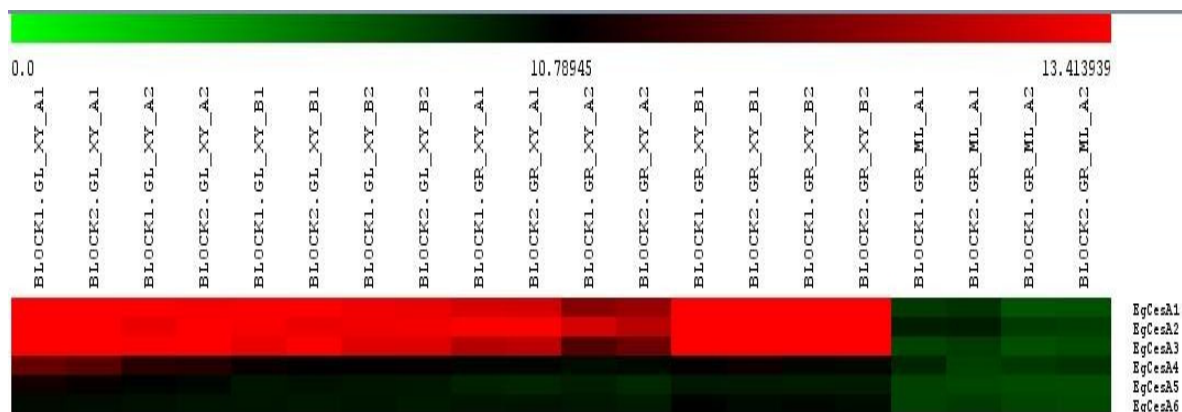


Figura 28. Perfil de expressão dos genes de celulose sintase a partir de dados de *microarrays*. Cada linha representa um gene, e as colunas representam os diferentes tratamentos; GL= *E. globulus*; GR= *E. grandis*; XY= xilema; ML= folhas maduras. O nível de expressão gênica é positivamente correlacionado à intensidade de sinal, que é crescente a partir da fluorescência de cor verde, até uma fluorescência de cor vermelha.

Considerando somente comparações sobre o perfil de expressão dos genes *EgCesA* expressos em xilema e folha de *E. grandis* (Figura 29), também houve uma significativa diferença de expressão dos genes *EgCesA1-EgCesA3* em xilema, de acordo com as análises feitas a partir de dados de *microarrays*. Os genes *EgCesA* apresentaram uma magnitude de expressão muito maior no xilema do que em tecidos foliares, enquanto o perfil de expressão desses em folhas foi aproximadamente constante.

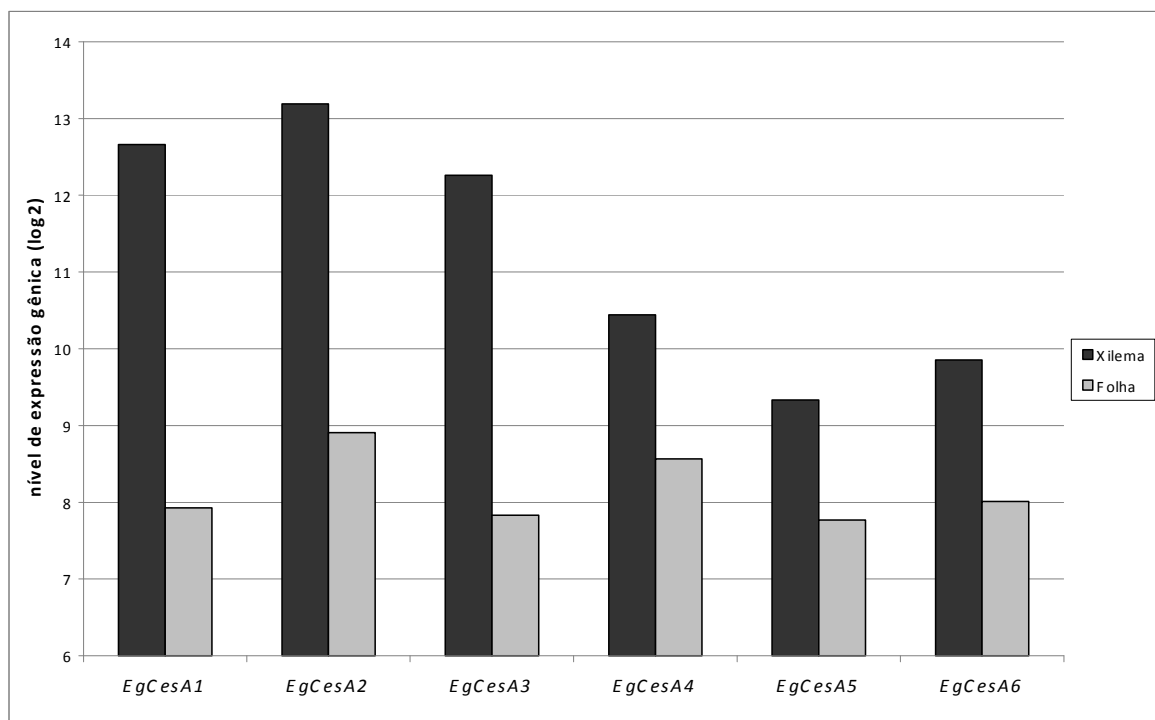


Figura 29. Nível de expressão gênica de dados de *microarrays* a partir de mRNA de xilema e folhas de *E. grandis*.

Tabela 18. Significância das diferenças nos níveis de expressão dos genes *EgCesA* no xilema e em tecidos foliares de *E. grandis* a partir de dados gerados com *microarrays*.

gene	Xilema		Folha		p-valor
	média	SE	média	SE	
<i>EgCesA1</i>	12,66	0,41	7,93	0,70	0,0000
<i>EgCesA2</i>	13,19	0,38	8,91	0,76	0,0001
<i>EgCesA3</i>	12,26	0,54	7,83	0,39	0,0000
<i>EgCesA4</i>	10,44	0,15	8,57	0,46	0,0003
<i>EgCesA5</i>	9,33	0,18	7,77	0,19	0,0000
<i>EgCesA6</i>	9,85	0,14	8,01	0,23	0,0000

Comparando-se os dados de *microarrays* em relação à expressão dos genes *EgCesA* em xilema de *E. grandis* com aqueles obtidos em xilema de *E. globulus* (Figuras 30 e 31), observou-se que o perfil de expressão dos genes foi semelhante para ambas as espécies. Isso contrasta com o que foi encontrado com o uso da abordagem denominada “*Digital Differential Display*”, que detectou expressão diferencial entre as duas espécies, sobretudo em relação aos genes *EgCesA* envolvidos na expressão de celulose da parede secundária.

