Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Caracterização de perfis transcricionais de folhas e da região cambial de
***Eucalyptus grandis* usando o SAGE**

Mayra Costa da Cruz Gallo de Carvalho

Tese apresentada para obtenção do título de Doutor em
Agronomia. Área de concentração: Genética e
Melhoramento de Plantas

Piracicaba

2006

Mayra Costa da Cruz Gallo de Carvalho
Bacharel em Ciências Biológicas

**Caracterização de perfis transcricionais de folhas e da região cambial de *Eucalyptus grandis*
usando o SAGE**

Orientador:
Prof. Dr. CARLOS ALBERTO LABATE

**Tese apresentada para obtenção do título de Doutor em Agronomia. Área de
concentração: Genética e Melhoramento de Plantas**

Piracicaba
2006

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Carvalho, Mayra Costa da Cruz Gallo de

Caracterização de perfis transcricionais de folhas e da região cambial de *Eucalyptus grandis* usando o SAGE / Mayra Costa da Cruz Gallo de Carvalho. - - Piracicaba, 2006.
153p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 2006.
Bibliografia.

1. Eucalipto 2. Expressão gênica 3. Folhas (plantas) 4. Fotossíntese 5. Parede celular vegetal 6. Respiração vegetal 7. Xilema I. Título

CDD 634.9734

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

Dedico este trabalho aos meus pais, Adalberto e Deisy, e ao meu marido, Paulo André.
Sem a compreensão e o amor de vocês este sonho não teria sido possível.

AGRADECIMENTOS

Ao CNPq pelo financiamento da minha bolsa de estudos e à Companhia Suzano de Papel e Celulose pelo apoio financeiro e material biológico.

Ao professor Dr. Carlos Alberto Labate, pela oportunidade, orientação, companheirismo, e pela imensa contribuição à minha formação acadêmica.

Ao Dr. David H. Moon, pelo empenho, ensinamentos, apoio, direcionamento e, principalmente, pela amizade.

À Lívia Maria Franceschini pela amizade, atenção e auxílio.

Aos amigos do laboratório Dra. Mônica Teresa V. Labate, Esteban Gonzáles, Guilherme Salvatierro, Eduardo Leal Oliveira Camargo, Felipe Boaretto, Gisele Lacerda, Inês Faraldo, Lilia Barbosa, Paola F. Celedon, Gabriela Conti, Ana Letícia Bertolo, Karem G. X. Meireles e Alexander de Andrade, pela alegria e apoio.

Aos meus amigos Danielle Gregório Gomes, Fernanda Salvato, Simone Bragatto, Juliano Bragatto, Raphael T. Carneiro, Danielle Gregório da Silva, Cristiano Medri, Moacyr Eurípedes Medri, Sheila Recepute Silveira e Mariane Murakami, agradeço por tudo que aprendi, pelo incentivo, alegria e amizade. Obrigada por fazer a minha vida ainda mais feliz!

Aos meus queridos agregados, Heloísa Helena D. de Carvalho e Rui Cláudio Celini de Carvalho, pelo carinho e por se fazerem sempre presente nos momentos de alegria e também nos de dificuldade. Aos meus queridos irmãos Vitor Costa da C. Gallo, Adalberto Jardim Gallo Filho e também Estevam Luís de Carvalho e Pedro Lucas de Carvalho, pelo carinho e apoio.

SUMÁRIO

RESUMO	7
ABSTRACT	9
LISTA DE FIGURAS	11
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 REVISÃO BIBLIOGRÁFICA	21
2.1 O eucalipto e o setor florestal	21
2.2 Genômica e pós-genômica	25
2.2.1 Transcrissômica	26
2.2.2 Análise serial da expressão gênica (SAGE)	29
2.3 Folhas – órgãos fotossintetizantes	33
2.4 A formação da madeira	38
3 OBJETIVOS	43
4 MATERIAL E MÉTODOS	45
4.1 Material vegetal	45
4.2 Extração do RNA total	46
4.3 Construção das bibliotecas SAGE de madeira e folhas	48
4.3.1 Isolamento do RNA mensageiro e síntese de cDNA	48
4.3.2 Digestão do cDNA com a endonuclease <i>Nla</i> III e ligação dos adaptadores	50
4.3.3 Digestão com a endonuclease <i>Bsm</i> F I e formação dos dímeros de cDNA	51
4.3.4 Otimização das reações de PCR para produção de dímeros em larga escala	52
4.3.5 Amplificação em larga escala e purificação dos dímeros de cDNA	53
4.3.6 Digestão dos dímeros de cDNA com <i>Nla</i> III	55
4.3.7 Polimerização/Concatenação e clonagem das <i>ditags</i>	56
4.4 Sequenciamento dos polímeros de <i>ditags</i>	58
4.5 Bioinformática e análise dos dados	59
4.5.1 Seleção de sequências para análise	59
4.5.2 Extração das SAGE <i>tags</i>	60
4.5.3 Comparação dos níveis de expressão das <i>tags</i> entre as bibliotecas	63

4.5.4 Encontrando as sequências de origem das <i>tags</i> e suas prováveis funções celulares	65
5 RESULTADOS E DISCUSSÃO	69
5.1 Resumo dos resultados produzidos através da metodologia SAGE	69
5.2 Perfil transcricional da biblioteca de folhas	72
5.3 Perfil transcricional da biblioteca de madeira	88
5.4 Perfil comparativo da expressão gênica entre folhas e madeira	107
6 CONSIDERAÇÕES FINAIS	125
REFERÊNCIAS	129
ANEXOS	147

RESUMO

Caracterização de perfis transcricionais de folhas e da região cambial de *Eucalyptus grandis* usando o SAGE

Apenas muito recentemente estratégias genômicas e pós-genômicas têm sido utilizadas em estudos de espécies arbóreas importantes na silvicultura. Nos países tropicais, a exemplo do Brasil, as espécies de eucalipto são especialmente importantes nos plantios comerciais principalmente destinados à indústria de papel e celulose. O eucalipto é caracterizado por elevadas taxas de crescimento e alta adaptabilidade resultante da interação de mecanismos moleculares e processos metabólicos ainda pouco conhecidos. Nesse sentido, o presente trabalho teve como objetivo principal o estabelecimento de perfis transcricionais comparativos para folhas e para o xilema em formação de árvores de *Eucalyptus grandis*. O uso do método SAGE permitiu analisar 5864 genes dos quais 2247 (38%) puderam ser identificados. Foram encontrados 464 genes diferencialmente expressos, sendo 47 exclusivamente expressos na região cambial e 64 nas folhas. Além disso, as estratégias adotadas na identificação *tags*-genes permitiram que as SAGE *tags* fossem eficientemente utilizadas na diferenciação de isoformas gênicas tecido-específicas e de localização celular específica. Genes relacionados à fotossíntese, fotorrespiração e destoxificação celular foram preferencialmente encontrados na biblioteca SAGE de folhas, enquanto genes relacionados à síntese da parede celular, organização do citoesqueleto e respiração, foram preferencialmente expressos na biblioteca SAGE de madeira. Os níveis de expressão dos transcritos sugeriram a ocorrência de possíveis mecanismos de controle transcricional comum para grupos de genes funcionalmente relacionados e foram também utilizados para sugestão de genes e processos que representam alvos potenciais para o melhoramento do eucalipto.

Palavras-chave: *Eucalyptus*, SAGE, folhas, xilema, fotossíntese, fotorrespiração, respiração, parede celular

ABSTRACT

Characterization of *Eucalyptus grandis* leaves and cambial region transcriptomes using SAGE

Recently genomic and post-genomic strategies have been used to study important tree species in planted forests. In the tropical countries, like Brazil, *Eucalyptus* species are especially important in commercial plantations destined for the paper and cellulose industry. *Eucalyptus* species are characterized by their high growth rates and great adaptability resulting from the interaction between molecular mechanisms and metabolic processes that are still uncharacterized. Thus, the main goal of this work was the comparison of the transcriptional profiles from leaves and the cambial region of *Eucalyptus grandis* trees. The use of SAGE allowed the evaluation of 5864 expressed tags of which 2247 (38%) could be identified. 464 differentially expressed tags were indicated, of which 47 were exclusively expressed in the cambial region library and 64 in the leaf library. Furthermore, the strategies used in the tag mapping process allowed an efficient use of the SAGE tags to distinguish gene isoforms with tissue-specific and/or specific cell localizations. Genes related to photosynthesis, photorespiration and cellular detoxification were preferentially found in the SAGE leaf library, while genes involved in cell wall biosynthesis, cytoskeletal organization and respiration were preferentially expressed in the developing xylem. Transcript expression levels suggested the presence of a common transcriptional control for a few functionally related genes, for example, some lignin related genes. Importantly, transcript expression levels were also used for the identification of target genes and processes potentially useful in future *Eucalyptus* breeding programs.

Key-words: *Eucalyptus*, SAGE, leaves, xylem, photosynthesis, photorespiration, respiration, cell wall

LISTA DE FIGURAS

Figura 1 - Esquema da estrutura dos cloroplastos e posicionamento dos complexos proteicos dos fotossistema I e II nas membranas dos tilacóides	34
Figura 2 - Transferência de elétrons e prótons na membrana dos tilacóides através dos complexos proteicos PSII, citocromo <i>b₆f</i> , PSI e ATP-sintase.....	36
Figura 3 - Coleta do material utilizado na construção da biblioteca SAGE de madeira	45
Figura 4 - Gel de eletroforese do produto do RT-PCR teste realizado com as amostras de RNA total de folhas (F) e madeira (M) usando <i>primers</i> para a UDP-glicose desidrogenase de <i>E. grandis</i>	47
Figura 5 - Esquema ilustrativo da metodologia SAGE (análise serial da expressão gênica)	48
Figura 6 - Gel de eletroforese dos produtos das reações de otimização da concentração de DNA para PCR	53
Figura 7 - Gel de eletroforese dos produtos da amplificação em larga escala dos dímeros de cDNA	54
Figura 8 - Gel de eletroforese dos produtos da digestão com <i>Nla</i> III	55
Figura 9 - Gel de eletroforese dos produtos da ligação das <i>ditags</i>	57
Figura 10 - Cromatogramas de dois eventos de sequenciamento de um mesmo SAGE clone	60
Figura 11 - Arquivo de saída do <i>software</i> SAGE2000 TM (versão 4.5) ou <i>Tag list</i> da biblioteca de folhas.....	61

Figura 12 - Arquivo de saída do <i>software</i> SAGE2000 TM (versão 4.5) ou <i>Tag list</i> comparando as bibliotecas de folhas e madeira.....	62
Figura 13 - Resumo das etapas de geração e análise das bibliotecas SAGE	63
Figura 14 - Planilha <i>Excel</i> do teste <i>Z</i> comparando os níveis de expressão de <i>tags</i> /genes entre as bibliotecas SAGE de folhas e madeira.....	65
Figura 15 - Seqüência ilustrativa de um EST utilizado na anotação de uma <i>tag</i>	66
Figura 16 - Fichas “eletrônicas” do banco de dados local de genômica funcional de <i>E. grandis</i> .	68
Figura 17 - Distribuição dos genes identificados a partir da biblioteca SAGE de folhas em categorias funcionais.....	73
Figura 18 - Distribuição dos genes por abundância ou número de transcritos identificados a partir da biblioteca SAGE de folhas em categorias funcionais.....	75
Figura 19 - Ciclo de Benson-Calvin. Assimilação de CO ₂ e regeneração de ribulose-1,5-bisfosfato.....	80
Figura 20 - Glicólise e Ciclo dos Ácidos Tricarboxílicos	84
Figura 21 - Oxigenação de ribulose-1,5-bisfosfato e Fotorrespiração	86
Figura 22 - Distribuição dos genes identificados a partir da biblioteca SAGE de madeira em categorias funcionais.....	89
Figura 23 - Distribuição dos genes por abundância ou número de transcritos identificados a partir da biblioteca SAGE de madeira em categorias funcionais.....	92

Figura 24 - Reações de interconverções de açúcares-NDP e síntese de hemiceluloses e pectinas.....	95
Figura 25 - Direcionamento de sacarose para síntese de celulose	97
Figura 26 - Síntese e polimerização dos monolignóis	100
Figura 27 - Reciclagem local de amônia durante a síntese e degradação de fenilalanina	106

LISTA DE TABELAS

Tabela 1 - Otimização das reações de PCR a partir de diferentes concentrações da amostra	52
Tabela 2 - Reação de ligação do vetor aos polímeros de <i>ditags</i>	58
Tabela 3 - Classificação Funcional de RISON <i>et al.</i> (2000)	67
Tabela 4 - Resumo dos dados gerais produzidos a partir das bibliotecas SAGE	69
Tabela 5 - Lista dos 50 genes mais expressos na biblioteca de Folhas	74
Tabela 6 - <i>Tags</i> relacionadas aos genes dos fotossistema I (PSI), fotossistema II (PSII), citocromo <i>b₆f</i> e ATP sintetase, encontradas nas bibliotecas SAGE.....	78
Tabela 7 - Lista das <i>Tags</i> que representam genes relacionados à proteção celular contra o estresse oxidativo encontradas nas bibliotecas SAGE.....	88
Tabela 8 - Lista dos 50 genes mais expressos na biblioteca de Madeira	90
Tabela 9 - <i>Tags</i> relacionadas aos genes nucleares da cadeia respiratória mitocondrial encontradas nas bibliotecas SAGE.....	108
Tabela 10 - Lista de genes diferencialmente expressos entre as bibliotecas de folhas e madeira ($P \leq 0,01$).....	109
Tabela 11 - Distribuição dos genes identificados nas bibliotecas de Madeira e Folhas em categorias e subcategorias funcionais.....	119

1 INTRODUÇÃO

As florestas são fundamentalmente importantes na manutenção dos ecossistemas e redução dos efeitos da atividade industrial sobre o clima mundial através do seqüestro do CO₂ produzido em altas quantidades com a queima de combustíveis fósseis. Alternativamente, as florestas representam uma importante fonte de recursos renováveis para geração de energia, produção de polpa e papel e para a construção. A demanda por produtos florestais, especialmente, polpa, papel e energia, têm aumentado nas últimas décadas e continuará em franca expansão nos próximos anos. A previsão de aumento é de 50% até 2010 (FAO, 2000) o que torna crescente a necessidade de produção de florestas plantadas cada vez mais produtivas para minimizar ao máximo a exploração das florestas naturais.

No Brasil, a redução da exploração sobre as florestas nativas foi inicialmente alcançada com a política de incentivos fiscais do governo brasileiro em 1995 para o estabelecimento em larga escala das florestas plantadas, principalmente constituídas de espécies de *Eucalyptus sp.* Atualmente o Brasil mantém cerca de 5,4 milhões de ha de plantios florestais, dos quais 3,4 milhões de ha são ocupados com eucalipto (SBS, 2005). Nas últimas décadas o setor florestal, principalmente relacionado à indústria de papel e celulose, apresentou um grande crescimento situando o Brasil em primeiro lugar no *ranking* mundial de produção de celulose fibra curta (de eucalipto). Os ganhos em produtividade nos plantios de eucalipto desde a década de 60, cerca de 300% (LEITE, 2005), foram alcançados através de ciclos sucessivos de melhoramento, hibridação interespecífica e segregação de clones superiores ou estabelecimento de plantios clonais, sendo a última a principal estratégia adotada.

Embora os ganhos em produtividade nos plantios florestais de eucalipto no Brasil tenham sido significativos, ainda não é possível suprir a necessidade de madeira no mercado. Neste cenário atual, o estabelecimento de estratégias que demandem menos tempo e áreas para plantio, como o emprego de métodos biotecnológicos, oferecem um potencial imenso para o uso de práticas mais sustentáveis ao melhoramento florestal. Obviamente, os objetivos dessas estratégias devem se concentrar no aumento do volume de madeira produzida nos plantios florestais, incremento na resistência a patógenos e na modificação de características da madeira que a torne mais adequada ao seu uso final.

O primeiro passo para alcançar estes objetivos é o conhecimento dos mecanismos moleculares relacionados ao crescimento vegetativo e formação da madeira em *Eucalyptus*. Entretanto, apenas poucos pesquisadores têm se dedicado ao estabelecimento de perfis funcionais do metabolismo de espécies de eucalipto, e os poucos trabalhos já disponíveis trazem informações muito pontuais do crescimento que estão estritamente relacionadas a um estágio do desenvolvimento das árvores e especificamente associadas à atividade do meristema cambial na formação de madeira (PAUX *et al.*, 2004, PAUX *et al.*, 2005; FOUCART *et al.*, 2006). Por outro lado, empresas do setor florestal e governos de diversos países têm se associado em grupos de pesquisa para gerar informações genômicas de espécies de eucalipto como é o caso, por exemplo, dos projetos de iniciativa Brasileira *Forests, Genolyptus* e, mais recentemente, do consórcio internacional do genoma de *Eucalyptus* (IEuGC) formado por pesquisadores de vários países. Estas iniciativas, associadas a disponibilização pública crescente de ESTs de *Eucalyptus sp.* nos bancos de dados mundiais, e a disponibilização de bancos genômicos de espécies arbóreas como o *Populus*, representam um importante passo fornecendo suporte aos projetos de pesquisa de genômica funcional de espécies florestais.