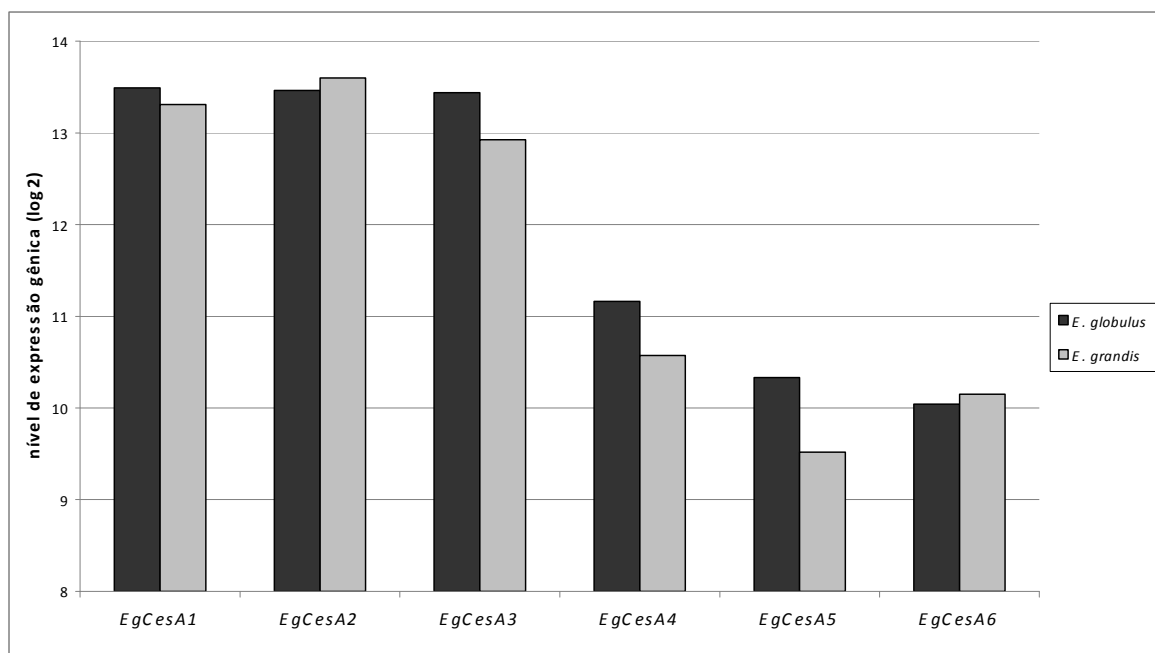


Figura 30. Nível de expressão gênica de dados de *microarrays* a partir de mRNA de xilema de *E. globulus* e de *E. grandis*.

Tabela 19. Significância das diferenças nos níveis de expressão dos genes *EgCesA* no xilema de *E. grandis* e *E. globulus* a partir de dados gerados por *microarrays*.

gene	<i>E. globulus</i>		<i>E. grandis</i>		p-valor
	média	SE	média	SE	
<i>EgCesA1</i>	13,49	0,25	13,31	0,75	0,5233
<i>EgCesA2</i>	13,46	0,29	13,60	0,51	0,5162
<i>EgCesA3</i>	13,44	0,26	12,92	0,80	0,1041
<i>EgCesA4</i>	11,16	0,41	10,57	0,23	0,0036
<i>EgCesA5</i>	10,33	0,46	9,52	0,24	0,0006
<i>EgCesA6</i>	10,04	0,19	10,15	0,35	0,4546

A explicação para esses resultados poderia estar no desenho das sondas de oligonucleotídeos, usadas para a implementação dos *arrays*. Existe uma crítica em relação aos resultados de *microarrays*, pois a detecção por hibridização, sobretudo em famílias gênicas é muito prejudicada pelas "reações cruzadas". Neste caso, as sondas, podem, dependendo do nível de estringência, hibridizar com cDNA de outros genes homólogos (mas não ortólogos). As sondas utilizadas neste estudo foram desenhadas de modo aleatório ao longo dos *contigs* e podem ser complementares a regiões pouco conservadas dentro de cada classe de genes *EgCesA*. O nível de expressão gênica poderia então ser subestimado, ou superestimado, para diferentes genes com produtos hibridizados

incorretamente com uma sonda não-ortóloga. A análise baseada em *microarrays*, nesse caso, apresentaria uma sensibilidade inferior àquela observada pela utilização da técnica de DDD. O desenho de sondas específicas, criadas para hibridizarem em regiões classe-específicas (CRSI e CRSII), poderia resolver o problema. De fato, quando foram feitas às comparações entre as seqüências de todas as sondas desenhadas para os *contigs* de celulose e os genes *EgCesA1-EgCesA6*, observou-se que várias sondas apresentaram um alto grau de identidade com diferentes classes dos genes *EgCesA*, pois essa complementação estava baseada em uma região conservada entre estes genes.

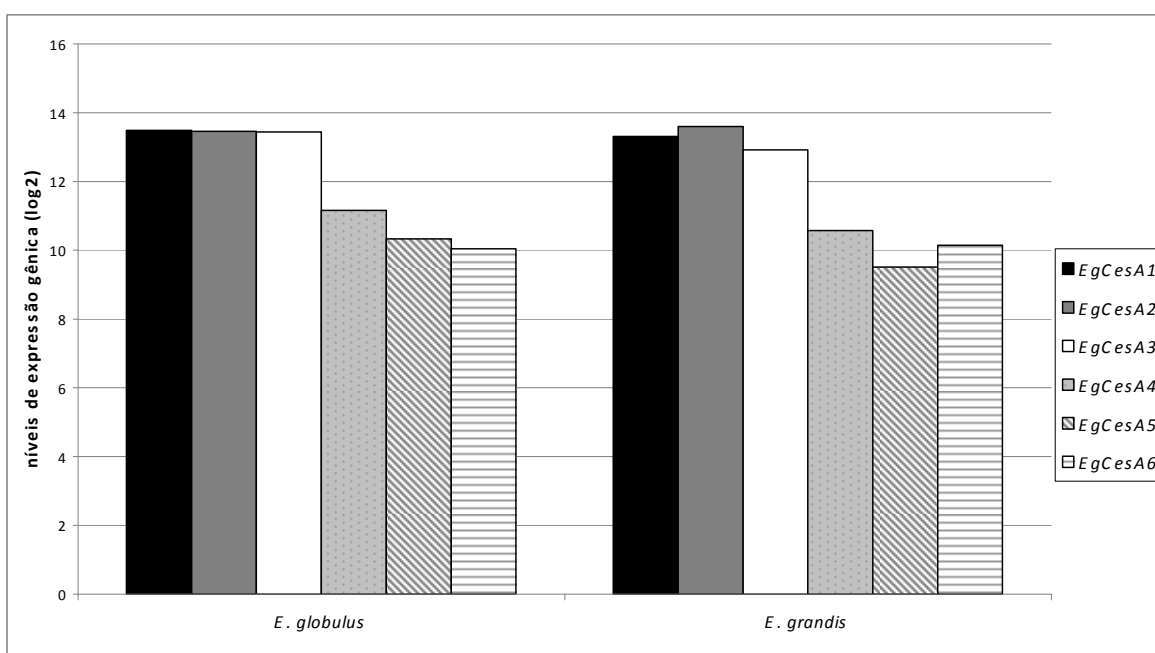


Figura 31. Nível de expressão gênica de dados de *microarrays* a partir de mRNA de xilema de *E. globulus* e de *E. grandis*.

Apesar dos dados de ensaios com *microarrays* não apresentarem a mesma sensibilidade de detecção de expressão gênica diferencial obtida com a análise DDD, é inegável a importância desse tipo de abordagem em estudos de expressão de genes envolvidos na formação da madeira, um processo complexo, que envolve muitos fenômenos biológicos, os quais são coordenados por uma ampla variedade de fatores endógenos (fitohormônios), exógenos (temperatura e fotoperíodo), e pela interação de ambos (Andrade, 2006). Nesse sentido, Hertzberg et al. (2001) investigaram a expressão de diferentes genes importantes para a formação da madeira, em diferentes estágios de desenvolvimento, a partir do meristema vascular de *Populus*. A análise de *microarray* usando um conjunto de 2995 genes únicos (*unigenes*) obtidos a partir de seqüências ESTs,

revelou que os genes que codificam enzimas envolvidas na biossíntese de celulose e lignina, bem como fatores de transcrição que regulam a xilogênese, estão sob estrita regulação transcricional, específica para cada fase de desenvolvimento.

5 CONCLUSÃO

- i. A análise de diversidade nucleotídica de fragmentos do gene *EgCesA2* revelou a ocorrência de três locos polimórficos numa região exônica do gene, composta por 427pb. Esses locos poderão ser utilizados em futuros projetos de mapeamento deste gene.
- ii. Foi feita uma estimativa do desequilíbrio de ligação entre os SNPs detectados no *contig* relacionado ao gene *EgCesA2*, tendo-se detectado um desequilíbrio significativo entre dois locos distantes 138pb um do outro. Existem poucos dados sobre a associação entre haplótipos em genes de *Eucalyptus* e, embora não se possa inferir a respeito da extensão do DL ao longo do genoma, sequer de todo o gene, considera-se que qualquer informação gerada é válida para que se possa alcançar uma compreensão maior a respeito do padrão de distribuição de polimorfismos.
- iii. A triagem realizada na biblioteca de BACs, construída no âmbito do projeto Genolyptus, revelou que somente um clone contém o gene *EgCesA2*. Essa informação deverá contribuir para a caracterização minuciosa do gene em *Eucalyptus*.
- iv. O seqüenciamento parcial do gene *EgCesA3* permitiu a formação de um *contig* de 770pb, composto por duas regiões exônicas, respectivamente, de 33 e 582pb, intercaladas por um íntron de 95pb de comprimento, além de 60pb da região 3'-UTR. Ao longo desse *contig*, foram encontrados nove sítios polimórficos, que poderão ser úteis na construção de mapas genéticos e em estudos de mapeamento por associação.

- v. Foi feita uma análise de diversidade no gene *EgCesA3* e conforme era esperado, a diversidade nucleotídica na região intrônica foi de uma magnitude de 7,5 vezes maior do que no éxon.
- vi. Análises de detecção do DL na extensão do gene *EgCesA3* indicaram que é possível detectar desequilíbrio em aproximadamente 600pb, superior ao encontrado no gene *ccoamt* da via de lignificação de *Eucalyptus*.
- vii. Todos os polimorfismos de um único nucleotídeo detectados nos genes *EgCesA2* e *EgCesA3* são sinônimos. Isso de fato corrobora com idéia de que a celulose é um polímero muito importante para a sobrevivência da planta, e, portanto, sua maquinaria de síntese apresenta características bastante conservadas, sendo que mutações em regiões efetivamente transcritas tenderiam a serem deletérias, e por isso não seriam fixadas.
- viii. A análise de expressão baseada em “*Digital Differential Display*” detectou expressão gênica diferencial quando se comparava xilema de diferentes espécies, ou mesmo de diferentes tecidos. *E. urophylla* apresentou o perfil de expressão mais distinto. Os perfis de expressão dos *E. globulus* e *E. grandis* são invertidos para os genes codificantes de celulose sintase de parede secundária. A análise baseada em dados de *microarrays* apresentou uma menor sensibilidade, não evidenciando os diferentes níveis de expressão destes genes em xilema de *E. grandis* e *E. globulus*.

6 REFERÊNCIAS

ALZATE, S. B. A. **Caracterização da madeira de árvores de clones de *Eucalyptus grandis*, *E. saligna*, e *E. grandis* x *urophylla***. 2004. 133f. Tese (Doutorado em Recursos Florestais)- Universidade de São Paulo, Piracicaba, 2004.

ALONI, Y.; COHEN, R.; BENZIMAN, M.; DELMER, D. P. Solubilization of the UDP-glucose: 1,4-beta-D-glucan 4-beta-D- glucosyltransferase (cellulose synthase) from *Acetobacter xylinum*. A comparison of regulatory properties with those of the membrane-bound form of the enzyme. **Journal of Biological Chemistry**, Bethesda, v. 258, n. 7, p. 4419-4423, 1983.

ALTSCHUL S. F.; GISH, W.; MILLER, W. MYERS, E. W. LIPMAN, D. J. Basic Local Alignment Search Tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403-410, 1990.

AMOR, Y.; HAIGLER, C. H.; JONHSON, S.; WAINSCOTT, M.; DELMER, D.P. A membrane-associated form of saccharose synthase and its potential role in synthesis of cellulose and callose in plants. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 92, n. 20, p. 9353-9357, 1995.

ANDRADE, A.; CELEDÓN, P. A. F.; LABATE, C. A. O proteoma da madeira. **Biotecnologia, Ciência e Desenvolvimento**, Brasília, v. 36, n. 1, p. 10-17, 2006.