Diferentes estratégias complementares podem ser adotadas no estabelecimento de perfis funcionais, entre elas, as que revelam o conjunto de genes transcritos num dado momento ou transcrissômica, as que identificam o conjunto de proteínas ou proteômica, ou ainda as que se preocupam com os metabólitos celulares ou metabolômica. Estas diferentes linhas de pesquisa vêm sendo implementadas e desenvolvidas no laboratório Max Feffer de Genética de Plantas do Departamento de Genética da ESALQ, através de projetos de pesquisa com diferentes tecidos, tais como raízes, folhas e madeira, em diferentes estágios do desenvolvimento e/ou tratamentos, como madeira de tração, árvores jovens, maduras e plântulas de *Eucalyptus grandis*.

Neste trabalho, a técnica SAGE (*Serial Analysis of Gene Expression*), que é um método de análise global do transcrissoma, foi adotada para gerar perfis transcricionais de folhas e da região cambial de árvores de *Eucalyptus grandis*. Estas informações foram produzidas com o objetivo maior de possibilitar o estabelecimento de um panorama global do crescimento vegetativo e formação da madeira no eucalipto. Assim, o presente trabalho está inserido na linha de pesquisa de transcrissômica de *Eucalyptus grandis* e teve como objetivos: (a) implementar o uso da técnica SAGE no estabelecimento de perfis transcricionais com a construção de um banco de dados local de transcritos, (b) estabelecer o perfil transcricional de folhas de árvores de

Eucalyptus grandis, (c) estabelecer o perfil transcricional da região cambial (madeira em formação) de árvores de *Eucalyptus grandis*, (d) comparar os perfis transcricionais de folhas e da região cambial, identificando processos e/ou genes candidatos para o melhoramento, (e) identificar genes exclusivos entre folhas e madeira que apresentam potencial para o desenvolvimento de promotores tecido-específicos a serem utilizados em eventos de transgenia.

2 REVISÃO BIBLIOGRÁFICA

2.1 O eucalipto e o setor florestal

O gênero *Eucalyptus* pertence à família *Myrtaceae* sendo formado por mais de 600 denominações entre espécies, variedades e híbridos naturais. O gênero teve sua origem na Austrália, mas atualmente encontra-se amplamente distribuído em países tropicais e subtropicais, com exemplares adaptados a diversas condições de clima e solo. As espécies do gênero são monóicas e preferencialmente alógamas (apresentam 30% de autogamia) e são polinizadas por insetos (PRYOR, 1976).

As primeiras sementes do gênero foram levadas para a Europa por volta de 1774 apenas como curiosidades botânica (ANDRADE, 1961). Na América do Sul, o Chile foi o primeiro país a introduzir o eucalipto em 1823 seguido pela Argentina, Uruguai e Brasil. Somente em 1850 as espécies de eucalipto começaram a ser testadas para a produção de madeira em países europeus e passaram a ser cultivadas em diversos países. Atualmente, as espécies mais utilizadas no estabelecimento de florestas plantadas são o *E. grandis*, *E. urophylla*, *E. camaldulensis*, *E. saligna*, *E. citriodora*, *E. globulus*, *E. viminalis* e *E. tereticornis* (MORA & GARCIA, 2000).

A área total reflorestada com o eucalipto no mundo chega a 18 milhões de ha (FAO, 2000) com grandes plantios comerciais na Ásia, principalmente Índia, na América do Sul, principalmente no Brasil, na África do Sul, Portugal e Espanha. Dos 18 milhões de ha mundiais plantados com eucalipto, 3,4 milhões de ha estão no Brasil (SBS, 2005). Os Estados brasileiros que apresentam a maior área plantada são Minas Gerais, São Paulo, Bahia e Espírito Santo, nesta ordem de importância (MORA & GARCIA, 2000).

A silvicultura intensiva teve início no Brasil no século passado com o estabelecimento de plantios florestais de espécies exóticas de eucalipto, em uma iniciativa liderada pela companhia Paulista de Estrada de Ferro em 1903, através do Agrônomo Edmundo Navarro de Andrade. Os primeiros plantios de eucalipto foram destinados à produção de energia para as locomotivas, construção de ferrovias e postes elétricos. Entretanto, a imensa variedade de espécies de eucalipto utilizadas nestes plantios acabou transformando o Horto Florestal de Rio Claro - São Paulo, no mais completo banco de germoplasma de eucalipto no Brasil (ANDRADE, 1961). Rapidamente estas áreas se tornaram grandes pólos de fornecimento de sementes para os programas de formação de florestas incentivados pelo governo federal brasileiro entre os anos de 1960 e 1970.

No final da década de 70 e durante os anos 80, foram instalados os primeiros testes de progênes e pomares de sementes para atender a demanda crescente de sementes.

O plantio de eucalipto no Brasil foi imensamente favorecido por diversas razões, entre elas: alta adaptabilidade, elevada produção de sementes e facilidade de propagação vegetativa, crescimento rápido, alta produtividade, boa resposta a tratos culturais de manejo e melhoramento, adequação aos mais diferentes usos e aceitação de mercado (MORA & GARCIA, 2000; SILVA, 2005).

O período de 1986 a 1990 foi caracterizado pelo fim do subsídio para o setor florestal e pela nova fase em que a eficiência nos plantios florestais era a palavra chave. Com isso, as pesquisas básicas de melhoramento foram abandonadas e o enfoque dos produtores passou a ser a propagação de árvores superiores, especialmente de híbridos como, por exemplo, o *E. urograndis* (*E. grandis* x *E. urophylla*) (FERREIRA, 1983). A propagação vegetativa de materiais genéticos superiores permite ganhos de produtividade pelo uso da herdabilidade no sentido amplo, maximizando resultados em florestas resistentes a doenças, com grande homogeneidade e ganhos na qualidade da madeira. O seu emprego nos plantios comerciais de eucalipto possibilitou um salto na produtividade de 8 m³/ha.ano para 45 m³/ha.ano, em um intervalo de 30 anos. Atualmente, as empresas do setor florestal têm recuperado seus programas de melhoramento, mas o plantio clonal continua sendo o carro chefe da eucaliptocultura no Brasil.

Somente nos anos 90 é que os diversos setores florestais passaram a questionar as características desejáveis que a madeira deveria apresentar para um determinado uso final e sua relação com o método de processamento, para assim definir as características que deveriam ser avaliadas nos programas de melhoramento. O termo qualidade da madeira surge então com uma função direta do seu uso final e passa a ser buscado pelo setor, visando uma aproximação da matéria-prima com o processo (RAYMOND, 2002; POT *et al.*, 2002). Assim, na indústria siderúrgica, as características mais importantes da madeira são aquelas que afetam positivamente a qualidade do carvão, tais como elevadas densidades e alto teor de lignina as quais somam maior poder calorífero ao carvão. Para a indústria moveleira, características como uniformidade no sentido medula-casca, baixa retratibilidade, desrama natural e boa usinabilidade, são desejáveis (FERREIRA, 2003). Já a madeira destinada à produção de celulose e papel deve apresentar alto conteúdo de celulose, elevada proporção relativa de fibras e baixo teor de lignina e extrativos, visando diminuir o custo de produção e aumentar a qualidade do produto final (POT *et al.*, 2002).

Entretanto, para a maior parte dos caracteres relacionados à qualidade da madeira há ainda pouca informação disponível sobre a herdabilidade e correlação com caracteres de fácil mensuração e não destrutivos como, por exemplo, a densidade (RAYMOND, 2002).

O setor brasileiro de base florestal participou, em 2004, com 4% do PIB nacional, com faturamento de US\$ 24,6 bilhões, e as exportações brasileiras ligadas ao setor atingiram US\$ 7 bilhões, correspondendo a 7,3% do total exportado pelo país (SBS, 2005). O setor florestal brasileiro é principalmente dividido nos segmentos de papel e celulose, siderurgia e carvão vegetal e indústria moveleira, e lidera o *ranking* dos países produtores de celulose fibra curta (SBS, 2005). Assim, deve-se atribuir a grande competitividade do setor florestal brasileiro no cenário mundial ao sucesso alcançado nas estratégias adotadas pelas empresas do setor.

Comparativamente a outras espécies cultivadas, a domesticação de espécies arbóreas teve início apenas muito recentemente. Além disso, o melhoramento convencional de espécies arbóreas é dificultado por uma série de fatores que em sua maior parte estão relacionados à biologia das árvores, tais como: longo ciclo de vida, maturação tardia, polinização cruzada, elevada heterozigosidade e conseqüente manutenção de alelos deletérios recessivos nas populações que levam a forte depressão por endogamia, além da necessidade de grandes áreas para os plantios experimentais (CAMPBELL *et al.*, 2003). Incrementos ainda maiores na produtividade do setor florestal podem ser alcançados com o emprego de métodos biotecnológicos auxiliares aos programas de melhoramento que facilitem a seleção e produção de genótipos superiores. Neste sentido, a genética molecular oferece uma importante alternativa aos problemas comuns do melhoramento clássico de espécies arbóreas através da seleção assistida, que usa marcadores moleculares para alelos relacionados aos genes alvo, facilitando a seleção de genótipos superiores, e através da transformação genética, que possibilita alterar características da madeira em uma única geração através, por exemplo, da introdução de alelos dominantes (CAMPBELL *et al.*, 2003).

A seleção assistida baseia-se no fato de que para um gene ou um bloco de genes ligado a um marcador genético, a seleção através do marcador deve ser mais eficiente do que para a própria característica. O uso de marcadores moleculares pode possibilitar a seleção precoce de caracteres identificados somente na fase adulta, redução do tempo para obtenção de indivíduos superiores, monitoramento da herança dos alelos favoráveis, caracterização de populações de melhoramento, discriminação e verificação de clones elite e, ainda, a determinação de

paternidade em populações de melhoramento (FERREIRA & GRATTAPAGLIA, 1995). A construção de mapas genéticos para espécies de *Pinus*, *Eucalyptus* e *Populus*, iniciou-se a partir da década de 90 baseando-se na combinação de marcadores dominantes de RAPD e AFLP (RITTER *et al.*, 1990; GRATTAPAGLIA & SEDEROFF, 1994), passando a ser incrementada com informações de marcadores co-dominantes RFLP (BYRNE *et al.*, 1995), e principalmente, de marcadores microssatélites, os quais possibilitaram a revelação de uma quantidade muito maior de polimorfismo permitindo uma maior eficiência na obtenção e validação de QTLs (BRONDANI *et al.*, 1998). BRONDANI *et al.* (2006), construíram o primeiro mapa de ligação de microssatélites para 234 *loci* cobrindo aproximadamente 90% do genoma de *Eucalyptus*. Neste estudo, foi demonstrado um conjunto de 41 microssatélites para genes e QTLs de características da madeira e florescimento representando, portanto, um importante marco na seleção assistida em espécies de eucalipto.

A maioria dos estudos em genética molecular de árvores tem sido focada na alteração da composição química da parede celular, principalmente na quantidade e qualidade do conteúdo de lignina, através da transformação genética (CAMPBELL *et al.*, 2003). Entretanto, estes trabalhos estão principalmente concentrados nas espécies de *Populus*, que possuem importância econômica equivalente à do eucalipto em diversos países europeus. O *Populus* foi a primeira árvore a ser regenerada *in vitro* já no final dos anos 60, e transformada geneticamente (FILLATTI *et al.*, 1987) e a primeira a ter o genoma totalmente sequenciado (TUSKAN *et al.*, 2006). A transformação genética do eucalipto tem sido o objetivo de vários grupos de pesquisa nos últimos anos, mas poucos trabalhos relatam a obtenção de plantas transgênicas de eucalipto (SERRANO *et al.*, 1996; MORALEJO *et al.*, 1998). Somente em 2002, GONZALEZ *et al.*, conseguiram obter plantas transgênicas para a principal espécie de interesse comercial no Brasil, o *E. grandis*, e para o híbrido *E. grandis* x *E. urophylla*. Espera-se agora que características de interesse na madeira do eucalipto possam ser geneticamente modificadas aumentando ainda mais a produtividade do setor florestal brasileiro. De outra forma, diversas plantas transgênicas de *Populus* com redução no conteúdo de lignina da madeira e resistentes á herbicidas já foram produzidas (Revisado em TZVI *et al.*, 1998 e CAMPBELL *et al.*, 2003).

2.2 Genômica e pós-genômica

A genômica pode ser definida como o grupo de estratégias utilizadas na identificação de genes e genomas. O sequenciamento dos primeiros genomas representou a construção de uma base para os passos posteriores de análise da expressão gênica. A era da genômica florestal começou com projetos de EST (*Expressed Sequence Tags*) em *Pinus* e *Populus* e teve como objetivo inicial a identificação de genes específicos de espécies arbóreas (BHALERAO *et al.*, 2003). Os trabalhos pioneiros sequenciaram 5692 ESTs de *Populus* (STERKY *et al.*, 1998) e 1000 ESTs de *Pinus* (ALLONA *et al.*, 1998). Desde então, já estão disponíveis nos bancos de dados públicos (<http://www.ncbi.nlm.nih.gov/>) mais de 359002 ESTs de *Pinus sp.* e mais de 369238 de *Populus sp.* Além disso, estão sendo também depositadas seqüências de ESTs de espécies de *E. globulus*, *E. camaldulensis* e *E. grandis* que já somam aproximadamente 15413 ESTs no NCBI. Outros projetos de sequenciamento de ESTs de *Eucalyptus* foram também concluídos no Brasil (*Genolyptus* e *Forests*), mas em ambos um grande número de empresas do setor florestal estiveram envolvidas inviabilizando a disponibilização destas informações em bancos de dados públicos.

Embora o sequenciamento de ESTs represente um método rápido e barato para identificar genes expressos quando comparado ao sequenciamento genômico, o sequenciamento de genomas completos é também necessário por muitas razões entre elas: genes que apresentam baixa expressão são dificilmente representados em bibliotecas de ESTs, possibilita a construção de mapas genéticos com identificação de QTLs e marcadores moleculares de uso na seleção assistida, permite a comparação entre espécies identificando genes que são comuns às espécies arbóreas (BHALERAO *et al.*, 2003). O primeiro genoma completo seqüenciado de uma espécie arbórea foi concluído para *Populus trichocarpa* (TUSKAN *et al.*, 2006). Para espécies de eucalipto já está disponível o sequenciamento do genoma do cloroplasto de *E. globulus* (STEANE, 2005), e foi iniciado o sequenciamento do genoma de espécies de eucalipto de interesse comercial (*E. globulus*, *E. camaldulensis* e *E. grandis*) pelo consórcio internacional do genoma de *Eucalyptus* (IEuGC) formado em 2004.

Paralelamente a identificação de genomas, a identificação da função de genes é fundamental para a compreensão de processos biológicos tais como a diferenciação celular, morfogênese, determinação fenotípica, resistência a patógenos e adaptabilidade. Assim, a grande quantidade de informações genômicas geradas para espécies de plantas entre elas o *arabidopsis*,

arroz e *Populus*, deram início a era das “ômicas” em plantas. Esta nova era na genética molecular de plantas compreende o estudo de transcritos ou transcrissômica, o estudo de proteínas ou proteômica (CHEN & HARMON, 2006), e o estudo dos metabólitos celulares ou metabolômica (BHALLA *et al.*, 2005), com o objetivo maior de traçar perfis funcionais da célula e/ou tecido, fornecendo uma visão holística da interação genes-fenótipo.

2.2.1 Transcrissômica

O estudo do conjunto de transcritos expressos em uma célula ou tecido representa o primeiro passo na caracterização funcional dos genes e identificação de blocos gênicos co-regulados. As metodologias utilizadas para estabelecer perfis transcricionais baseiam-se na geração de populações de cDNA a partir da população de RNAm expressa, e podem ser divididas em métodos abertos ou globais de análise ou em métodos fechados ou limitados de análise (ALBA *et al.*, 2004). Os métodos globais permitem o acesso à potencialmente todos transcritos expressos em um determinado momento porque não requerem informações prévias de seqüências gênicas para a espécie, já os métodos fechados são limitados pela quantidade de informação disponível para a espécie uma vez que utilizam sondas e/ou *primers* específicos.