ARIOLI, T.; PENG, L.; BETZNER, A. S.; BURN, J.; WITTKE, W.; HERTH, W.; CAMILLERI, C.; HÖFTE, H.; PLAZINSKI, J.; BIRCH, R.; CORK, A.; GLOVER, J.; REDMOND, J.; WILLIAMSON, R. E. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. **Science**, Washington, v. 279, n. 5351, p. 717-719, 1998.

ASSOCIAÇÃO BRASILEIRA DE CELULOSE E PAPEL (Bracelpa). **Informe Anual**, 2006. Disponível em <<http://www.bracelpa.org.br/anual/perfil2006.pdf>>. Acesso em: 23 ago. 2006.

ASSOCIAÇÃO BRASILEIRA DE CELULOSE E PAPEL (Bracelpa). **Relatório Socioambiental**, 2005. Disponível em <<http://www.bracelpa.org.br/br/social/socioambient.pdf>>. Acesso em: 10 de set. 2006.

ASSOCIAÇÃO BRASILEIRA DE CELULOSE E PAPEL (Bracelpa). **Reflorestamento: espécies mais plantadas de *Eucalyptus***, 2004. Disponível em <<http://www.bracelpa.org.br/br/numeros/reflorestamento/06-Eucalyptus.pdf>>. Acesso em: 10 de set. 2006.

AUDIC, S.; CLAVERIE, J. M. The significance of digital gene expression profiles. **Genome Research**, Cold Spring Harbor, v. 7, n. 10, p. 986-995, 1997.

BARRET, J. C.; FRY, B.; MALLER, J.; DALY, M. J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, Oxford, v. 21, n. 2, p. 263-265, 2005.

BERGELSON, J.; STAHL, E.; DUDEK, S. KREITMAN, M. Genetic variation within and among populations of *Arabidopsis thaliana*. **Genetics**, Pittsburgh, v. 148, n. 3, p. 1311-1323, 1998.

BHATTRAMAKKI, D.; DOLAN, M.; HANAFEY, M.; WINELAND, R.; VASKE, D.; REGISTER, J. C.; TINGEY, S. V.; RAFALSKI, A. Insetion-deletion polymorphims in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. **Plant Molecular Biology**, Springer Netherlands, v. 48, n. 5, p. 256-259, 2002.

BROOKES, A. J. The essence of SNPs. **Gene**, Amsterdam, v. 234, p. 177-186, 1999.

BROWN, R. M.; SAXENA, I. M.; KUDLICKA, K. Cellulose biosynthesis in higher plants. **Trends in Plant Science**, London, v.1, n. 5, p. 149-156, 1996.

BROWN JR., M. R.; SAXENA, I.M. Cellulose biosynthesis: a model for understanding the assembly of biopolymers. **Plant Physiology and Biochemistry**, Amsterdam, v. 38, n. 1, p. 57-67, 2000.

BURTON, R. A.; SHIRLEY, N. J.; KING, B. J.; HARVEY, A. J.; FINCHER, G. B. The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. **Plant Physiology**, Rockville, v. 134, n. 1, p. 224-236, 2004.

CARNEIRO, N. P.; CARNEIRO, A. A.; GUIMARÃES, C. T.; PAIVA, E. Desvendando o código genético. **Biociência, Ciência e Desenvolvimento**, Brasília, v. 17, p. 50-58, 2000.

CHAKRAVARTI, A. Population genetics-making sense out of sequence. **Nature Genetics**, New York, v. 21, n. 1, p. 56-60, 1999.

CHEN, Z.; HONG, X.; ZHANG, H.; WANG, Y.; LI, X.; ZHU, J.K.; GONG, Z. Disruption of the cellulose synthase gene, *AtCesA8/IRX1*, enhances drought and osmotic stress tolerance in *Arabidopsis*. **The Plant Journal**, Malden, v. 43, n. 2, p. 273-283, 2005.

CHING A.; CALDWELL, K.; JUNG, M.; DOLAN, M.; SMITH, H.; TINGEY, S.; MORGANTE, M.; RAFALSKI, A. SNP frequency and haplotype structure of 18 maize genes. **BMC Genetics**, London, v. 3, n. 19, p. 1-14, 2002.

CHO, R. J; MINDRINOS, M.; RICHARDS, D.R.; SAPOLSKY, R. J; ANDERSON, M.; DRENKARD, E.; DEWDNEY, J.; REUBER, T. L.; STAMMERS, M.; FEDERSPIEL, N. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. **Nature Genetics**, New York, v. 23, n. 3, p. 203-207, 1999.

COLLINS, F. S.; GUYER, M. S.; CHAKRAVARTI, A. Variation on a theme: cataloging human DNA sequence variation. **Science**, Washington, v. 278, n. 5343, p. 1580-1581, 1997.

DELMER, D. P.; Amor, Y. Cellulose biosynthesis. **The Plant Cell**, Rockville, v. 7, n. 7, p. 987-1000, 1995.

DESPREZ, T.; Vernhettes, S.; Fagard, M.; Refrégier, G.; Desnos, T.; Aletti, E.; Py, N.; Pelletier, S.; Höfte, H. Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6. **Plant Physiology**, Rockville, v. 128, n. 2, p. 482-490, 2002.

DOYLE, J. J.; DOYLE J. L. Isolation of plant DNA from fresh tissue. **Focus**, Carlsbad, v. 12, n. 1, p. 13-15, 1991.

DUARTE, F. A. S. **Avaliação da madeira de *Betula pendula*, *Eucalyptus globulus* e de híbrido de *Eucalyptus grandis* x *Eucalyptus urophylla* à produção de polpa celulósica Kraft**. 2007. 107f. Dissertação (Mestrado em Recursos Florestais- Tecnologia de Produtos Florestais)- Universidade de São Paulo, Esalq, Piracicaba, 2007.

DUAN, J.; WAINWRIGHT, M. S.; COMERON, J. M.; SAITOU, N.; SANDERS, A R.; GELERNTER, J.; GEJMAN, P. V. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. **Human Molecular Genetics**, Oxford, v. 12, n. 3, p. 205-216. 2003.

ELDRIDGE, K.; DAVIDSON, J.; HARWOOD, C. **Eucalypt domestication and breeding**. New York: Oxford University Press, 1994. 288p.

EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. **Arlequin ver. 3.1: An integrated software package for population genetics**. 2006. Disponível em: <<http://cmpg.unibe.ch/software/arlequin3>>. Acesso em 30 de julho de 2007.

EWING, B.; HILLIER, L.; WENDL, M. C. & GREEN, P. Base-calling of automated sequencer traces using PHRED. I. Accuracy Assessment. **Genome Research**, Cold Spring Harbor, v. 8, n. 3, p. 175-185, 1998.

FEARNSIDE, P. M. Forests and global warming mitigation in Brazil: Opportunities in the Brazilian forest sector for responses to global warming under the "Clean Development Mechanism". **Biomass and Bioenergy**, Amsterdam, v. 16, n. 3, p. 171-189, 1999.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Brasília: Empresa Brasileira de Pesquisa Agropecuária. Embrapa, 1998. 220p.

FIGUEIREDO, R. C.; BRITO, M. S.; FIGUEIREDO, L. H. M.; QUIAPIN, A.C.; VITORELLI, P. M.; SILVA, L. R.; SANTOS, R. V.; MOLFETTA, J. B.; GOLDMAN, G. H.; GOLDMAN, M. H. S. Dissecting the sugarcane expressed sequence tag (SUCEST) database: unraveling flower-specific genes. **Genetics and Molecular Biology**, São Paulo, v. 24, n. 1-4, p. 77-84, 2001.

FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER IV, E. S. Structure of linkage disequilibrium in plants. **Annual Review Plant Biology**, Palo Alto, v. 54, n. 1, p. 257-374, 2003.

FOELKEL, C. E. B.; MORA, E.; MENOCELLI, S. Densidade básica: sua verdadeira utilização como índice de qualidade da madeira de eucalipto para a produção de celulose. **O papel**, São Paulo, v. 53, n. 5, p. 35-40, 1992.

FUKUDA, H. Xylogenesis: initiation, progression, and cell death. **Annual Review Plant Physiology**, Palo Alto, v. 47, n. 1, p. 299-325, 1996.

GAUT, B. S.; LONG, A. D. The lowdown on linkage disequilibrium. **The Plant Cell**, Rockville, v. 15, n. 7, p. 1502-1506, 2003.

GONZÁLEZ, E.R. **Transformação genética de *Eucalyptus grandis* e do híbrido *E. grandis* x *E. urophylla* via *Agrobacterium***. 2002. 93f. Tese (Doutorado em Agronomia: Genética e Melhoramento de Plantas)- Escola de Agronomia Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2002.

GRATTAPAGLIA, D.; BRADSHAW JR., H. D. Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. **Canadian Journal of Forest Research**, Ottawa, v. 24, p. 1074-1078, 1994.

GUIMARÃES, P. E. M.; COSTA, M. C. R.; SNPs: sutis diferenças de um código. **Biotecnologia Ciência & desenvolvimento**, Brasília, v. 26, n. 5, p. 24-27, 2002.

GUT, I. G., Automation in genotyping single nucleotide polymorphisms. **Human Mutation**, Hoboken, v. 17, n. 6, p. 475-492, 2001.

HERTZBERG, M.; ASPEBORG, H.; SHRADER, J.; ANDERSSON, A.; ERLANDSSON, R.; BLOMQUIST, K.; BHALERAO, R.; UHLE'N, M.; TEERI, T. T.; LUNDEBERG, J.; SUNDBERG, B.; NILSSON, P.; SANDBERG, G. O. A transcriptional roadmap to wood formation. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 98, n. 25, p. 14732-14737. 2001.

HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, Heidelberg, v. 38, n. 6, p. 226-231, 1968.

HOLLAND, N.; HOLLAND, D.; HELENTJARIS, T.; DHUGGA, K. S.; XOCONOSTLE-CAZARES, B. & DELMER, D. P. A comparative analysis of the plant cellulose synthase (*CesA*) gene family. **Plant Physiology**, Rockville, v. 123, n. 4, p. 1313-1324, 2000.

HOUSE, S. M. Reproductive biology of *Eucalyptus*. In: WILLIAMS, J.; WOINARSKI, J. (Ed.) **Eucalypt Ecology**, Cambridge University Press, 1997. cap. 4, p. 56-92.

INSTITUTO DE PESQUISAS E ESTUDOS FLORESTAIS (IPEF). **Chave de Identificação de Espécies Florestais**, 2006. Disponível em < <http://www.ipef.br/identificacao/cief/espécies>. Acesso em 12 de novembro de 2006.

JOHNSON, G. C.; ESPOSITO, L.; BARRATT, B. J.; SMITH, A. N.; HEWARD, J.; DI GENOVA, G.; UEDA, H.; CORDELL, H. J.; EAVES, I. A.; DUDBRIDGE, F. Haplotype tagging for the identification of common disease genes. **Nature Genetics**, New York, v. 29, n. 2, p. 233-237, 2001.

JORDE, L. B. Linkage disequilibrium and the search for complex disease genes. **Genome Research**, Cold Spring Harbor, v. 10, n. 10, p. 1435-1444, 2000.

KAPPOR, M.L.; SHARMA, V.K. Experimentally synthesized allotetraploides in *Eucalyptus*. **Silvae Genetica**, Frankfurt, v. 34, n. 1, p. 19-22, 1985.

KIMURA, S.C.; LAOSINCHAI W.; ITOH, T.; CUI, X.; LINDER, R.; BROWN JR., R.M. Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. **The Plant Cell**, Rockville, v. 11, n. 11, p. 2075–2085, 1999.

KIRST, M.; MYBURG, A. A.; LEÓN, J. P. G., KIRST, M. E.; SCOTT, J. Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. **Plant Physiology**, Rockville, v. 135, n. 4, p. 2368-2378, 2004.

KIRST, M.; CHRISTOPHER, J. B.; MYBURG, A. A.; ZENG, Z.; SEDEROFF, R. R. Genetic architecture of transcript-level variation in differentiating xylem of a *Eucalyptus* hybrid. **Genetics**, Pittsburgh, v. 169, n. 4, p. 2295-2303, 2005.

KLEYN, P. W.; VESELL, E. S. Genetic variation as guide to drug development. **Science**, Washington, v. 281, n. 5384, p. 1820-1821, 1998.

KWOK, P. Methods for genotyping single nucleotide polymorphisms. **Annual Review Genomics and Human Genetics**, Palo Alto, v.2, n. 1, p. 235-258, 2001.

LEWONTIN, R. C. The interaction of selection on linkage I. **Genetics**, Pittsburgh, v. 49, n. 1, p. 49-67, 1964.

LI, W. **Molecular Evolution**. Sunderland: Sinauer Associates, 1997. 487p.

LIMA, W.P. **Impacto ambiental do *Eucalyptus***. 2. ed. São Paulo: Edusp, 1996. 301p.

LIPSHUTZ, R. J.; FODOR, S. P. A.; GINGERAS, T. R.; LOCKHART, D. J. High density synthetic oligonucleotide arrays. **Nature Genetics**, New York, v. 21, n. 1, p. 20-24, 1999.