Diferentes metodologias estão atualmente disponíveis que permitem uma análise global do transcrissoma, entre elas o sequenciamento de grandes quantidades de ESTs (RONNING *et al.*, 2003), SAGE (VESLCULESCU *et al.*, 1995), *Differential display* (LIANG & PARDEE, 1992), *Supression subtractive hybridization* (SSH) (DIATCHENKO *et al.*, 1996), *Amplified restriction fragment polymorphism* baseado em cDNA (cDNA-AFLP) (BACHEM *et al.*, 1996), e *Massively parallel signature sequencing* (MPSS) (BRENNER *et al.*, 2000).

O sequenciamento de grandes quantidade de ESTs foi o primeiro método utilizado para identificação de genes expressos e para o estabelecimento de perfis de expressão gênica (ADAMS *et al.*, 1991). Os ESTs, no entanto, não representam qualitativamente o conjunto de RNAm expressos já que transcritos com baixo número de cópias não são representados em bibliotecas de ESTs (SUN *et al.*, 2004). Por outro lado, o sequenciamento de ESTs e o acúmulo de informações genômicas são fundamentais em outras abordagens como, por exemplo, para a construção de plataformas de análise baseadas em *arrays*.

O método de *Differential display* foi desenvolvido para aplicação em estudos comparativos de populações de RNAm em diferentes células ou condições. Basicamente, o

differential display consiste na amplificação de RNAm a partir de oligo(dT) *primers* e *primers* 5' curtos de sequência arbitrária, marcação radioativa dos cDNAs amplificados, eletroforese em gel de poliacrilamida, autoradiografia e identificação de transcritos diferencialmente expressos (LIANG & PARDEE, 1992). Embora permita a comparação de um grande número de transcritos entre populações de interesse, o uso dos *primers* pequenos e arbitrários torna a confirmação das sequências de interesse bastante difícil.

Alternativamente, o método AFLP (VOS *et al.*, 1995) está também sendo utilizado para a geração de perfis transcricionais em plantas (BACHEM *et al.*, 1996, RANIK *et al.*, 2006) a partir da síntese de cDNA dupla-fita. A técnica cDNA-AFLP representa um método de identificação de genes diferencialmente expressos de precisão e confiabilidade muito maior que o *differential display* (BACHEM *et al.*, 1996, ALBA *et al.*, 2004). Este sistema pode ser resumido na síntese de uma população de cDNA, restrição, ligação de adaptadores com sítios para amplificação, PCR de alta estrringência, devido ao uso de adaptadores, e seletiva uma vez que uma ou mais bases aleatórias são adicionadas ao *primer*, eletroforese em gel de poliacrilamida, excisão e sequenciamento das bandas de interesse. Uma grande quantidade de transcritos pode ser precisamente acessada neste sistema de análise, embora sua sensibilidade às sequências de baixa frequência seja também muito baixa (ALBA *et al.*, 2004).

Recentemente, um novo método para geração de informação de milhões de sequências expressas denominado *Massively parallel signature sequencing* (MPSS) foi desenvolvido por BRENNER *et al.* (2000). Neste método milhões de *microbeads* são ligados a adaptadores, associados a sequências de cDNA da amostra, e posicionadas em uma plataforma de análise. O sequenciamento das amostras é produzido a partir de passos sucessivos de clivagem com uma enzima de restrição tipoII, que reconhece seu sítio na sequência dos adaptadores, e cliva adiante na sequência da amostra, liberando 4 nucleotídeos em cada passo de clivagem para o sequenciamento. Segundo os autores da técnica, a amplitude do MPSS na geração de dados é muito similar à alcançada através do SAGE, e os testes preliminares realizados com o primeiro método indicam que sequências de 16-20 bases são produzidas a partir da amostra, o que potencialmente resultaria em uma redução da ambigüidade na identificação de genes.

O primeiro estudo que estabeleceu um perfil transcricional para *Eucalyptus* foi apenas recentemente realizado a partir do método SSH (PAUX *et al.*, 2004). Através do uso de uma segunda metodologia bastante precisa de análise quantitativa da expressão de genes específicos, o

Real time-PCR, os autores demonstraram que a construção de bibliotecas por SSH é um método eficiente para a identificação de genes preferencialmente ou exclusivamente expressos em um determinado tecido. O método de subtração de DNA baseado em PCR (SSH) foi desenvolvido para amplificar seletivamente fragmentos de cDNA alvo (diferencialmente expressos) e ao mesmo tempo inibir a amplificação do cDNA que é expresso tanto na amostra quanto no controle (DIATCHENKO *et al.*, 1996). A construção de bibliotecas SSH envolve produção de híbridos de cDNA diferencialmente expressos entre duas amostras, que são formadas somente após uma etapa de hibridização entre a amostra e o controle, ou seja, somente as moléculas que não hibridizaram no primeiro passo (as diferencialmente expressas) estarão disponíveis para hibridização e seleção no segundo passo, onde uma nova quantidade de cDNA desnaturado da amostra é fornecida para hibridização. As populações de cDNA da amostra e controle são ligadas a adaptadores com sítios de anelamento para *primers* específicos no início do processo de forma que os híbridos-alvos formados são especificamente amplificados via PCR, clonados e então seqüenciados.

De todas as metodologias disponíveis de análise de transcritos a metodologia de *microarrays* ou plataformas de chips de nucleotídeos é certamente a mais utilizada em plantas (HERTZBERG *et al.*, 2001, YANG *et al.*, 2003, KO *et al.*, 2004, MOREAU *et al.*, 2005, ANDERSSON-GUNNERÅS *et al.*, 2006). Dois diferentes tipos de *microarrays* podem ser utilizados: os *microarrays* de cDNA ou de oligonucleotídeos. O uso dos *microarrays* de cDNA é muito mais freqüente porque neste caso a plataforma pode ser preparada diretamente a partir de seqüências de cDNA disponíveis nos bancos de dados públicos (DRMANAC & DRMANAC, 1999). Uma outra vantagem desta abordagem é que uma mesma plataforma pode ser utilizada para hibridizar com duas ou mais amostras diferentes usando fluorescências de cores diferentes para produção de perfis comparativos (AHARONI & VORST, 2001). De outra forma, os *microarrays* de oligonucleotídeos apresentam a grande vantagem sobre os *arrays* de cDNA por possibilitarem que um número muito maior de genes sejam avaliados em uma plataforma, e por permitir a hibridização extremamente específica de um único gene com cada oligonucleotídeo. A última vantagem da técnica, no entanto, representa também sua maior desvantagem já que a especificidade da hibridização é gerada pela necessidade de uma grande quantidade de informações de seqüências para o desenho de oligonucleotídeos específicos tornando este método até 5 vezes mais custoso que os *arrays* de cDNA (ALBA *et al.*, 2004).

Ao mesmo tempo em que as metodologias baseadas em *microarrays* oferecem um excelente potencial para análise de milhares de genes, permitindo a identificação de genes diferencialmente expressos, estabelecimento de perfis transcricionais, e quantificações relativas de transcritos, estas metodologias são tanto mais amplas quanto maior a quantidade de informações disponíveis para a espécie em questão. Possivelmente este seja o principal fator limitante para o uso dessa metodologia em espécies de eucalipto, já que nenhum estudo realizado com *microarrays* em espécies de eucalipto foi publicado até então.

2.2.2 Análise serial da expressão gênica (SAGE)

O SAGE (*Serial Analysis of Gene Expression*) é um método de estudo da expressão gênica baseado em sequenciamento que permite identificar quais genes são expressos em um dado momento e quantificar seu nível absoluto de expressão. Este método foi desenvolvido por VELCULESCU e colaboradores em 1995 e resume-se na geração de seqüências curtas de cDNA (14 pb) ou *tags* a partir de uma posição única de cada RNAm da amostra, e na concatenação destas seqüências para clonagem e sequenciamento. A concatenação das *tags* representa um importante passo no processo uma vez que permite que o sequenciamento da biblioteca SAGE seja realizado mais rapidamente e com maior economia quando comparado ao sequenciamento de bibliotecas de ESTs. O método de estudo de expressão gênica a partir de ESTs baseia-se na clonagem de seqüências expressas que representam um único transcrito por clone. Alternativamente, cada seqüência gerada pelo SAGE é o produto de múltiplas *tags* ou transcritos seqüenciados em uma única reação por clone. Assim, quando comparadas bibliotecas de mesmo tamanho, o método SAGE terá permitido identificar um número muitas vezes maior de transcritos (SUN *et al.*, 2004).

O padrão de expressão gênica para qualquer população de transcritos pode então ser estimado qualitativamente e quantitativamente através da identificação dos genes correspondentes às *tags*, e da determinação da abundância de cada *tag* na biblioteca, que nada mais é do que o número de vezes que uma determinada *tag* foi seqüenciada. O ponto crítico do método é a associação de cada *tag* ao transcrito de origem que deve ser realizada através de um banco de dados de seqüências, processo denominado de *tag mapping* (LASH *et al.*, 2000).

Em síntese, a produção de bibliotecas SAGE, baseia-se na extração do RNA poliadenilado da amostra e síntese de cDNA dupla fita, seguida por uma série de etapas posteriores de digestão,

ligação e amplificação, clonagem e sequenciamento (Ver material e métodos). As *tags* são produzidas a partir da clivagem de um ponto único na molécula de cDNA que é definido pelo último sítio de corte da enzima de restrição que antecede o terminal poli(A) fixado às esferas magnéticas. Assim, todo o processo de *tag mapping* deve ser baseado neste fundamento do método, validando-se positivamente somente os *matches* nos quais a *tag* é encontrada imediatamente após o último sítio de reconhecimento da enzima utilizada, antes do terminal poli(A).

Preferencialmente, as *tags* devem ser mapeadas em bancos de seqüências completas de cDNA. Entretanto, além de não haver bancos completos para nenhum organismo, os projetos já realizados priorizam o estudo de humanos, *D. melanogaster* e outros. Um importante recurso que é comumente utilizado para o *tag mapping* é o NCBI SAGEmap (<http://www.ncbi.nlm.nih.gov/sage>). Como descrito por LASH *et al.* (2000), o mapeamento é conduzido cruzando os dados do SAGE a um conjunto de *tags* extraídas de ESTs e seqüências de cDNAs depositadas no NCBI UniGene clusters (<http://www.ncbi.nlm.nih.gov/UniGene/>). Para a extração das *tags*, o SAGEmap utiliza apenas as seqüências expressas (ESTs e cDNAs) com a terminação poli(A) ou com o sinal de poliadenilação. As SAGE *tags* são extraídas da porção mais próxima ao terminal 3' de cada seqüência (determinada pela presença do sítio de restrição da enzima utilizada na geração das *tags*), e as *tags* “virtuais” podem então ser comparadas ao conjunto de *tags* obtido experimentalmente (PLEASANCE *et al.*, 2003).

O estudo do conjunto de *tags* produzidos experimentalmente revela um padrão geral de expressão que pode ser representado como um transcissoma (VELCULESCU *et al.*, 1997). Embora perfis transcricionais de expressão gênica possam ser estudados também através de técnicas de *microarrays* e ESTs, o emprego da tecnologia SAGE torna-se inovador uma vez que fornece acesso a potencialmente todos os genes expressos na amostra, incluindo aqueles ainda não conhecidos, e revela valores absolutos de níveis de expressão; permitindo realizar comparações diretas dos padrões de expressão para diferentes tratamentos de um mesmo tecido ou estádios diferentes do desenvolvimento. Adicionalmente, o uso do SAGE propicia uma maior sensibilidade à análise, permitindo acessar transcritos de baixa abundância, os quais muito provavelmente devem representar reguladores essenciais da expressão gênica, atuantes tanto na transcrição quanto na tradução, e que são facilmente perdidos em outras abordagens (SUN *et al.*, 2004; GIBBINGS *et al.*, 2003). Além disso, as SAGE *tags* podem ser utilizadas como *primers* na

identificação de genes ainda desconhecidos, ou como “sondas” para identificar genes desconhecidos em genomas (COEMANS *et al.*, 2005, CHEN *et al.*, 2002). Por outro lado, é importante salientar que a sensibilidade, precisão e potencial de identificação do SAGE são limitados pela quantidade de clones/*tags* seqüenciados e pelo processo de *tag mapping*, ou seja, pela disponibilidade de seqüências para a espécie em questão. Assim, embora o método não necessite do conhecimento prévio de seqüências para geração das *tags*, outras metodologias devem ser agregadas à técnica para possibilitar a identificação de genes para os quais ainda não há informação disponível.

Atualmente alguns pesquisadores têm usado bancos genômicos para permitir um sucesso maior na identificação das SAGE *tags*. Para isso, é construída uma lista de todas as *tags* possíveis a partir dos dados disponíveis do genoma (*tags* virtuais) da espécie em questão, e esta lista é comparada com a lista de *tags* produzidas experimentalmente, dando ênfase às *tags* virtuais localizadas na região 3' de cada gene. Além de alcançar uma porcentagem maior de identificação de genes, esta abordagem permite ainda evidenciar regulação por *splicing* (FIZAMES *et al.*, 2004; ROBINSON *et al.*, 2004).

As SAGE *tags* podem também ser utilizadas para evidenciar a ocorrência de regulação anti-sense como foi demonstrado em alguns casos (GIBBINGS *et al.*, 2003, GE *et al.*, 2006). Nestes casos, a *tag* identifica uma sequência complementar invertida da fita anti-sense. Os resultados iniciais dos autores acima citados indicam que a ocorrência de *tags* anti-sense é relativamente comum entre as *tags* com baixo número de cópias (1-3 cópias) o que, segundo os autores, é condizente com uma provável função regulatória.

A maior parte dos trabalhos realizados com o SAGE refere-se a estudos de tumores cancerígenos em humanos (NACHT *et al.*, 1999, LEE *et al.*, 2003). A ampla utilização desta metodologia em estudos de câncer deve-se principalmente: a grande quantidade de informações de sequência disponíveis para humanos, o que torna o *tag mapping* eficiente e preciso, à possibilidade de identificar genes tumor-específico que podem ser usados como marcadores no prognóstico e/ou tratamento do câncer, e a possibilidade de compreensão dos eventos envolvidos no desenvolvimento de tumores desde a origem até a metástase.

Analogamente aos estudos de câncer em humanos, o estudo de padrões de desenvolvimento de um determinado organismo requer, da mesma forma, uma metodologia que permita acessar um grande número de genes conhecidos e não conhecidos simultaneamente.

Processos biológicos complexos, tais como os relacionados ao desenvolvimento e diferenciação em plantas, são determinados por um grande número de genes cuja expressão altera a expressão de outros genes de diferentes formas e em diferentes estádios do desenvolvimento.

Apenas poucos pesquisadores empregaram o método SAGE ao estudo da expressão gênica em plantas superiores (MATSUMURA *et al.*, 1999, 2003; LORENZ & DEAN, 2002; GIBBINGS *et al.*, 2003; COEMANS *et al.*, 2005, BAO *et al.*, 2005), e os primeiros trabalhos realizados com esta metodologia na espécie modelo *Arabidopsis thaliana* foram publicados apenas recentemente (JUNG *et al.*, 2003; LEE & LEE, 2003; FIZAMES *et al.*, 2004). O trabalho realizado por MATSUMURA *et al.* (1999) em arroz foi o primeiro trabalho onde o SAGE foi empregado no estudo da expressão gênica em plantas. Os autores foram pioneiros também em demonstrar um método para uma identificação de novos genes mais eficiente e rápido que o reportado no trabalho original de VELCULESCU *et al.* (1995).

A primeira abordagem utilizando o SAGE no estudo de árvores foi publicada por LORENZ & DEAN (2002). Os autores utilizaram para o estudo amostras do xilema do ápice e base do tronco de uma árvore de *Pinus taeda* de 10 anos. Para o ápice, foram geradas 85.000 *tags*, representando 27.398 genes, e para a base do tronco, 65.000 *tags* ou 25.983 genes foram produzidas. 33 *tags* apresentaram níveis de expressão diferentes em pelo menos 5 vezes entre as bibliotecas. Quando analisadas em conjunto, as bibliotecas produziram 150.855 *tags* sendo que destas 42.641 representavam genes distintos.

Os padrões de expressão obtidos a partir do SAGE têm sido confirmados em muitos trabalhos através de outras metodologias tais como o *Northern blotting* e *Microarrays* (JUNG *et al.*, 2003, POROYKO *et al.*, 2005). Em um trabalho realizado com células normais e cancerígenas em humanos onde foram produzidos perfis transcricionais, através do SAGE, e perfis proteicos das duas linhagens de células, foi demonstrado que para 50% dos transcritos superexpressos e reprimidos nas células cancerígenas o mesmo padrão de regulação foi encontrado para as respectivas proteínas (SOUZA *et al.*, 2006).