MARCOLIN, M; SANTOS, A. A.; SALATI, E. Emissões e remoções de dióxido de carbono por mudanças nos estoques de florestas plantadas. **Brazilian Foundation of Sustainable Development**. Brasília. Ministério de Ciência e Tecnologia, 2002. 47p. Disponível em < <http://www.mct.gov.br/index.php/content/view/21456.html> >. Acesso em: 06 de fev. 2007.

MATTHYSSE, A. G.; THOMAS, D. L.; WHITE, A. R.; Mechanism of cellulose synthesis in *Agrobacterium tumefaciens*. **Journal of Bacterology**, Washington, v. 177, n. 4, p.1076-1081, 1995.

MOOT, R. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. **Computer Applications in the Biosciences**, Oxford, v. 13, n. 4, p. 477-478, 1997.

MOURA, V. P. G. O germoplasma de *Eucalyptus urophylla* S. T. Blake no Brasil. **Comunicado técnico**, n.111. Brasília: Embrapa-Recursos Genéticos e Biotecnologia, 2004. Disponível em < <http://www.cenargen.embrapa.br/publica/trabalhos/cot111.pdf>>. Acesso em: 10 de jun. 2007.

MYBURG, A. A. **Genetic Architecture of Hybrid Fitness and Wood Quality Traits in a Wide Interspecific Cross of *Eucalyptus* Tree Species**. 2001. 249 f. Tese (Doutorado em Silvicultura e Genética) – Universidade da Carolina do Norte, Raleigh, NC, 2001. Disponível em < <http://www.lib.ncsu.edu/theses/available/etd-20010723-175234/>>. Acesso em: 27 de julho de 2007.

MYBURG, A. A.; VOLGL, C.; GRIFFIN, R.; SEDEROFF, R. R.; WHETTEN, R. W. Genetics of postzygotic isolation in *Eucalyptus*: whole- genome analysis of barriers to introgression in a wide interespecific cross of *Eucalyptus grandis* and *E. globulus*. **Genetics**, Pittsburgh, v. 166, n. 1, p. 1405-1418, 2004.

MUELLER S. C.; BROWN JR, R. M. Evidence for an intramembrane component associated with a cellulose microfibril synthesizing complex in higher plants. **Journal of Cell Biology**, New York, v. 84, n. 2, p. 314-326, 1980.

NAIRN, J. C.; HASELKORN, T. Three loblolly pine *CesA* genes expressed in developing xylem are orthologous to secondary cell wall *CesA* genes of angiosperms. **New Phytologist**, Oak Ridge, v. 166, n. 3, p. 907-915, 2005.

NASU, S.; SUZUKI, J.; OHTA, R.; HASEGAWA, H.; YUI, R.; KITAZAWA, N. Search for and analysis of single nucleotide polymorphisms (SNPs) in Rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP Markers. **DNA Research**, Oxford, v. 9, n. 5, p. 163-171, 2002.

NICKERSON, D. A.; TOBE, V. O.; TAYLOR, S. L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. **Nucleic Acids Research**, Oxford, v. 25, n. 14, p. 2745-2751, 1997.

NICOL, F.; HIS, I.; JAUNEAU, A.; VERNHETTES, S.; CANUT, H.; HÖFTE, H. A plasma membrane-bound putative endo-1,4 β -D- glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. **The EMBO Journal**, New York, v. 17, n. 19, p. 5563-5576, 1998.

NEI, M.; W. LI. Mathematical model for studying genetic variance in terms of restriction endonucleases. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 76, n. 10, p. 5269-5273, 1979.

NORDBORG, M.; BOREVITZ, J. O.; BERGELSON, J.; BERRY, C. C.; CHORY, J.; HAGENBLAD, J.; KREITMAN, M.; MALOOF, J. N.; NOYES, T.; OEFNER, P. J.; STAHL, E. A.; WEIGEL, D. The extent of linkage disequilibrium in *Arabidopsis thaliana*. **Nature Genetics**, New York, v. 30, n. 2, p. 190-193, 2002.

OHMIYA, Y.; SAMEJIMA, M.; SHIROISHI, M.; AMANO, Y.; KANDA, T.; SAKAI, F.; HAYASHI, T. Evidence that endo-1,4 β -glucanases act on cellulose in suspension-cultured poplar cells. **The Plant Journal**, Malden, v. 24, n. 2, p. 147-158, 2000.

PAREDEZ, A.R.; SOMERVILLE, C.R.; EHRDARDT, D.W. Visualization of cellulose synthase demonstrates functional association with microtubules. **Science**, Washington, v. 9, n. 5779, p. 1482-1483, 2006.

PAGE, R. D. TreeView: an application to display phylogenetic trees on personal computers. **Computer Applications in the Biosciences**, Oxford, v. 12, n. 4, p. 357-358, 1996.

PEAR, J. R.; KAWAGOE, Y.; SCHRECKENGOST, W. E.; DELMER, D. Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 93, n. 22, p. 12637-12642, 1996.

POTTS, B.M.; WILTSHIRE, R.J.E. Eucalypt genetics and genecology. In: WILLIAMS, J.; WOINARSKI, J.(Ed.) **Eucalypt Ecology**. Cambridge University Press, 1997. cap.4, p. 56-92.

POT, D.; MCMILLAN, L.; ECHT, C.; PROVOST, G. L.; GARNIER-GÉRÉ, R.; CATO, S.; PLOMION, C. Nucleotide variation in genes involved in wood formation in two pine species. **New Phytologist**, Oak Ridge, v. 167, n. 1, p. 101-112, 2005.

PÓVOA, A. M. **Frequência de SNPs, estrutura de haplótipos e desequilíbrio de ligação para os genes CAD2 e COMT2 da via de lignificação em *Eucalyptus***. 2005. 126f. Dissertação (Mestrado em Ciências Genômicas e Biotecnologia)- Universidade Católica de Brasília, Brasília, 2005.

PRIMMORE, S. B. **Princípios de Análise do Genoma: um guia para mapeamento e seqüenciamento de DNA de diferentes organismos**. 2 ed. Ribeirão Preto: FUNPEC, 2003. 193p.

PRYOR, L. **Biology of *Eucalyptus***. London: Edward Arnold, 1976. 82p.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Opinion in Plant Biology**, Philadelphia, v. 5, n. 2, p. 94-100, 2002.

RANIK, M.; MYBURG, A. Six new cellulose synthase genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. **Tree Physiology**, Victoria, v. 26, n. 7, p. 545-556, 2006.

RAVEN, P. H.; EVERT, R. F.; EICHHORN, S. E. **Biologia Vegetal**. 6 ed. Rio de Janeiro: Guanabara Koogan, 2002. 906p.

READ, S. M. BACIC, T. Prime time for cellulose. **Science**, Washington, v. 295, n. 5552, p. 59-60, 2002.

REMLINGTON, D. L; THORNSBERRY, J. M; MATSUOKA, Y.; WILSON, L. M. WHITT, S. R.; DOEBLEY, J.; KRESOVICH, S.; GOODMAN, M. M. BUCKLER, E. S. Structure of linkage disequilibrium and phenotypic associations in the maize genome.

Proceedings of the National Academy of Sciences of the United States of America, Washington, v. 98, n. 20, p.11479-11484, 2001.

RIBEIRO, F. S.; ZANI FILHO, J. Variação da densidade básica da madeira em espécies/procedências de *Eucalyptus* spp. **Instituto de Pesquisa e Estudos Florestais**, Araraquara, v. 1, n. 46, p. 76-85, 1993.

ROMUALDI, C.; BORTOLUZZI, S.; DANIELI, G. A. Detecting differentially expressed genes in multiple tag sampling experiments: comparative evaluation of statistical tests. **Human Molecular Genetic**, Oxford, v.10, n. 19, p. 2133-2141, 2001.

ROZAS, J.; ROZAS, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis, **Bioinformatics**, Oxford, v. 2, n. 15, p. 174-175, 1999.

ROZEN, S. & SKALETSKY, H. J. Primer 3 on WWW for general users and for biologist programmers. In: KRAWETZ, S.; MISENER, S. (eds) **Bioinformatics methods and protocols: methods in molecular biology**. Totowa: Human Press, 2000. p. 365-386. Disponível em: <http://fokker.wi.mit.edu/primer3/>.

RUSSO, C. A. M. Como escolher genes para problemas filogenéticos específicos. In: Matioli, S. R. (Ed). **Biologia Molecular e Evolução**. Ribeirão Preto: Holos Editora, 2001, cap. 12, p. 130-136.

SAITOU, N. & NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular biology and evolution**, Oxford, v. 4, n. 4, p. 405-425, 1997.

SAEED, A.I.; SHAROV, V.; WHITE, J.; LI, J.; LIANG, W.; BHAGABATI, N.; BRAISTED, J.; KLAPA, M.; CURRIER, T.; THIAGARAJAN, M.; STURN, A.; SNUFFIN, M.; REZANTSEV, A.; POPOV, D.; RYLTSOV, A.; KOSTUKOVICH, E.; BORISOVSKY, I.; LIU, Z.; VINSAVICH, A.; TRUSH, V.; QUACKENBUSH, J. TM4: a free, open-source system for microarray data management and analysis. **BioTechniques**, London, v. 34, n. 2, p. 374-378, 2003.

SALAZAR, M. M. **Anotações e caracterização preliminar de genes de celulose sintase em diferentes espécies de Eucalyptus**. 2006. 78f. Dissertação (Mestrado em Biologia Molecular e Celular)- Universidade Federal de Goiás, Goiânia, 2006.

SANTOS, S. N. **Genes de lignificação em Eucalyptus: estrutura e diversidade genética dos genes 4cl e ccoaomt**. 2005. 208f. Dissertação (Mestrado em Ciências Genômicas e Biotecnologia)- Universidade Católica de Brasília, Brasília, 2005.

SAMUGA, A.; JOSHI, C. P. A new cellulose synthase gene (*PtrCesA2*) from aspen xylem is orthologous to Arabidopsis *AtCesA7* (*irx3*) gene associated with secondary cell wall synthesis. **Gene**, Amsterdam, v. 334, n. 1, p. 37-44, 2002.

SAXENA, I.M.; LIN, F. C.; BROWN JR, R. M. Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. **Plant Molecular Biology**, Springer Netherlands, v. 15, n. 5, p. 673-683, 1990.

SAXENA, I.M.; BROWN JR., I. M. Cellulose biosynthesis: current views and evolving concepts. **Annals of Botany**, Oxford, v. 96, n. 1, p. 9-21, 2005.

SCANAVA JÚNIOR, L. **Caracterização silvicultural, botânica e tecnológica do *Eucalyptus urophylla* S. T. Blake e de seu potencial para utilização em serraria.** 2001.107f. Dissertação (Mestrado em Ciências Florestais) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2001.

SHEN, L. X.; BASILION, J. P.; STANTON JR, V. P. Single-nucleotide polymorphism can cause different structural folds of mRNA. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 96, n. 14, p. 7871-7876, 1999.

SILVA JÚNIOR F. G. Qualidade da matéria prima – Um tema atual. Coluna/Artigo. **Celulose Online**, 2002/2003. Disponível em:<
<http://www.celuloseonline.com.br/colunista/colunista.asp?IDAssuntoMateria=2&iditem=96> >. Acesso em: 10 de maio de 2007.

SOCIEDADE BRASILEIRA DE SILVICULTURA. **Área plantada com *Pinus e Eucalyptus* no Brasil (Ha)**, 2001. Disponível em
<http://www.sbs.org.br/area_plantada.htm >. Acesso em: 10 jul. 2006.

STADEN, R.; BEAL, K. F. & BONFIELD, J. K. The staden package. **Methods in Molecular Biology**, v. 132, p. 115-130. 1998.

TAIZ L.; ZEIGER, E. **Plant Physiology**. 3. ed. Sunderland: Sinauer Associates, 2002. 650p.

TARAMINO, G.; TINGEY, S. Simple sequence repeats for germplasm analysis and mapping in maize. **Genome Research**, Cold Spring Harbor, v. 39, n. 2, p. 277-287, 1996.

TAYLOR, N.G.; SHEIBLE, W.; CUTLER, S.; SOMERVILLE, C.R.; TURNER, S. R. The irregular xylem 3 locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. **The Plant Cell**, Rockville, v. 11, n. 5, p. 769-779, 1999.

TAYLOR N. G.; HOWELLS, R. M.; HUTTLY, A. K.; VICKERS, K.; TURNER, S. R. Interactions among three distinct *CesA* proteins essential for cellulose synthesis. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 100, n. 3, p. 1450-1455, 2003.

TANAKA, K.; MURATA, K.; YAMAZAKI, M.; ONOSATO, K.; MIYAO, A.; HIROCHIKA, H. Three distinct rice cellulose synthase catalytic subunit genes required of cellulose synthesis in the secondary wall. **Plant Physiology**, Rockville, v.133, n. 1, p. 73-73, 2003.

TENAILLON, M. I.; SAWKINS, M. C.; LONG, A. D.; GAUT, R. L.; DOEBLEY, J. F. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 98, n. 16, p. 9161-9166, 2001.