Alguns autores ainda questionam o potencial de uso das SAGE *tags* na identificação de genes únicos principalmente para genes que pertencem a famílias gênicas (GE *et al.*, 2006), enquanto outros pesquisadores encontram apenas taxas baixíssimas de ambiguidade (BAO *et al.*, 2005). A identificação de mais de um gene para uma única *tag* ocorre basicamente quando seqüências genômicas são utilizadas no *tag mapping* e a *tag* é identificada em regiões mais

internas e não na região 3' da sequência (WEI *et al.*, 2004). Variações do método SAGE originalmente descrito vêm sendo desenvolvidas para facilitar a identificação precisa de genes através do *tag mapping* e para favorecer a caracterização de novos genes através do uso direto da própria sequência da *tag* como *primers* (MATSUMURA *et al.*, 2003, WEI *et al.*, 2004, GOWDA *et al.*, 2004). Inicialmente foi desenvolvido o *SuperSAGE* que se baseia no uso de uma enzima de restrição tipoIII para gerar *tags* maiores que 25 pb (MATSUMURA *et al.*, 2003). Posteriormente, foi desenvolvido o *LongSAGE*, onde *tags* de 20pb são produzidas tanto a partir da região 5' quanto da região 3' dos transcritos, aumentando a especificidade de identificação de genes. Outras modificações foram também propostas para aumentar a eficiência da clonagem e tamanho dos insertos, reduzindo o custo das metodologias que produzem *tags* de maior tamanho (GOWDA *et al.*, 2004).

2.3 Folhas – órgãos fotossintetizantes

As folhas são formadas a partir da diferenciação de células do meristema apical do caule ou *shoot apical meristem* (SAM). O grupo de células do SAM passa por uma série de divisões ao longo de um grande período de tempo, dando origem às células que se diferenciarão nos primórdios foliares das plantas. As células do SAM são heterotróficas, ou seja, não contém clorofila ou cloroplastos e não são capazes de realizar fotossíntese. Ainda não se conhecem os mecanismos moleculares que determinam quais células do SAM se diferenciarão em primórdios foliares, mas sabe-se que este processo é acompanhado pela imediata diferenciação dos plastídios em cloroplastos e aquisição de capacidade fotossintética (FLEMING, 2006).

Os cloroplastos são as organelas onde as reações da fotossíntese ocorrem. A estrutura dos cloroplastos é caracterizada pela presença de um extenso sistema interno de membranas denominado tilacóides. As membranas dos tilacóides quando isoladas formam as lamelas do estroma, e nos muitos pontos de sobreposição formam estruturas denominadas lamelas dos *grana* (Figura 1). Toda clorofila foliar encontra-se posicionada neste sistema de membranas que representa, portanto, o local onde as reações de luz da fotossíntese ocorrem. De outra forma, as reações de redução do carbono assimilado ou ciclo de Benson-Calvin ocorrem no espaço delimitado entre a membrana do cloroplasto e os tilacóides, denominado estroma (LINCOLN & ZEIGER, 2004).

Figura 1 - Esquema da estrutura dos cloroplastos e posicionamento dos complexos proteicos dos fotossistema I e II nas membranas dos tilacóides. Figura adaptada de ALLEN & FORSBERG (2001)

As reações de luz da fotossíntese são catalisadas por quatro complexos proteicos de múltiplas subunidades associados à membrana: fotossistema I (PSI), fotossistema II (PSII), complexo citocromo *b₆f*, e F-ATPase. O centro de reação do PSII, clorofilas do sistema coletor de luz e as proteínas transportadoras de elétrons associadas ao PSII, estão principalmente distribuídas nas lamelas dos *grana*. Já o centro de reação PSI, seu sistema coletor de luz e proteínas transportadoras de elétrons associadas, bem como proteínas do complexo F-ATPase, são quase exclusivamente encontradas nas lamelas do estroma e extremidades das lamelas dos *grana*. O complexo citocromo *b₆f*, que conecta ambos fotossistemas, está distribuído nas lamelas dos *grana* e estroma (NELSON & YOCUM, 2006) (Figura 1).

Ambos PSI e PSII são formados por muitas subunidades proteicas, codificadas por genes nucleares e genes cloroplastidiais, que formam os centros de reação P700 (PSI) e P680 (PSII) e sistema coletor de luz LHCB. O sistema coletor de luz funciona na transferência direta de energia para o centro de reação ao qual está associado (LINCOLN & ZEIGER, 2004). Nos organismos eucariotos fotossintéticos que contém clorofilas *a* e *b*, as proteínas mais abundantes do sistema coletor de luz são membros de uma grande família de proteínas estruturalmente relacionadas. Algumas dessas proteínas estão associadas ao PSII e são denominadas proteínas do complexo coletor de luz II ou *light-harvesting complex II* (LHCII); e outras proteínas são associadas ao PSI e são denominadas LHCI. Conjuntamente estes complexos são também conhecidos como *chlorophyll a/b antenna proteins* (LINCOLN & ZEIGER, 2004).

As reações de luz da fotossíntese podem ser resumidas; (i) na oxidação de duas moléculas de H₂O à O₂ no PSII, produzindo 4 elétrons e 4 prótons, (ii) transferência dos elétrons do PSII

para o complexo *b₆f* e em seguida para o PSI, (iii) PSI reduz NADP^+ à NADPH através da ação da ferredoxina (Fd) e flavoproteína ferredoxina-NADP redutase (FNR), (iiii) e finalmente o gradiente de prótons gerado através da membrana dos tilacóides atua como força propulsora para a síntese de ATP (Figura 2).

O PSII utiliza um *cluster* de íons inorgânicos formados por quatro Mn, um Ca^{2+} , e um Cl^- para oxidar a água e assim produzir elétrons para redução da plastoquinona (Revisado em NELSON & YOCUM, 2006). Os prótons produzidos pela oxidação da água são inicialmente liberados no lúmen dos tilacóides e são posteriormente liberados no estroma através do complexo F-ATPase. Um tipo especial de clorofila, a feofitina, é o primeiro receptor de elétrons a partir do PSII que são então transportados para duas plastoquinonas ligadas ao centro de reação do PSII. A plastoquinona reduzida se dissocia do centro de reação e transfere seus elétrons para o complexo citocromo *b₆f*. O complexo citocromo *b₆f* por sua vez, é formado por múltiplas subunidades proteicas associadas a vários grupos prostéticos e uma proteína denominada *Rieske iron-sulfur*, na qual dois átomos de ferro são ligados a dois átomos de enxofre. Durante a transferência de elétrons, a proteína *Rieske* oxidada recebe elétrons da plastoquinona reduzida e os transfere ao citocromo *f*. O citocromo *f* transfere um elétron para a plastocianina, que reduz PSI. Os receptores de elétrons do PSI são uma série de três ferredoxinas. Finalmente, a flavoproteína ferredoxina-NADP redutase reduz NADP^+ a NADPH, completando a sequência de transporte de elétrons que se inicia com a oxidação da água (LINCOLN & ZEIGER, 2004).

Figura 2 - Transferência de elétrons e prótons na membrana dos tilacóides através dos complexos proteicos PSII, citocromo b_6/f , PSI e ATP-sintase. Figura adaptada de LINCOLN & ZEIGER (2004)

Além da redução de $NADP^+$ a NADPH, a energia solar é também utilizada para síntese de ATP em um processo conhecido como fotofosforilação. O princípio da fotofosforilação baseia-se no mecanismo quimiosmótico, ou seja, as diferenças na concentração de íons e potencial elétrico através dos sistemas de membrana, gerados durante o transporte de elétrons, é utilizado com força propulsora para movimentação de prótons através do complexo F-ATPase e síntese de ATP (LINCOLN & ZEIGER, 2004).

Os produtos das reações de luz da fotossíntese, NADPH e ATP, são consumidos no estroma dos cloroplastos durante as reações de redução do CO_2 assimilado a moléculas de carboidrato através do Ciclo de Belson-Calvin. Durante a fotossíntese, o CO_2 é convertido a trioses-fosfato, que são transportadas para o citoplasma e dirigidas para síntese de sacarose. Parte do carbono reduzido, entretanto, é retida nos cloroplastos para síntese de compostos de reserva tais como o amido, que são utilizados no metabolismo foliar durante a noite e em outros tecidos (MARTIN & SMITH, 1995).

As reações de redução que ocorrem durante a fotossíntese são importantes na regulação do fluxo de carbono para a síntese de açúcares, amido, lipídeos e também para o metabolismo de nitrogênio (Revisado em GEIGENBERGER *et al.*, 2005). Por exemplo, as tiorredoxinas *f* e *m* quando reduzidas ativam a ADP-glicose pirofosforilase (AGPase) que é uma importante enzima

para a síntese de amido nos cloroplastos. Da mesma forma, a enzima acetil-CoA carboxilase (ACCase) que catalisa o primeiro passo na síntese de lipídeos nos plastídeos é também ativada através de modulação-redox. Outro importante exemplo é a ativação por modulação-redox de enzimas envolvidas na reciclagem de N e síntese de aminoácidos durante a fotorrespiração tais como a glutamina sintetase (GS) e glutamato sintase (GOGAT) (Revisado em GEIGENBERGER *et al.*, 2005). Outras evidências indicam também que a ativação da rubisco pela rubisco ativase é também regulada pela relação ADP/ATP e pela luz através da tiorredoxina *f* (Revisado em GRACIET *et al.*, 2004).

Nas plantas C_3 o carbono é inicialmente fixado pela rubisco em duas moléculas de 3 átomos de carbono, o 3-fosfoglicerato, nos cloroplastos das células do mesofilo foliar, como ocorre, por exemplo, nas plantas de *Eucalyptus*. A rubisco (ribulose bifsosfato carboxilase/oxigenase) catalisa a carboxilação de ribulose-1,5-bisfosfato, precursor inicial do ciclo de Benson-Calvin, mas também é capaz de catalisar a oxigenação de ribulose-1,5-bisfosfato em 2-fosfoglicolato e 3-fosfoglicerato durante a fotorrespiração. A atividade carboxilase da rubisco é favorecida pela sua ativação, através da rubisco ativase e tiorredoxina *f*, e também por concentrações maiores de CO_2 . No entanto, na concentração atmosférica atual de O_2 e CO_2 , a fotorrespiração nas plantas C_3 chega a ser responsável pela dissipação de mais de 25% do carbono fixado na fotossíntese (SHARKEY, 1988). Assim, a fotorrespiração é um processo bastante representativo nas plantas C_3 , e embora esteja relacionada à síntese de aminoácidos e reciclagem de amônia, sua vantagem adaptativa ainda é muito discutida (KEYS, 2006).

Como a fotossíntese produz ATP e metabólitos intermediários que são utilizados na síntese dos componentes bioquímicos celulares que compõem as plantas, a modulação de fatores que aumentam a eficiência desse processo pode resultar em aumentos de produtividade. Uma das formas de aumentar a eficiência fotossintética, por exemplo, é reduzindo-se a fotorrespiração (REYNOLDS *et al.*, 2000).

O uso de tecnologias moleculares que possibilitem estudar a importância relativa de cada gene relacionado a estes processos em plantas, bem como comparar perfis transcricionais entre plantas que apresentam variação na capacidade fotossintética e/ou estão sob diferentes condições ambientais, permitirá uma maior compreensão dos mecanismos que regulam a fotossíntese favorecendo a sua modulação.

2.4 A formação da madeira

Uma das características mais importantes que torna a biologia das espécies lenhosas diferente das espécies não lenhosas é o crescimento secundário que leva à formação da madeira. A madeira pode ser definida como o conjunto de células mortas do xilema secundário que são acumuladas ao longo do desenvolvimento das espécies lenhosas. Sua formação resulta de um processo contínuo de diferenciação que se inicia a partir das células do cambio vascular (LARSON, 1994). O câmbio vascular é geralmente descrito como uma camada única de células constituída pelas células iniciais radiais e iniciais fusiformes que se dividem repetidamente em direção à medula, dando origem ao xilema, e em direção à casca dando origem ao floema. Quando uma célula inicial se divide, uma célula mantém as características meristemáticas, enquanto a outra, mais interna, se torna uma célula mãe do xilema (LARSON, 1994). Tanto as iniciais fusiformes quanto as iniciais radiais estão envolvidas na formação do xilema através da formação dos elementos longitudinais, entre eles, os elementos de vaso ou traqueídeos, parênquima longitudinal, e fibras, e dos elementos transversais do parênquima radial, respectivamente (LARSON, 1994).

Uma vez formada, a nova célula do xilema passará por um processo de diferenciação que pode ser dividido em quatro etapas principais: expansão celular, síntese de parede celular (biossíntese e deposição de: celulose, hemiceluloses, proteínas de parede celular, e lignina), lignificação e morte celular programada (LARSON, 1994).

O xilema em desenvolvimento, ao contrário dos tecidos verdes, é caracterizado por um intenso metabolismo heterotrófico que produz a energia necessária para a manutenção das elevadas taxas de crescimento deste tecido, observadas principalmente durante o crescimento secundário. Nos tecidos heterotróficos, as moléculas de carboidratos altamente reduzidas que são produzidas nos tecidos fotossintetizantes, são oxidadas por uma via sequencial de reações dividida em glicólise, ciclo dos ácidos tricarboxílicos (TCA) e respiração aeróbia. Essa sequência de eventos, entretanto, pode apresentar muitas variações devido à necessidade de adaptação das plantas a diferentes condições ambientais (PLAXTON & PODESTÁ, 2006). Por exemplo, a habilidade observada em muitos tecidos de dirigir o piruvato, produto da glicólise, para produção de acetaldeído e etanol, em condições de baixa pressão de oxigênio, tem sido frequentemente associada com a adaptação a condições de hipoxia (KREUZWIESER *et al.*, 2004).

Atividades moderadas das enzimas envolvidas nesse processo, tais como a piruvato descarboxilase e álcool desidrogenase, têm sido também encontradas na região cambial de várias espécies arbóreas em condições normais de crescimento (KIMMERER & SPRINGER, 1988; MacDONALD & KIMMERER, 1991). KIMMERER & SPRINGER (1988) sugeriram que uma baixa pressão de oxigênio é observada no meristema cambial como resultado da forte restrição à difusão de O₂ através da casca das árvores. Assim, a energia para a formação da madeira nas espécies lenhosas é possivelmente suprida pela respiração anaeróbia ou fermentação alcoólica. Os autores sugerem ainda que a participação da álcool desidrogenase na degradação do etanol e conseqüente aumento na produção de ATP seja, possivelmente, um importante mecanismo para otimizar a quantidade de energia liberada pela fermentação alcoólica na região cambial.

A maior parte do carbono fotossinteticamente fixado nas plantas é incorporada aos polímeros que constituem a parede celular (REITER, 2002). Assim, tem-se que aproximadamente 40% a 50% da madeira é formada por celulose, 25% por hemicelulose, 25% a 35% por lignina, e uma pequena porcentagem responde por proteínas de parede celular. A parede celular é formada por diversas camadas sintetizadas em diferentes momentos da diferenciação. A primeira camada a se desenvolver após a divisão celular é a lamela média (PLOMION *et al.*, 2001). A parede celular primária é a segunda camada a ser depositada concomitantemente com a expansão celular (REITER, 2002).

Com o término da expansão celular, inicia-se a síntese da parede celular secundária (REITER, 2002). Esta nova parede é considerada a mais importante em termos de resistência mecânica (PLOMION *et al.*, 2001). Seu componente mais característico, a lignina, é um polímero que estabelece ligações com os demais polímeros da parede celular, reforçando a impermeabilidade das paredes celulares, fundamental à função de transporte dos elementos de vaso, proteção contra patógenos, e suporte mecânico (FUKUDA, 1996). A síntese da parede celular secundária nos elementos de vaso é acompanhada pelo início do processo de morte celular programada durante o qual há perda de todo conteúdo celular levando a formação de uma unidade única de vasos condutores do xilema (GROOVER & JONES, 1999).

A xilogênese é, portanto, um complexo processo de diferenciação controlado basicamente por fatores exógenos que modulam a atividade do cambio, e pela expressão coordenada de uma série de genes (fatores endógenos) estruturais e regulatórios envolvidos nas diferentes etapas da

origem e diferenciação dos elementos celulares do xilema (PLOMION *et al.*, 2001, HERTZBERG *et al.*, 2001).

Os estudos moleculares mais recentes da xilogênese ou formação da madeira têm sido focados em quatro abordagens principais: (i) identificação de novos genes associados à formação do xilema primário e secundário através da indução da xilogênese tanto em *Arabidopsis* quanto em *Zinnia elegans*, (ii) uso de mutantes de *A. thaliana* na caracterização da função de genes sabidamente envolvidos na formação da madeira, (iii) caracterização de padrões de expressão em função de etapas específicas do processo de diferenciação e/ou estádios do desenvolvimento em plantas (*Arabidopsis*, *Populus* e *Pinus*), (iiii) indução da formação de madeira de reação em angiospermas (madeira de tração) e em gimnospermas (madeira de compressão).

O uso crescente de plantas de *Arabidopsis* nos estudos de diferenciação do xilema é estimulado pela observação de que tais plantas, mesmo sendo herbáceas (não apresentam crescimento secundário), são induzidas a formar xilema secundário através da remoção das inflorescências (CHAFFEY *et al.*, 2002; OH *et al.*, 2003). CHAFFEY *et al.* (2002), demonstraram que as fibras e elementos de vaso do xilema formados por indução em *Arabidopsis* apresentavam características estruturais e ultraestruturais similares aos produzidos por uma angiosperma arbórea, o *Populus*, mesmo embora em *Arabidopsis* não sejam detectados raios no xilema secundário nem tão pouco a diferenciação entre lenho inicial e tardio.

OH *et al.* (2003), utilizando *microarrays* em xilema de *A. thaliana*, fizeram a importante observação de que uma série de fatores de transcrição esteve superexpressa no xilema secundário quando comparado ao xilema primário. Esta observação levou os autores a concluir que o crescimento secundário em plantas é uma função da regulação da expressão gênica e não da presença de genes estruturais diferenciados nas plantas lenhosas, o que torna possível a indução do crescimento secundário em espécies não lenhosas. Similarmente, em outro experimento, KO *et al.* (2004), encontraram 700 genes diferencialmente expressos durante a transição do xilema primário para xilema secundário em *Arabidopsis*.

Além da identificação de genes e proteínas envolvidas na diferenciação do xilema, o uso de mutantes em *Arabidopsis* tem propiciado o conhecimento mais profundo das funções destes genes. ZHONG *et al.* (2001), publicaram uma revisão de trabalhos onde mutantes de *Arabidopsis* para a diferenciação das fibras do xilema foram utilizados como modelo para o estudo da diferenciação celular, expansão, e síntese de parede celular. Um exemplo é a brusca redução na

deposição da parede celular secundária nas fibras e células do xilema dos mutantes *irx*. TURNER & SOMERVILLE, (1997) demonstraram que esta redução é um produto direto da redução na deposição de celulose. Mais recentemente, os genes *IRX1* e *IRX3* foram clonados e demonstrou-se que estes genes codificam classes distintas de subunidades catalíticas do complexo celulose sintase (Revisado em ZHONG *et al.*, 2001).

Adicionalmente aos estudos da formação da madeira em plantas de *Arabidopsis*, foi também estabelecido um sistema *in vitro* de cultura de células de *Zinnia elegans* onde células do mesofilo são induzidas a se diferenciar em traqueídeos (FUKUDA, 1996). Mais de 8000 clones de cDNA foram isolados dessas culturas de células e utilizados em uma análise de *microarray* possibilitando a identificação de uma série de genes relacionados à síntese da parede secundária e morte celular programada (DEMURA *et al.*, 2002).

Conforme as seqüências de ESTs de espécies arbóreas começaram a ser disponibilizadas nos bancos de dados públicos, foram realizados os primeiros experimentos de *microarrays* para espécies arbóreas. Certamente, o trabalho desenvolvido por HERTZBERG *et al.* (2001) foi fundamental aos estudos posteriores, uma vez que foi pioneiro em demonstrar que muitos genes relacionados à formação da madeira em *Populus* estão sob controle transcricional restrito a estádios específicos da diferenciação do xilema. Os resultados produzidos por HERTZBERG *et al.* (2001) foram alcançados através do emprego da metodologia de microsecção de células do xilema em diferentes estádios de diferenciação, e análise dos perfis transcricionais das diferentes amostras por *microarrays*. A partir deste trabalho, muitos outros foram realizados utilizando a mesma abordagem de estudo para identificar genes relacionados à formação da madeira em estádios específicos do processo de diferenciação do xilema (OH *et al.*, 2003, SCHRADER *et al.*, 2004, MOREAU *et al.*, 2005).

Mais recentemente, a indução da formação de madeira de reação em espécies arbóreas tem sido utilizada como modelo para estudar as alterações na expressão gênica (PAUX *et al.*, 2005, ANDERSSON-GUNNERÅS *et al.*, 2006). A madeira de reação é normalmente induzida a partir do estímulo de inclinação da árvore, sendo produzida como um mecanismo do tronco para corrigir sua posição original. Nas angiospermas, a madeira de reação é denominada de madeira de tração e nas gimnospermas é denominada como madeira de compressão (TIMELL, 1986). Alterações marcantes nos conteúdos de celulose e lignina são observadas na madeira reação de

forma que este mecanismo natural representa um excelente modelo de estudo sobre a formação da madeira.

Da vasta gama de genes presumivelmente requeridos para a formação da madeira, apenas uma pequena quantidade tem sido aferida em estudos moleculares. O estudo da formação da madeira em árvores em nível molecular teve início apenas recentemente em espécies de *Pinus* (ALLONA *et al.*, 1998), *Populus* (STERKY *et al.*, 1998), *Robinia pseudoacacia* (YANG *et al.*, 2003), e *Eucalyptus* (PAUX *et al.*, 2004), sendo a maior parte dos trabalhos desenvolvidos com a planta herbácea modelo *A. thaliana* (ZHAO *et al.*, 2000; CHAFFEY *et al.*, 2002 e OH *et al.*, 2003).

A identificação de genes envolvidos na formação e determinação das propriedades da madeira deverá aumentar a eficiência e/ou reduzir o tempo e custo necessário para a seleção de genótipos superiores (tradicionalmente requer ensaios químicos e físicos de alto custo inviabilizando o processo de seleção), através da disponibilização de métodos de seleção precoce (FERREIRA & GRATTAPAGLIA, 1995), e de ferramentas para alterar as propriedades da parede celular em árvores geneticamente modificadas (LEE *et al.*, 1998; TSAI *et al.*, 1998; HU *et al.*, 1999). Nesta perspectiva, a identificação dos genes e mecanismos envolvidos na formação da madeira, associada às tecnologias de seleção assistida e modificação genética, representam uma grande oportunidade para a manipulação e agregação de fatores que elevam a qualidade da madeira.

3 OBJETIVOS

O presente trabalho teve como objetivo principal o estabelecimento de um panorama geral do crescimento vegetativo e formação da madeira em *Eucalyptus grandis* e abrangeu também os seguintes objetivos específicos:

- (a) Implementar o uso da técnica SAGE no estabelecimento de perfis transcricionais com a construção de um banco de dados local de transcritos;
- (b) Estabelecer o perfil transcricional de folhas de árvores de *Eucalyptus grandis*;
- (c) Estabelecer o perfil transcricional da região cambial (madeira em formação) de árvores de *Eucalyptus grandis*;
- (d) Comparar os perfis transcricionais de folhas e da região cambial, identificando processos e/ou genes candidatos para o melhoramento;
- (e) Identificar genes exclusivos entre folhas e madeira que apresentam potencial para o desenvolvimento de promotores tecido-específicos.

4 MATERIAL E MÉTODOS

4.1 Material vegetal

O material vegetal foi obtido a partir de árvores jovens (3 anos e 6 anos) de uma população de meio-irmãos de *Eucalyptus grandis* proveniente de um plantio comercial da companhia Suzano – Papel e Celulose. Tanto o material vegetal utilizado para a construção da biblioteca SAGE de madeira, quanto o utilizado para a construção da biblioteca de folhas, foram coletados durante o verão do ano de 2003 em Itapetininga-SP. Um total de 40 indivíduos dessa população foi utilizado para compor as amostras de madeira e folhas. Para a construção da biblioteca de madeira, coletou-se a região cambial das árvores, compreendendo, portanto, não só o meristema cambial, mas também as células do floema e xilema, primário e secundário, mais distantes da medula e consequentemente, potencialmente ainda em diferenciação. O tecido foi coletado utilizando um método não destrutivo que consistiu na remoção de uma janela de 20 x 15 cm da casca das árvores no DAP (diâmetro à altura do peito) seguida pela imediata raspagem do tecido úmido presente na superfície do tronco e na face interna da casca. Como a amostra era rapidamente oxidada após a remoção da casca, o material raspado foi imediatamente acondicionado em nitrogênio líquido (Figura 3).

Figura 3 - Coleta do material utilizado na construção da biblioteca SAGE de madeira

Por fim, todas as amostras da região cambial das 40 árvores foram reunidas em um *bulk* e armazenadas em freezer à -80°C para posterior extração do RNA total. Para a construção da biblioteca de folhas, foram utilizadas folhas jovens e aparentemente saudáveis (aproximadamente 10 folhas) de cada um dos 40 indivíduos. As folhas coletadas foram imediatamente congeladas em nitrogênio líquido e quebradas manualmente, o que otimizou a construção do *bulk* de folhas.

Posteriormente, no laboratório, o *bulk* de folhas foi também armazenado em freezer à -80°C até a extração do RNA total.

4.2 Extração do RNA total

Devido à vulnerabilidade do RNA quando extraído do ambiente celular, todo material de laboratório incluindo vidrarias, tubos plásticos e reagentes foram previamente tratados para eliminação de RNases (ANEXO A). A extração do RNA total das amostras de madeira e folhas foi realizada seguindo o protocolo de extração baseado em fenol descrito por SALZMAN *et al.* (1999) (ANEXO B) com algumas pequenas modificações.

As amostras foram primeiramente maceradas em nitrogênio líquido com o auxílio de um almofariz e pistilo até a obtenção de um pó fino. Um total de 2g dos tecidos macerados foi então utilizado para a extração do RNA total que se inicia com a solubilização do macerado no tampão de extração (ANEXO C). Posteriormente, foram realizadas lavagens com clorofórmio:álcool isoamílico (24:1) até a obtenção de uma interfase límpida o que totalizou três lavagens. As amostras foram então precipitadas com etanol e NaCl para eliminação dos compostos fenólicos ligados ao PVP do tampão de extração.

Outros contaminantes presentes como proteínas e carboidratos foram removidos em passos posteriores através da extração com fenol e precipitação com LiCl, respectivamente. O protocolo originalmente recomenda a solubilização da amostra em 10 mL do tampão de extração, mas, testes de extração anteriores realizados no laboratório, indicaram um melhor rendimento das amostras de madeira quando solubilizadas em 20 mL deste tampão. Além disso, como a construção das bibliotecas SAGE baseia-se na síntese de cDNA, foram conduzidos testes simples de RT-PCR nas amostras de RNA obtidas a partir de madeira e folhas, para verificar o seu potencial para uso em reações enzimáticas. E, embora o dobro do volume do tampão de extração tenha sido utilizado, o RNA obtido a partir da amostra de madeira ainda não apresentava produto de amplificação nítido quando submetido a uma reação teste com *primers* para o gene de eucalipto que codifica a UDP-glicose desidrogenase (previamente desenvolvidos no laboratório). Este resultado foi um forte indicativo da presença de inibidores de reação na amostra como, por exemplo, compostos fenólicos os quais foram eliminados com a introdução de pequenas modificações no tampão de extração.

O tampão utilizado para a extração do RNA da amostra de madeira passou a conter o dobro do volume originalmente recomendado dos agentes antioxidantes, PVP (Polivinilpirrolidona solúvel) e de 2-mercaptoetanol, que atuam de forma a impedir a oxidação dos compostos fenólicos na amostra, impedindo ou minimizando sua ligação aos ácidos nucleicos. Com a introdução destas modificações, o RT-PCR teste, acima mencionado, passou a fornecer um resultado positivo (Figura 4).

Figura 4 - Gel de eletroforese do produto do RT-PCR teste realizado com as amostras de RNA total de folhas (F) e madeira (M) usando *primers* para a UDP-glicose desidrogenase de *E. grandis*. (P) Padrão de DNA (1Kb)

Para a amostra de folhas, nenhuma modificação no protocolo de extração de RNA total foi necessária. Anteriormente à realização das reações de RT-PCR teste, as amostras de RNA total obtidas foram avaliadas quanto a sua integridade, concentração e qualidade. A integridade das amostras foi verificada através de eletroforese em gel de agarose (1%) enquanto sua concentração e qualidade foram estimadas por espectrofotometria nos comprimentos de onda 260nm e 280nm. A concentração de RNA das amostras foi estimada em função da equação determinada por SAMBROOK *et al.* (1989) como segue: $[RNA] = (A_{260nm} \times Fc \times Fd)/1000 \mu g/\mu l$, onde Fc representa o fator de conversão, que indica a equivalência de 1 unidade de absorbância no comprimento de onda 260nm correspondente a 40 μg RNA/ml, e Fd é o fator de diluição da amostra. A ausência de contaminantes nas amostras foi avaliada pela razão entre as absorbâncias (A_{260nm}/A_{280nm}). Quando a razão variava entre 1,8 a 2,0 o RNA obtido era classificado como de boa qualidade, ou seja, livre de contaminantes. Todas as amostras de RNA total obtidas foram separadas em alíquotas de 50 μg , precipitadas em uma solução de 1% acetato de sódio 2,5M e 2X o volume de etanol 100%, e armazenadas em freezer à -80°C até sua utilização na construção das bibliotecas SAGE.

4.3 Construção das bibliotecas SAGE de madeira e folhas

A construção das bibliotecas SAGE de madeira e folhas foi realizada seguindo-se as instruções da metodologia I-SAGE™ *Kit* da Invitrogen a qual é baseada na metodologia de análise serial da expressão gênica (SAGE) desenvolvida originalmente por VELCULESCU *et al.* (1995). O método utilizado, no entanto, inclui algumas modificações para aumentar a eficiência das várias reações enzimáticas do processo e permitir a análise a partir de quantidades menores de RNA. Na Figura 5 é apresentado um esquema geral ilustrativo da metodologia SAGE sendo a descrição de todas as etapas do processo apresentada nos itens subseqüentes.

Figura 5 - Esquema ilustrativo da metodologia SAGE (análise serial da expressão gênica). (A) ligação de moléculas de RNAm poliadeniladas a terminais oligo(dT) ligados a esferas magnéticas e síntese de cDNA dupla fita, (B) digestão com a enzima *NlaIII* que reconhece e corta todo sítio CATG gerando extremidades coesivas utilizadas na posterior ligação de adaptadores que contém um sítio de reconhecimento da enzima *BsmFI* e sítios de anelamento para *primers* específicos, (C) Os fragmentos são divididos igualmente e cada parte é ligada a um adaptador diferente (A e B). (D) Digestão com *BsmFI* proporcionando a clivagem das seqüências 10pb *downstream* em relação ao seu sítio de reconhecimento, (E) ligação das *tags* formando as *ditags*, amplificação via PCR e digestão dos produtos novamente com a enzima *NlaIII* liberando as *ditags* dos adaptadores, (F) concatenação das *ditags* em seqüências de 400 pb - 1 Kb, clonagem e sequenciamento. Esta figura foi adaptada de STOLLBERG *et al.* (2000)

4.3.1 Isolamento do RNA mensageiro e síntese de cDNA

Um total de 50 µg de RNA total de madeira e de folhas foi diretamente utilizado na construção das bibliotecas SAGE. O primeiro passo da metodologia é o isolamento do RNA mensageiro das amostras utilizando partículas esféricas magnéticas associadas à *primers* oligo(dT) (desoxitimidina). Inicialmente 100 µl da solução de partículas magnéticas foram

transferidos para um tubo que foi então posicionado no suporte magnético por 2 minutos para concentração das partículas na parede do microtubo. Removeu-se o sobrenadante e adicionou-se 500 µl do tampão de lise/ligação (*Lysis/binding buffer* ANEXO D). Enquanto isso, as amostras de RNA foram preparadas através da solubilização do *pellet* de RNA em 1000 µl do tampão de lise/ligação. Após o preparo das amostras, o tubo contendo as partículas magnéticas equilibradas no tampão de lise/ligação foi posicionado no suporte magnético e o tampão removido seguido pela adição imediata da amostra de RNA equilibrada no mesmo tampão. Retirou-se imediatamente o tubo do suporte magnético e a solução foi mantida em agitação suave por 30 minutos, à temperatura ambiente, para promover a ligação das moléculas de RNAm da amostra aos oligos(dT) presentes nas partículas magnéticas.

Após a reação de ligação, o tubo foi novamente posicionado no suporte magnético para a realização dos passos de lavagem das partículas associadas ao RNAm que foram realizadas duas vezes com o tampão de lavagem A (*Washing buffer* A ANEXO E), e uma última vez com o tampão de lavagem B (*Washing buffer* B ANEXO F). Em seguida, a amostra foi preparada para a síntese da primeira fita das moléculas de cDNA através de 4 passos de lavagens sucessivos com o tampão de síntese da primeira fita (1X *First strand buffer* ANEXO G). No quarto passo da lavagem, o sobrenadante só foi removido no momento em que a mistura de reação da primeira fita de cDNA (ANEXO H) já estava pronta para adição do RNAm.