THOMPSON, J. D.; GIBSON, T. J.; PLEWNIAK, F.; JEANMOUGIN, F. & HIGGING, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Research**, Oxford, v. 24, n. 25, p. 4876-4882, 1997.

THUMMA, B. R.; NOLAN, M. F.; EVANS, R.; MORAN, G. F. Polymorphisms in *Cinnamoyl CoA Reductase (CCR)* are Associated with Variation in Microfibril Angle in *Eucalyptus* spp. **Genetics**, New York, v. 171, n. 2, p. 1257-1268, 2005.

TURNER, S. R.; SOMERVILLE, C. R. Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary cell wall. **The Plant Cell**, Rockville, v. 9, n. 5, p. 689-701, 1997.

VAZ, A. R. C. **Posicionamento de microssatélites em clones BAC: rumo a integração do mapa genético e mapa físico (BACs) de *Eucalyptus grandis***. 2005. Xf. Monografia - Trabalho de Conclusão de Curso (Graduação em Biologia) – Universidade Católica de Goiás, Goiânia, 2005.

WATTERSON, G. A. On the number of segregating sites in genetical models without recombination. **Theoretical Population Biology**, New York, v. 7, n. 2, p.256-276, 1975.

WEIR, B.S. **Genetic Data Analysis II - Methods for Discrete Population Genetic Data**. Sinauer Associates, Sunderland, 1996.

WILLIAMS, J.; WOINARSKI, J. **Eucalypt Ecology**. Cambridge University Press, 1997. 442p.

WONG, H. C.; FEAR, A. L.; CALHOON, R. D.; EICHINGER, E. H.; MAYER, R.; AMIKAM, D.; BENZIMAN, M.; GELFAND, D. H.; MEADE, J. H.; EMERICK, A. W.; BRUNER, R.; BASSAT, A.; TAL, R. Genetic organization of the cellulose synthase operon in *Acetobacter xylinum*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 87, n. 20, p. 8130-8134, 1990.

WU, L.; JOSHI, C. P.; CHIANG, V. A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. **The Plant Journal**, Malden, v. 22, n. 6, p. 495-502, 2000.

ZHU, Y.L, SONG, Q. J. HYTEN, D.L. ; VAN TASSEL, C. P. ; MATUKUMALLI, L.K. ; GRIMM, D. R.; HYATT, S. M.; FICKUS, E. W.; YOUNG, N. D.; GREGAN, P. B. Single nucleotide polymorphism in soybean. **Genetics**, Pittsburgh, v. 163, n. 3, p. 123-1134, 2003.

ANEXO I

Estimativa da frequência dos haplótipos na população de *E.urophylla*.

	Haplótipos	Frequência
1	? ? ? ? ? ? ? C ?	0,000000
2	? ? ? ? ? ? ? C ? ?	0,031250
3	? ? ? ? ? ? ? C C C	0,015625
4	? ? ? ? ? ? ? T ? ?	0,125000
5	? ? ? ? ? ? ? T C C	0,015625
6	? ? A G ? ? T ? ?	0,031250
7	? ? G T ? ? ? ? ?	0,031250
8	? A G ? ? ? C ? ?	0,015625
9	? A G ? ? ? T ? ?	0,015625
10	? G G T G A C A ?	0,015625
11	? G G T G A T C ?	0,015625
12	C ? ? ? ? ? C ? ?	0,031250
13	C ? G T G A T ? ?	0,015625
14	C ? G T G G C ? ?	0,015625
15	C A A G ? G T A C	0,031250
16	C A A G G G T ? ?	0,031250
17	C A A G G G T A C	0,015625
18	C A A G T G ? A C	0,031250
19	C A G T ? ? ? ? ?	0,015625
20	C A G T G A ? ? ?	0,015625
21	C A G T G A C ? C	0,015625
22	C A G T G G C ? ?	0,031250
23	C G ? ? ? ? C ? ?	0,015625
24	C G ? ? ? ? T ? ?	0,015625
25	C G G T ? ? T ? ?	0,015625
26	C G G T G ? C A C	0,015625
27	C G G T G ? T C C	0,015625
28	C G G T G A ? ? ?	0,015625
29	C G G T G A ? C A	0,015625
30	C G G T G A ? C C	0,015625
31	C G G T G A C A A	0,015625
32	C G G T G A C C ?	0,031263
33	C G G T G A C C A	0,015625
34	C G G T G A C C C	0,062500
35	C G G T G A T ? C	0,015625
36	T G G T G G T ? ?	0,015625
37	T ? G T G A C ? ?	0,015625
38	T ? G T G G T ? ?	0,015625
39	T A A G G G C A C	0,031250
40	T A A G G G T A ?	0,031250
41	T A A G G G T A C	0,015625
42	T A G T G G T ? ?	0,015625


```

67003910|gb|DQ014507.1|
contig
ruleM
*****
2550
10
2410.....2420.....2430.....2440.....2450.....2460.....2470.....2480.....2490.....2500.....2510.....2520.....2530.....2540.....2550

*****
67003910|gb|DQ014507.1|
contig
ruleM
*****
2605
159
2560.....2570.....2580.....2590.....2600.....2610.....2620.....2630.....2640.....2650.....2660.....2670.....2680.....2690.....2700

*****
67003910|gb|DQ014507.1|
contig
ruleM
*****
2705
159
2710.....2720.....2730.....2740.....2750.....2760.....2770.....2780.....2790.....2800.....2810.....2820.....2830.....2840.....2850

*****
67003910|gb|DQ014507.1|
contig
ruleM
*****
2905
459
2860.....2870.....2880.....2890.....2900.....2910.....2920.....2930.....2940.....2950.....2960.....2970.....2980.....2990.....3000

*****
67003910|gb|DQ014507.1|
contig
ruleM
*****
3055
609
3010.....3020.....3030.....3040.....3050.....3060.....3070.....3080.....3090.....3100.....3110.....3120.....3130.....3140.....3150

*****
67003910|gb|DQ014507.1|
contig
ruleM
*****
3205
759
3110.....3120.....3130.....3140.....3150.....3160.....3170.....3180.....3190.....3200.....3210.....3220.....3230.....3240.....3250.....3260.....3270.....3280.....3290.....3300.....3310.....3320.....3330.....3340.....3350

*****
67003910|gb|DQ014507.1|
contig
ruleM
*****
3355
782
3310.....3320.....3330.....3340.....3350.....3360.....3370.....3380.....3390.....3400.....3410.....3420.....3430.....3440.....3450

*****
67003910|gb|DQ014507.1|
contig
ruleM
*****
3452
782
3460.....3470.....3480.....3490.....3500.....3510.....3520.....3530.....3540.....

```