As partículas magnéticas ligadas às moléculas de RNAm foram então ressuspensas na solução de reação que foi mantida a 37°C por 2 minutos para pré-aquecimento e em seguida foram adicionados 3 µl da transcriptase reversa *SuperScriptTM* II (ANEXO I) à reação e o tubo foi novamente colocado em banho maria a 37°C por mais uma hora. A reação de síntese da primeira fita foi finalizada colocando o tubo no gelo por 2 minutos antes de iniciar a reação de síntese da segunda fita de cDNA.

Para a síntese da segunda fita, adicionou-se ao tubo 465 µl de água DEPC, 150 µl de tampão para síntese da segunda fita (5X *Second strand buffer* ANEXO J), 15 µl de dNTP *mix* (ANEXO K), 5 µl de *E. coli* DNA ligase (ANEXO L), 20 µl de *E. coli* DNA polimerase (ANEXO M) e 5 µl de *E. coli* RNase H (ANEXO N). A reação foi incubada por 2 horas à 16°C e após o término desse período, finalizou-se a reação através da adição de 45 µl de EDTA 0,5 M. O tubo foi então novamente posicionado no suporte magnético para manter as partículas magnéticas, agora ligadas à moléculas dupla-fita de cDNA, fixas na parede do tubo, enquanto a

mistura de reação foi removida. Adicionou-se às partículas 750 µl do tampão de lavagem C (*Washing buffer* C ANEXO O) pré-aquecido à 75°C para inativar a DNA polimerase. Após 12 minutos, o tampão C foi descartado e uma nova lavagem com o mesmo tampão foi realizada. Em seguida, as partículas foram lavadas 4 vezes sucessivas com 750 µl do tampão de lavagem D (*Washing buffer* D ANEXO P). Posicionou-se o tubo no suporte magnético para remoção do tampão D e adição de 200 µl de 1X tampão 4 (ANEXO Q). A amostra foi então transferida para um novo tubo e armazenada a 4°C por 16 horas antes de seguir para a próxima reação.

4.3.2 Digestão do cDNA com a endonuclease *Nla* III e ligação dos adaptadores

A amostra de cDNA solubilizada em 1X tampão 4 foi novamente precipitada com o auxílio do suporte magnético, e ressuspensa na mistura de reação de digestão contendo 172 µl de LoTE (ANEXO R), 2 µl de 100X BSA (ANEXO S), 20 µl de 10X *buffer* 4 e 6 µl de *Nla* III (ANEXO T). A reação foi mantida por 1 hora a 37°C. Após este período, a mistura de reação foi descartada usando o suporte magnético e as partículas magnéticas ligadas às moléculas de cDNA foram lavadas duas vezes com 750 µl do tampão de lavagem C para inativar a enzima, seguida de quatro lavagens com 750 µl do tampão de lavagem D.

Em seguida, foi realizada a ligação dos adaptadores às moléculas de cDNA. Para isto, as partículas magnéticas na amostra foram primeiramente lavadas duas vezes com 150 µl do tampão de ligação 1X (ANEXO U) e após a última lavagem, a amostra foi homogeneizada em 100 µl do mesmo tampão e dividida em dois novos tubos A e B. Ambos tubos foram posicionados no suporte magnético para descarte do tampão. No gelo, foi adicionado às amostras a mistura de ligação dos adaptadores contendo 14 µl de LoTE, 2 µl do tampão para ligase 10X concentrado, 1,5 µl do adaptador A (20 ng/µl) no tubo A e 1,5 µl do adaptador B (20 ng/µl) no tubo B. Os tubos foram aquecidos por 2 minutos a 50°C, resfriados por 15 minutos à temperatura ambiente e imediatamente colocados no gelo para a adição de 2,5 µl de T4 DNA ligase (ANEXO V).

As amostras foram incubadas por 2 horas a 16°C e após este período foram lavadas quatro vezes com 500 µl do tampão de lavagem D. Posteriormente, as amostras foram lavadas duas vezes com 200 µl do tampão 4 1X concentrado, ressuspensas neste mesmo tampão e armazenadas a 4°C por 16 horas até o próximo passo que consistiu na clivagem dos cDNA + adaptadores com a endonuclease tipo II *BsmF* I (ANEXO W).

4.3.3 Digestão com a endonuclease *BsmF I* e formação dos dímeros de cDNA

Os tubos A e B foram posicionados no suporte magnético e o sobrenadante descartado. Em seguida, adicionou-se a mistura de digestão contendo 174 µl de LoTE, 20 µl do tampão 4 10X concentrado e 4 µl de BSA 100X concentrado. As amostras foram pré-incubadas a 65°C por 2 minutos. Imediatamente após, adicionou-se 2 µl da enzima *BsmF I* e a mistura foi então incubada a 65°C por uma hora.

A digestão com *BsmF I* resulta na liberação dos cDNA + adaptadores das partículas magnéticas, assim, após o término da digestão, os tubos foram posicionados no suporte magnético e o sobrenadante foi recuperado em dois novos tubos. Um PCR-teste (ANEXO X) foi realizado a partir do produto da reação de digestão que permaneceu ligado às partículas magnéticas para testar a eficiência da digestão com *BsmF I*. Como esperado, não houve sinal de banda no gel após a digestão com *BsmF I*.

O sobrenadante recuperado no passo anterior contém seqüências curtas de cDNA + adaptadores a partir das quais serão produzidas as *ditags* ou etiquetas duplas de cDNA futuramente utilizadas na anotação dos genes encontrados. Assim, foi adicionado às amostras um volume de fenol : clorofórmio (1:1) para extração dos cDNAs. Os tubos foram centrifugados em velocidade máxima por 5 minutos a temperatura ambiente e a fase aquosa de cada tubo foi transferida para dois novos tubos. Em seguida, precipitou-se o cDNA com 133 µl de acetato de amônia 7,5 M, 3 µl de *mussel glycogen* e 1000 µl de etanol 100%. As amostras foram precipitadas em gelo seco por somente 10 minutos e o *pellet* recuperado após a centrifugação (16000g por 40 minutos a 4°C) foi lavado duas vezes com etanol 70%. Removeu-se todo o etanol dos tubos, os *pellets* foram ressuspensos em 10 µl de LoTE e incubados a 37°C por 3 minutos para garantir a solubilização.

A cada um dos tubos, A e B, adicionou-se 5 µl do tampão *Klenow* (ANEXO Y), 1 µl de BSA 100X, 2,5 µl de uma mistura de dNTP (10 mM de cada), 30,5 µl de água DEPC e 1,0 da polimerase *Klenow* (ANEXO Z). As misturas foram incubadas por 30 minutos a 37°C e imediatamente reunidas em um único tubo com volume final de 100 µl para promover a ligação das seqüências de cDNA em dímeros. Posteriormente, foi realizada uma nova etapa de extração e precipitação dos dímeros com fenol:clorofórmio (1:1) e acetato de amônia 7,5M/etanol 100%, respectivamente, como descrito anteriormente. Durante a extração com fenol : clorofórmio foram removidos 100 µl da fase aquosa e transferidos para um novo tubo. Estes 100 µl de amostra

foram também precipitados e utilizados mais adiante como controle negativo no próximo PCR-teste.

O *pellet* final nos dois tubos (amostra e controle) contendo os dímeros de cDNA + adaptadores foram ressuspensos em 1,5 µl de LoTE e incubados a 37°C por 15 minutos para solubilização. Em seguida, foram adicionados a mistura de reação de ligação 1,25 µl de Tris-HCl 3mM pH 7,5 no tubo da amostra e 2,25 µl no tubo do controle negativo, 0,75 µl do tampão de ligase 10X concentrado, 0,75 µl de água DEPC e 1 µl de T4 DNA Ligase somente no tubo contendo a amostra. Ambos tubos foram mantidos *overnight* a 16°C.

4.3.4 Otimização das reações de PCR para produção de dímeros em larga escala

Após o período de incubação, foi adicionado 14 µl de LoTE a cada mistura de reação que foram então diretamente utilizadas nas reações de otimização de PCR. As reações de otimização foram realizadas com 3 diferentes diluições da amostra objetivando determinar a melhor diluição que resulte em quantidade suficiente de amplicons para seguir para o próximo passo de polimerização de dímeros. A escolha da diluição adequada foi realizada comparando-se a intensidade do produto de amplificação com o controle positivo fornecido no *kit* o qual continha uma concentração de 50-250 ng/50 µl que representa justamente a concentração necessária do produto após 27 ciclos de amplificação para garantir uma eficiente amplificação do produto. Assim, o controle negativo foi diluído 1/20 e a amostra foi diluída em 1/20, 1/40 e 1/80. As reações de otimização foram realizadas seguindo a Tabela 1.

Tabela 1- Otimização das reações de PCR a partir de diferentes concentrações da amostra

Regentes	Tubos					
	1	2	3	4	5	6
10X BV Buffer (ANEXO AE)	5 µl	5 µl	5 µl	5 µl	5 µl	5 µl
DMSO	3 µl	3 µl	3 µl	3 µl	3 µl	3 µl
dNTP mix (10 mM cada)	7,5 µl	7,5 µl	7,5 µl	7,5 µl	7,5 µl	7,5 µl
DTP-1(<i>primer</i> de anelamento ao adaptador, 175 ng/µl)	2 µl	2 µl	2 µl	2 µl	2 µl	2 µl
DTP-2	2 µl	2 µl	2 µl	2 µl	2 µl	2 µl
Água DEPC	29 µl	29 µl	29 µl	29 µl	29 µl	29 µl
Platinum®Taq DNA polimerase	0,5 µl	0,5 µl	0,5 µl	0,5 µl	0,5 µl	0,5 µl
Amostra 1/20	1 µl	-	-	-	-	-
Amostra 1/40	-	1 µl	-	-	-	-
Amostra 1/80	-	-	1 µl	-	-	-
Controle positivo	-	-	-	1 µl	-	-
Controle negativo (sem ligase) 1/20	-	-	-	-	-	1 µl

*Tubos 1-3: diferentes diluições da amostra, 1/20, 1/40 e 1/80; tubo 4 : controle positivo fornecido no *kit*, tubo 5: controle negativo (sem DNA) e tubo 6: controle negativo (sem ligase). Foram realizados 27 ciclos seguindo o programa: um ciclo de 2 minutos a 95°C; 27 ciclos com 30 segundos a 95°C, um minuto a 55°C e um minuto a 70°C; e por fim um ciclo de 5 minutos a 70°C.

Os produtos de PCR (5 μ l) foram separados em gel de poliacrilamida 12% (ANEXO AA), corados com brometo de etídeo (0,5 μ g/ml) e visualizados sob luz ultravioleta. O produto de reação que mais se aproximou ao produto do controle positivo foi o obtido a partir da diluição 1/40 (segunda coluna no gel) (Figura 6).

Figura 6 - Gel de eletroforese dos produtos das reações de otimização da concentração de DNA para PCR. (1) *Ditags* diluídas 1/20. (2) *Ditags* diluídas 1/40. (3) *Ditags* diluídas 1/80. (P) Padrão de DNA. (C) Controle positivo fornecido no *kit*

4.3.5 Amplificação em larga escala e purificação dos dímeros de cDNA

Foram realizadas 200 reações de PCR com volume final de 50 μ l cada uma contendo 5 μ l de tampão BV 10X concentrado (ANEXO AB), 3 μ l de DMSO, 7,5 μ l de dNTP *mix* (10 mM de cada), 2 μ l de cada *primer* DTP-1 e DTP-2 (175 ng/ μ L em água DEPC para cada *primer*), 24 μ l de água DEPC, 0,5 μ l de *Platinum®Taq* DNA polimerase e 6 μ l da amostra diluída 1/40. O programa de PCR utilizado foi o mesmo descrito anteriormente. Os produtos das 200 reações foram reunidos em um único tubo e procedeu-se com as etapas de extração e precipitação dos dímeros usando fenol:clorofórmio (1: 1) e acetato de amônia 7,5M/etanol 100% conforme já descrito anteriormente. O *pellet* resultante da precipitação foi ressuscitado em 300 μ l de LoTE e incubado por 10 minutos a 37°C.

Em seguida, foi realizada a purificação do volume total de dímeros em 9 colunas de um gel de poliacrilamida 12 %. Durante a etapa de formação dos dímeros, diferentes tipos de dímeros poderiam ser formados na reação como, por exemplo, dímeros de cDNA contendo somente adaptadores A ou B nas extremidades, dímeros de cDNA contendo o adaptador A em uma extremidade e o B em outra, ou ainda dímeros formados apenas pelos adaptadores. No entanto, na etapa de amplificação dos dímeros foram usados *primers* específicos, um para o adaptador A e outro para o adaptador B, o que levou a amplificação somente dos dímeros formados por ambos

adaptadores garantindo assim que os dímeros amplificados fossem resultantes das duas populações de cDNA A e B, inicialmente separadas para ligação dos adaptadores. O tamanho esperado dos dímeros adequados é de aproximadamente 100 pb (41 pb de cada adaptador + 10 pb de uma molécula de cDNA + 10 pb da outra molécula de cDNA) e baseando-se neste princípio selecionou-se no gel a banda de aproximadamente 100 pb para posterior polimerização. No gel é possível ainda observar duas bandas mais fracas correspondendo a 80 pb e 60 pb aproximadamente, e que representam produtos com falha na ligação dos adaptadores (Figura 7).

Figura 7 - Gel de eletroforese dos produtos da amplificação em larga escala dos dímeros de cDNA. Todas as bandas de 100 pb (círculo) foram excisadas do gel para geração das *ditags*. (1 – 9) Amostras de dímeros de cDNA. (P) Padrão de DNA

As bandas correspondentes a 100 pb foram recortadas do gel usando um bisturi e colocadas em tubos de 0,7 ml (furados no fundo) encaixados em tubos de 1,7 ml para centrifugação (14000 g por 3 minutos) o que levou ao rompimento da porção de gel em frações menores facilitando assim a dissolução da poliacrilamida. Os fragmentos de gel foram dissolvidos em 150 µl de LoTE e acetato de amônia 7,5M na proporção 125:25.

Após homogeneização, as amostras foram incubadas por 2 horas a 65°C. Após este período, as amostras foram combinadas três a três e passadas em uma coluna S.N.A.PTM que foi centrifugada a velocidade máxima por 2 minutos. A solução eluída a partir da centrifugação de cada coluna (aproximadamente 300 µl) foi recuperada em um tubo para precipitação com acetato de amônia 7,5M/etanol 100%. Os *pellets* resultantes da precipitação foram então ressuspensos em 14 µl de LoTE, reunidos em um único tubo com volume final de aproximadamente 126 µl e armazenados a -20°C *overnight*.

4.3.6 Digestão dos dímeros de cDNA com *Nla* III

Nesta etapa, a digestão dos dímeros cDNA + adaptadores com a enzima *Nla* III promove a liberação dos adaptadores em ambas extremidades do dímero já que a mesma enzima foi utilizada no primeiro passo de clivagem do cDNA e assim, seu sítio de reconhecimento estará posicionado imediatamente antes das seqüências de adaptadores em ambas extremidades do dímero. A digestão resulta na liberação de seqüências de 28 pb, os dímeros de etiquetas ou *ditags*.

O primeiro passo nesta fase consistiu na divisão dos 126 µl de amostra em três diferentes tubos com o mesmo volume final seguido pela adição da mistura de reação de digestão contendo: 42 µl de *ditags*, 15 µl de 10X tampão 4, 2 µl de 100X BSA, 12 µl de *Nla* III e 79 µl de água DEPC. Os três tubos com a mistura de reação de digestão foram incubados a 37°C por 2 horas. Após este período o volume de cada tubo foi ajustado para 200 µl com LoTE. Em seguida, procedeu-se com a extração e precipitação das *ditags* usando fenol:clorofórmio (1:1) e acetato de amônia 7,5M/etanol 100%, respectivamente, conforme já descrito em passos anteriores. Os *pellets* dos três tubos foram ressuspensos em LoTE e combinados em um único tubo com volume final de 32 µl.

As *ditags* de 28 pb foram separadas dos produtos não digeridos em de gel de poliacrilamida 12%. Para isto, foi adicionado 7 µl do tampão de amostra TBE (ANEXO AC) ao tubo contendo 32 µl da solução de *ditags* e todo volume da solução foi aplicado em três colunas no gel (Figura 8).

Figura 8 - Gel de eletroforese dos produtos da digestão com *Nla* III. (1 - 3) amostras de madeira. (P) Padrão de DNA. (4 - 6) amostras de folhas. Todas as bandas de 28 pb foram excisadas do gel para formação dos cancatâmeros

Após a eletroforese, o gel foi corado com brometo de etídio (0,5 µg/ml) e visualizado sob luz UV permitindo assim a excisão das bandas de 28 pb. Os fragmentos de gel foram quebrados em pequenos pedaços, usando o mesmo procedimento já descrito para a purificação dos dímeros de cDNA, e homogeneizados em uma solução contendo 150 µl de LoTE/acetato de amônia 7,5 M (125:25). O tubo contendo os fragmentos em solução foram incubados a 37°C *overnight*.

4.3.7. Polimerização/Concatenação e clonagem das *ditags*

A concatenação das *ditags* resulta na formação de seqüências maiores de cDNA que serão utilizadas para clonagem propiciando, portanto, que em um único evento de sequenciamento, cerca de 14 - 58 moléculas de cDNA diferentes sejam acessadas. Para isto, as *ditags* eluídas do gel foram purificadas, precipitadas e então submetidas a reação de ligação como segue.

A solução eluída do gel foi purificada em colunas S.N.A.PTM e o volume resultante foi dividido em volumes de 200 µl aliqüotados em diferentes tubos para precipitação com etanol. A cada tubo foram adicionados 90 µl de acetato de amônia 7,5 M, 3 µl de *mussel glycogen* e 850 µl de etanol 100% sendo a precipitação conduzida em gelo seco por 20 minutos. Posteriormente, os tubos foram centrifugados a velocidade máxima por 30 minutos a 4°C e os *pellets* resultantes foram lavados com etanol 70% e ressuspensos em um único tubo com 7,75 µl de LoTE.

Adicionou-se a solução com as *ditags* 1 µl do tampão de ligase 10X concentrado e 1,25 µl de T4 DNA ligase. As amostras foram incubadas por 3 horas a 16°C e após este período, foram adicionados 12 µl do tampão de amostra TBE 5X concentrado à mistura de reação que foi então mantida por 10 minutos a 65°C. Todo o volume da amostra foi submetido a eletroforese em gel de poliacrilamida 8% em uma única coluna. No gel já corado com brometo de etídio foi possível visualizar um rastro que se inicia com um produto de aproximadamente 100 pb indo até 1 Kb (Figura 9).

Figura 9 - Gel de eletroforese dos produtos da ligação das *ditags*. (F) Concatâmeros da amostra de folhas. (P) Padrão de DNA. (M) Concatâmeros da amostra de madeira. As frações excisadas do gel foram de 500 pb – 1 Kb

Foram selecionados para excisão no gel os fragmentos contendo polímeros de 500 pb - 1 kb e estes foram quebrados em pedaços menores através de centrifugação como mencionado anteriormente. Cada fração eluída do gel foi então solubilizada em dois tubos contendo 200 µl de LoTE/acetato de amônia 7,5 M (5:1) e incubada a 65°C por 2 horas. O conteúdo dos dois tubos (de cada amostra) foi purificado em uma coluna S.N.A.P.TM, centrifugado por 2 minutos a velocidade máxima, e a solução purificada foi dividida em volumes de 200 µl para extração e precipitação dos concatâmeros de *ditags*. Os passos de extração e precipitação foram realizados conforme já descrito em passos anteriores e o *pellet* resultante foi solubilizado em 6 µl de LoTE.

Anteriormente a clonagem dos concatâmeros de *ditags* procedeu-se com a linearização do vetor pZerO[®]-1 usando a enzima *Sph* I na seguinte mistura de reação: 2 µl do vetor (1 µg/µl ANEXO AD), 2,5 µl do tampão 2 10X concentrado (ANEXO AE), 1,4 µl da enzima *Sph* I (ANEXO AF) e 19,1 µl de água estéril. A solução foi incubada a 37°C por 30 minutos e após este período procedeu-se com a extração e precipitação do vetor linearizado. O *pellet* resultante foi ressuspendido em 60 µl de LoTE que foram então utilizados na reação de ligação do vetor aos concatâmeros. A reação de ligação foi realizada com dois controles negativos conforme mostrado na Tabela 2.

Tabela 2- Reação de ligação do vetor aos polímeros de *ditags*

Reagentes	Amostra	Controle negativo (sem DNA)	Controle negativo (sem ligase)
Concatâmeros	6 µl	-	-
pZER [®] -1/ <i>Sph</i> I	1 µl	1 µl	1 µl
Tampão de ligase 10X	1 µl	1 µl	1 µl
T4 DNA ligase	2 µl	2 µl	-
Água estéril	-	6 µl	8 µl

Os três tubos com a mistura de ligação foram incubados por 3 horas a 16°C. Após o término da reação o volume dos tubos foi ajustado para 200 µl e procedeu-se diretamente com a extração e precipitação dos vetores usando fenol:clorofórmio (1:1) e etanol 100%. Os *pellets* foram ressuspendidos em 12 µl de LoTE e utilizados na transformação das células de *E. coli* para construção das bibliotecas SAGE.

Um volume de 2 µl da mistura de ligação foi utilizado para transformar 50 µl de células de *E. coli* eletrocompetentes *One Shot TOP10* (Invitrogen). A mistura de células mais vetores foi transferida para uma cuveta de 0,1 cm e eletroporada com um pulso elétrico de 2,5 kV por 4 milissegundos. Após a eletroporação, adicionou-se 250 µl do meio SOC (ANEXO AG) à mistura que foi então mantida sob agitação constante (200 rpm) a 37°C por 1 hora. Após este período, adicionou-se ao tubo mais 750 µl do mesmo meio e a solução resultante foi dividida em volumes de 100 µl e plaqueada em placas contendo meio LB sólido (ANEXO AH) e o agente de seleção ZeocinaTM (50 µg/ml ANEXO AI).

A solução remanescente com as células transformadas foi armazenada a 4°C até o dia seguinte para verificar o resultado do plaqueamento. As placas foram mantidas a 37°C *overnight* e no dia seguinte pode-se observar um grande número de clones somente nas placas com as amostras (aproximadamente 150 clones/placa). Assim, todo o volume restante foi também plaqueado e as colônias formadas foram selecionadas para o sequenciamento.

4.4 Sequenciamento dos polímeros de *ditags*

A extração de DNA plasmidial e sequenciamento dos plasmídeos recombinantes foram realizados em colaboração com o laboratório de Biologia Celular e Molecular do CENA-USP sob supervisão da Dra. Siu Mui Tsai. As extrações foram realizadas usando a metodologia *PureLink Plasmid Purification System* (Invitrogen) e o sequenciamento foi realizado através do sistema *BigDye[®] Terminator* usando o *ABI PRISM[®] 3100 Genetic Analyzer* (Applied Biosystems).

4.5 Bioinformática e análise dos dados

O estabelecimento de perfis de expressão gênica a partir dos dados de sequenciamento de bibliotecas SAGE depende primeiramente do estabelecimento de uma associação correta entre cada *tag* e a sequência de cDNA ou EST de origem. Assim, foram utilizados neste trabalho 4 critérios para análise dos dados: (a) seleção de sequências com qualidade de sequenciamento igual ou superior a $q \geq 20$, (b) seleção das *tags* com número de cópias igual ou superior a 3 em cada uma das bibliotecas, (c) utilização de um banco de dados de ESTs de origem específica de espécies de eucalipto, (d) anotação somente daqueles ESTs que apresentaram o último sítio de reconhecimento da enzima *Nla* III (CATG) imediatamente seguido pela *tag* e sequência de poliadenilação 3'. Os critérios adotados na geração dos dados bem como na sua análise serão apresentados nas seções que seguem.

4.5.1 Seleção de sequências para análise

Os dados de sequenciamento gerados foram avaliados quanto a sua confiabilidade através do programa *Phred* (EWING *et al.*, 1998). O programa *phred* lê os sinais de fluorescência nos cromatogramas gerados pelo sequenciador, determina as posições ideais dos picos de cada base, compara as localizações dos picos detectados com as dos picos idealizados, quantifica a qualidade do sinal (baseando-se na intensidade do sinal, espaçamento entre os picos e na presença de picos mal resolvidos em volta da base em questão) e atribui valores de qualidade às bases previstas nos arquivos de saída. O valor de qualidade está relacionado com a probabilidade de erro nas atribuições das bases através da fórmula: $q = -10 \times \log(p)$, onde p é a probabilidade de erro estimado (EWING *et al.*, 1998). Para o presente trabalho foram utilizadas apenas as sequências com qualidade igual ou superior a 20, o que significa que 1 a cada 100 pb sequenciados possa estar errada, assim, a probabilidade de atribuição de bases corretas é de 99%. As sequências de qualidade inferior foram re-sequenciadas até a obtenção de sequências com qualidade adequada (Figura 10).

Figura 10 - Cromatogramas de dois eventos de sequenciamento de um mesmo SAGE clone. (a) Cromatograma obtido a partir de um primeiro sequenciamento de baixa qualidade onde é possível observar a ocorrência de sobreposição de picos com áreas pouco definidas. (b) Cromatograma referente ao re-sequenciamento da mesma sequência de (a), porém, com alta qualidade, sendo possível observar picos bem delimitados e não sobrepostos

4.5.2 Extração das SAGE tags

Após a análise de qualidade dos dados de sequenciamento, somente as sequências com qualidade $q \geq 20$ e com no mínimo 500 pb (tamanho mínimo dos concatâmeros de *ditags* selecionados para clonagem) foram submetidas ao *software* de análise SAGE2000TM versão 4.5 (<http://www.invitrogen.com/sage>). Este *software* é um programa apropriado para análise dos dados de sequenciamento dos concatâmeros de *ditags* usados na construção das bibliotecas SAGE. O programa importa os arquivos de sequências, identifica os sítios de restrição da enzima *Nla* III (CATG), extrai as *tags* usando os sítios reconhecidos como referência, e contabiliza a ocorrência de cada *tag* na biblioteca. Ao final, o programa lança um arquivo de saída ou *tag list* contendo as sequências de 10 pb das *tags* seguidas do seu nível de expressão que nada mais é do que o número de vezes em que aquela *tag* foi sequenciada na biblioteca (Figura 11).

```

                                Date: 08-29-2006
Report File Name = C:\Documents and Settings\Guillermo\Meus documentos\SAGE\R2006\rep-5.rpt

Project # 1
Project File = C:\Documents and Settings\Guillermo\Meus documentos\SAGE\R2006\R_2006.sum
SearchName = R_2006
Anchoring Enzyme = NlaIII - CATG
Tag Length = 10
DiTag Length = 26
Total Files = 1154
Total Tags = 39964
Total Duplicate Dimers = 1146

Total tags in selected Projects = 39964

Tag Abundance Report
MS Access File Name = C:\Documents and Settings\Guillermo\Meus documentos\SAGE\SAGE.MDB

Total tags after excluding tags = 39964 → Nº de cópias da tag/gene ou nível de expressão

Count      Percent      Tag Sequence      Tag BaseFour Number → Sequência da tag/gene
518         1,2961      TACTTCAGCT        818728
Genes in Class = 1 * Tags in Class = 518
Cumulative Gene Count = 1 * Cumulative Tag Count = 518

338         0,8457      CAAAAATTAC        262386
Genes in Class = 1 * Tags in Class = 338
Cumulative Gene Count = 2 * Cumulative Tag Count = 856

274         0,6856      GTGGATCGTG        762735
Genes in Class = 1 * Tags in Class = 274
Cumulative Gene Count = 3 * Cumulative Tag Count = 1130

273         0,6831      CACATCCCTT        281952
Genes in Class = 1 * Tags in Class = 273
Cumulative Gene Count = 4 * Cumulative Tag Count = 1403

238         0,5955      TCGCGGTACT        891592
Genes in Class = 1 * Tags in Class = 238
Cumulative Gene Count = 5 * Cumulative Tag Count = 1641

```

Figura 11 - Arquivo de saída do *software* SAGE2000™ (versão 4.5) ou *Tag list* da biblioteca de folhas. A *Tag list* contém uma lista das *tags* sequenciadas na biblioteca com seu respectivo nível de expressão

Para a comparação de perfis de expressão entre duas ou mais bibliotecas, o programa armazena os dados individualmente de cada biblioteca e permite o lançamento de um arquivo de saída com a *tag list* contendo *tags* encontradas em todas bibliotecas bem como *tags* exclusivas, seguidas do seu nível de expressão em cada uma das bibliotecas (Figura 12).

Date: 08-29-2006

This Compare Report was saved as C:\Documents and Settings\Guillermo\Meus documentos\SAGE\com-16.cmp

A Microsoft Access file was saved as C:\Documents and Settings\Guillermo\Meus documentos\SAGE\SAGE.MDB

Sorted by sum of tags.

Library # 1 = 30503 tags from C:\Documents and Settings\Guillermo\Meus documentos\SAGE\R2006\R_2006.sum

Library # 2 = 34327 tags from C:\Documents and Settings\Guillermo\Meus documentos\SAGE\Dani-Mayra\TAGBANK.sum

Total tags in all projects = 64830

Output = Observed tags, NonNormalized

Lib # 1	Lib # 2	Total	Tag Sequence
518	122	640	TACTTGAGCT
88	472	560	AATTTCCCAG
338	56	394	CAAAAATTAC
50	309	359	CAGTTAAAAG
274	62	336	GTGGATCGTG
164	163	327	TGTGTGTTTG
209	97	306	TAATACCATC
273	28	301	CACATCCCTT
22	273	295	TACTTATTTG
52	231	283	ATTAGTGGTG
74	195	269	GCATTGCATA
238	30	268	TCGCGGTACT

Diagram annotations:

- Arrow from "Output = Observed tags, NonNormalized" to "Nº de cópias da tag/gene na biblioteca de folhas"
- Arrow from "Output = Observed tags, NonNormalized" to "Nº de cópias da tag/gene na biblioteca de madeira"
- Arrow from "Tag Sequence" to "Seqüência da tag/gene"

Figura 12 - Arquivo de saída do *software* SAGE2000TM (versão 4.5) ou *Tag list* comparando as bibliotecas de folhas e madeira. A *Tag list* contém uma lista das *tags* sequenciadas em ambas bibliotecas com seu respectivo nível de expressão em cada biblioteca

As *tags* lançadas nos arquivos de saída do *software* SAGE2000TM foram então utilizadas para encontrar as sequências de cDNA ou ESTs que provavelmente lhes deram origem e assim, estabelecer o perfil de expressão de cada tecido (Figura 13).

Figura 13 - Resumo das etapas de geração e análise das bibliotecas SAGE. A análise das bibliotecas SAGE inicia-se com o sequenciamento dos clones e geração das *tag lists* que são então utilizadas nas associações *tag/gene*. Por fim, foi construído um banco local de dados para armazenamento de informações de sequências expressas e posterior estabelecimento de um perfil comparativo da expressão gênica entre amostras de folhas e madeira

Somente as *tags* com número de cópias ≥ 3 em cada uma das bibliotecas foram utilizadas nas análises, já que o número mínimo de cópias para resultar em diferenças significativas entre as duas bibliotecas foi igual a 3, segundo o teste estatístico aplicado (ver seção seguinte).

4.5.3 Comparação dos níveis de expressão das *tags* entre as bibliotecas

Duas características dos dados de experimentos realizados com a metodologia SAGE foram consideradas na escolha do teste estatístico para análise: (a) a ausência de repetições o que, consequentemente, impossibilita estimativas de desvio e variância e (b) a geração de bibliotecas de tamanhos (número total de *tags* sequenciadas) diferentes. Assim, foi escolhido para a análise dos dados o teste Z (KAL *et al.*, 1999) que é um teste estatístico que considera o número de cópias de uma determinada *tag* de um tecido como uma proporção do número total de *tags* observadas “N” para aquele tecido, permitindo, portanto, a comparação entre bibliotecas de tamanhos diferentes.

Com esta abordagem, o número de *tags* encontrado segue uma distribuição binomial uma vez que resulta da proporção de cada *tag* na biblioteca (KAL *et al.*, 1999). No entanto, quando um grande número de *tags* é sequenciado, esta distribuição binomial tende a se aproximar muito de uma distribuição normal com a proporção estimada para uma dada *tag* ou “p” representando uma média e o desvio da média (S_p) representado por:

$$S_p = \sqrt{p(1-p)}$$

A acurácia da estimativa das proporções depende do número total de *tags* sequenciadas onde $SE_p = S_p / \sqrt{N}$, sendo o intervalo de confiança da abundância observada para cada *tag* “n” calculado como 99% ($P \leq 0,01$) IC de $n = (p \pm 2,67 \times SE_p) \times N$. Finalmente, a diferença no número de cópias de uma dada *tag* entre duas bibliotecas $p_1 - p_2$ e a estimativa de $SE_{p_1 - p_2}$ são utilizadas para calcular o teste estatístico como demonstrado na equação 1:

$$Z = \frac{p_1 - p_2}{\sqrt{p_0(1-p_0)/N_1 + p_0(1-p_0)/N_2}}$$

Onde p_0 é a proporção calculada da *tag* quando não existe diferença entre as duas bibliotecas, ou seja, a hipótese nula é verdadeira. p_0 é calculado como $p_0 = (n_1 + n_2) / (N_1 + N_2)$. Na Figura 14 está representado um resumo da planilha utilizada para comparar os perfis de expressão das duas bibliotecas através do teste Z.

6	Tag_Sequence	Folha	Madeira	Total	p1	p2	p0	z	decisão 95%	decisão 99%
7	TACTTGAGCT	518	122	640	0,0130	0,0028	0,0079	16,530	Rejeita ho	Rejeita ho
8	AATTTCCAG	88	472	560	0,0022	0,0109	0,0066	-15,543	Rejeita ho	Rejeita ho
9	CAAAAATTAC	338	56	394	0,0085	0,0013	0,0049	14,829	Rejeita ho	Rejeita ho
10	CAGTTAAAAG	50	309	359	0,0013	0,0071	0,0042	-13,128	Rejeita ho	Rejeita ho
11	GTGGATCGTG	274	62	336	0,0069	0,0014	0,0041	12,173	Rejeita ho	Rejeita ho
12	TGTGTGTTTG	164	163	327	0,0041	0,0038	0,0039	0,782	FALSO	FALSO
13	TAATACCATC	209	97	306	0,0052	0,0022	0,0037	7,066	Rejeita ho	Rejeita ho
14	CACATCCCTT	273	28	301	0,0068	0,0006	0,0037	14,609	Rejeita ho	Rejeita ho
15	TACTTATTTG	22	273	295	0,0006	0,0063	0,0034	-14,193	Rejeita ho	Rejeita ho
16	ATTAGTGGTG	52	231	283	0,0013	0,0053	0,0033	-10,111	Rejeita ho	Rejeita ho
17	GGATTGCATA	74	195	269	0,0019	0,0045	0,0032	-6,792	Rejeita ho	Rejeita ho
18	TCGCGGTACT	238	30	268	0,0060	0,0007	0,0033	13,181	Rejeita ho	Rejeita ho

Figura 14 - Planilha *Excel* do teste Z comparando os níveis de expressão de *tags*/genes entre as bibliotecas SAGE de folhas e madeira. Quando o teste acusa diferença significativa entre o número de cópias de uma determinada *tag* entre as bibliotecas, a hipótese nula é rejeitada, “Rejeita ho”, quando o teste acusa que não há diferença significativa no número de cópias entre as bibliotecas a hipótese nula é aceita, “FALSO”

4.5.4 Encontrando as sequências de origem das *tags* e suas prováveis funções celulares

As sequências das *tags* ≥ 3 cópias por biblioteca foram utilizadas para encontrar as sequências de cDNA de origem no banco de dados de ESTs do projeto genolyptus e/ou no banco de dados públicos do NCBI. O projeto genolyptus foi lançado em 2002 como uma parceria de 12 empresas do setor de papel e celulose, 7 Universidades, a Embrapa e o Ministério da Ciência e Tecnologia. O genolyptus foi formado a partir do sequenciamento de 70893 ESTs que resultaram na construção de 23292 *contigs* representativos de amostras de folhas, plântulas, raízes, xilema, floema e flores das espécies de eucalipto: *E. globulus*, *E. grandis*, *E. pellita* e *E. urophila*.

O contraste das sequências das *tags* e as sequências de ESTs foi realizado usando a ferramenta de bioinformática BLAST-N que compara sequências de nucleotídeos com uma base de dados também formada por sequências de nucleotídeos (ALTSCHUL *et al.*, 1997). O resultado do BLAST-N fornece uma lista das sequências de ESTs onde é possível encontrar toda a sequência usada no BLAST-N, ou seja, os 4 pb do sítio de reconhecimento da enzima *Nla* III (CATG) seguidos pelos 10 pb da *tag*. No entanto, não fornece informações importantes como a presença de outros sítios CATG após a *tag* e se há ou não uma cauda poli(A) 3' indicando término da sequência. Assim, todas as sequências de ESTs buscadas foram analisadas manualmente e somente aquelas que obedeceram aos critérios acima mencionados foram utilizadas na anotação das *tags* (Figura 15).

```

>Contig940gr-ml (whole contig)
5'CTCCTCTTTTGGAGTTGTCAGGAGAGGAGCTGCAGAAGGGAAAATGGCTGCCGCAACGGCATCTC
CCATGGCCAGTCAGCTCAAGAGCAGCTTTGCATCATCATCCAGGAGCCTGATTGCTCCCAAGG
GCCTCTCTGGTTTCCTTCAGAGCCTCGCCACCAGGAGGACTCGCTCCTTCACTGTCAAGGCTAT
CCAATCCGAGAAGCCAACCTTCCAAGTGGTTCAACCCCTCAATGGTGACCCTTTCATTGGAAGCCT
CGAGACTCCGGTCACATCCAGCCCTCTCATTGCATGGTACCTCTCCAACCTCCCCGCCTACAGGAC
TGCCGTGAACCCGCTCCTCCGAGGCGTCGAGGTGGGCTCGCCACGGATTCTCTTGGTGGGCCC
CTTCGTCAAGGCCGGCCCTTTGAGGAACACGGAGTATGCAGGAGCAGCAGGATCATTGGCCGCCG
GAGGCTGGTTGCGATCCTGAGCATCTGCTTGACAATGTATGGGATCGCCTCATTCAATGAAGGAG
AGCCCTCGACAGCACCGTCGCTAACCTTGACCGGGAGGAAGAAGGAGCCTGATCAGTTGCAGACC
GCAATGGATGGGCCAAGTTCACCGCGGGTCTTCTTCGGCGGGATCTCTGGGGTCACATGGGCC
TACTTTCTCCTCTATGTGCTCAACCTTCCTTACTTTGTGAAGTAAATTCATACAGAAGAGAGAGATG
AAGTTGAAGAGTCATGTCTGTATATCTGAACCGTTTGGGAAGGAATGGGATTGAGGAAATTTCAT
GTCATTTAGCTCTGTATAAGCCGGAATTTCCCTGGACTTTGTCAAAGAATGTGCATTTCGTAGAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
GG 3'

```

Figura 15 - Sequência ilustrativa de um EST utilizado na anotação de uma *tag*. Em negrito estão todos os possíveis sítios de restrição da enzima *Nla* III, em negrito e sublinhada está a sequência da *tag* específica utilizada no blast(n) e na extremidade 3' da sequência é possível observar a presença de um terminal poli(A)

Além disso, a biblioteca de madeira foi anotada quase que exclusivamente a partir de ESTs de amostras de xilema e floema enquanto a biblioteca de folhas foi anotada apenas preferencialmente a partir de ESTs de folhas devido a pequena representatividade de ESTs de folhas depositados no genolyptus e ausência nos bancos de dados públicos. Todo o procedimento de anotação das *tags* foi realizado como dois eventos separados, primeiramente na anotação das *tags* da biblioteca de madeira e depois na anotação das *tags* da biblioteca de folhas, garantindo assim que os resultados fossem também utilizados na análise do potencial de uso de sequências curtas de 10 pb como “etiquetas” para a associação única e correta entre a *tag* e o gene.

As sequências de ESTs representativas de cada *tag* foram então usadas no programa BLAST-X (ALTSCHUL *et al.*, 1997) tendo como banco de referência o NCBI. O BLAST-X é usado para comparar sequências de nucleotídeos traduzidas em todos os três quadros possíveis de leitura contra sequências de proteínas. O resultado do BLAST-X é uma lista de alinhamentos possíveis ordenados de acordo com a confiabilidade do alinhamento a qual é medida pelos valores de *E*. Estes valores indicam a probabilidade de um dado alinhamento de pontuação igual ou maior ocorrer ao acaso numa dada base de dados, ou seja, quanto maior o valor de *E* maior a probabilidade do alinhamento ser um produto do acaso (ALTSCHUL *et al.*, 1997). Neste trabalho, apenas alinhamentos com valores de $E \leq e-07$ foram utilizados na anotação das *tags*.

Para a realização de uma análise funcional e mais dinâmica dos genes expressos em cada uma das bibliotecas foi realizada também uma classificação funcional dos genes anotados em categorias seguindo a classificação originalmente proposta por RISON *et al.* (2000) (Tabela 3)

com pequenas modificações como a inclusão da subcategoria transporte de vesículas e fatores transportadores na categoria transporte, e a inclusão da subcategoria citoesqueleto na categoria estrutura e organização da estrutura celular. Além disso, os ESTs que representaram *tags* anotadas com a mesma função foram utilizados em análises de alinhamento (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>) para verificar a ocorrência de isoformas gênicas ou genes distintos.

Tabela 3 - Classificação Funcional de RISON *et al.* (2000). Os números representam as funções-chave ou principais

1 Metabolismo	4 Estrutura e organização da estrutura celular
1.1 Energia	4.1 Envoltórios celulares/Membranas
1.1.1 Metabolismo autotrófico	4.1.1 Membrana celular
1.1.2 Metabolismo de carbono	4.1.2 Parede celular
1.1.3 Transferência de energia/ATP	4.2 Matrix extracelular
1.2 Macromoléculas	4.2.1 Estruturas de superfície
1.2.1 Polissacarídeos/Lipossacarídeos	4.3 Ribossomos
1.2.2 Proteoglicanas/Glicoproteínas	4.3.1 RNAs ribossomais
1.2.3 Lipídeos/Lipoproteínas	4.3.2 Proteínas ribossomais
1.3 Pequenas moléculas	4.4 Citoesqueleto
1.3.1 Metabolismo de aminoácidos	4.4.1 Relacionados ao cromossomo
1.3.2 Metabolismo de Nitrogênio	4.4.2 Relacionados às organelas
1.3.3 Metabolismo de nucleotídeos	
1.3.4 Metabolismo de fósforo	5 Vias de Informação
1.3.5 Metabolismo de carboidratos	5.1 Relacionada ao DNA
1.3.6 Metabolismo de Lipídeos	5.1.1 Síntese e Replicação
1.3.7 Vitaminas/co-fatores/grupos prostéticos	5.1.2 Restrição/modificação/reparo
1.3.8 Metabolismo Secundário	5.1.3 Recombinação
	5.1.4 Degradação
2 Processos	5.2 Relacionada ao RNA
2.1 Processos celulares	5.2.1 Síntese
2.1.1 Divisão celular	5.2.2 Transcrição
2.1.2 Tradução de sinais	5.2.3 Modificação
2.1.3 Direcionamento de proteínas	5.2.4 Degradação
2.1.4 Regulação celular	5.3 Relacionada a proteínas
2.2 Processos do organismo	5.3.1 Síntese de proteínas
2.2.1 Adaptação	5.3.2 RNAt
2.2.2 Respostas de proteção/detoxificação	5.3.3 Relacionada à tradução
2.2.3 Respostas a estímulos	5.3.4 Modificação/fosforilação
	5.3.5 Modificação/estrutura quaternária
3 Transporte	5.3.6 degradação
3.1 Grandes moléculas	
3.1.1 proteínas	6 Outros
3.1.2 Ácidos nucleicos	6.1 Não classificado (sequência expressa)
3.2 Pequenas moléculas	6.2 Não classificado (proteína de função hipotética)
3.2.1 Canais de íons/porinas	6.3 Sem função conhecida (sem homologia)
3.2.2 Transportadores de açúcares e carboidratos	
3.2.3 Transportadores de aminoácidos	
3.2.4 Transportadores de nucleotídeos/nucleosídeos	
3.2.5 Fatores Transportadores	
3.2.6 Transporte de Vesículas	

Por fim, os dados gerados no sequenciamento e análise das bibliotecas SAGE foram armazenados como “fichas eletrônicas” para *tags* dentro de grupos em um banco de dados para análise de genômica funcional de *E. grandis*, construído localmente com o auxílio de Livia Maria Franceschini (bio-informata do laboratório de Max Feffer de genética fisiológica) (Figura 16).

Laboratory Max Feffer - Ensembl Database - SAGE Data Base - Search Result

USP ESALQ

[exit](#) [contact](#)

a) Search Result

TAG:

Registered Groups: Groups that TAG is registered:

Description:

Blast Result:

Functional Domain:

Category:

Sequence:
 GGACGCTGCGGTACGGTCCGGACGCGTG
 GGCGGACGCGTGGCGGACGCGTCCGGAG
 AGGCTAAGAGGGAGGGAAGCTTTGAG

[Alter](#) [Delete](#) [List](#)

b) Search Result

TAG:

Registered Groups: Groups that TAG is registered:

Description:

Blast Result:

Functional Domain:

Category:

Sequence:
 GAGAGAGGCTAAGAGGGAGGGAGAGCT
 TTGAGAAAGATAAGGTAAACACAGTTTT
 GCTTGCCGGGCTGTGCTCGAGTTTGGCCT

[Alter](#) [Delete](#) [List](#)

Figura 16 - Fichas “eletrônicas” do banco de dados local de genômica funcional de *E. grandis*. (a) Ficha da *tag* com informações cadastradas para a biblioteca de madeira. (b) Ficha da *tag* com informações cadastradas para a biblioteca de folhas. Para a mesma *tag* foram criadas duas fichas diferentes buscando a melhor associação *tag*/gene para cada um dos tecidos avaliados. Observe que a ficha do banco de madeira foi construída a partir de um EST com a mesma origem (gl-xy: xilema de *E. globulus*) enquanto a ficha do banco de folhas foi construída a partir de um EST sequenciado a partir de folhas maduras de *E. grandis* (gr-ml)

5 RESULTADOS E DISCUSSÃO

5.1 Resumo dos resultados produzidos através da metodologia SAGE

As bibliotecas SAGE de Madeira e Folhas foram produzidas a partir de 1440 e 1154 clones, respectivamente. Cada clone sequenciado gerou em média 32 sequências de *tags* resultando em um total de 43304 *tags* sequenciadas a partir da biblioteca de madeira, e 39964 *tags* sequenciadas a partir da biblioteca de folhas. Como a identificação foi realizada somente para as *tags* com número igual ou superior a 3 cópias por biblioteca, somente 26958 *tags* da biblioteca de madeira e 22981 *tags* da biblioteca de folhas foram consideradas. Neste grupo final de *tags* apenas 3066 corresponderam a *tags* diferentes ou “genes” na biblioteca de madeira, e 2798 na biblioteca de folhas. Estas *tags*/genes foram então buscadas em sequências públicas de ESTs de *Eucalyptus* (NCBI) e entre as sequências de *Eucalyptus* produzidas pelo consórcio Genolyptus, permitindo assim a identificação de 1272 genes na biblioteca de madeira e 975 na biblioteca de folhas, ou seja, 38% do número total de *tags*/genes observados foram identificados (Tabela 4).

Tabela 4 - Resumo dos dados gerais produzidos a partir das bibliotecas SAGE

Dados gerais	Biblioteca Madeira	Biblioteca Folhas	Total
Total de clones sequenciados	1440	1154	2594
Tamanho da biblioteca (tags sequenciadas)	43304	39964	83268
Tags com $3 \geq$ cópias	26958	22981	43939
Total de genes com $3 \geq$ cópias	3066	2798	5864
Genes anotados	1272	975	2247
Total de genes diferencialmente expressos ($P \leq 0,01$)	-	-	464
Genes exclusivos ($P \leq 0,01$)	47	64	111
Genes preferencialmente expressos na biblioteca de madeira ($P \leq 0,01$)	-	-	252
Genes preferencialmente expressos na biblioteca de folhas ($P \leq 0,01$)	-	-	212

Trabalhos anteriores que também empregaram a metodologia SAGE para análise de transcritos em outras espécies alcançaram 50% de identificação em folhas e sementes de arroz (GIBBINGS *et al.*, 2003), 28% em folhas de *arabidopsis* (JUNG *et al.*, 2003), 45% em grãos de pólen de *arabidopsis* (LEE & LEE, 2003), 33% a partir do sistema nervoso periférico em humanos (DE JONGE *et al.*, 2003) e 70% de identificação a partir de células gástricas cancerígenas em humanos (LEE *et al.*, 2003). Esta baixa porcentagem de identificação de genes através do método SAGE está relacionada à pequena representatividade de ESTs de baixa frequência nos diferentes bancos de dados, que pode ser ilustrada pelo exemplo da experiência em humanos. No início do sequenciamento de ESTs em humanos um grande número de novos

genes puderem ser identificados (cerca de 10,4% dos ESTs sequenciados em 1996 representavam novos genes), mas, a medida que o sequenciamento progredia, o número de novos genes identificados passou a decrescer (2,7% dos ESTs sequenciados em 1998 e 1,6% dos ESTs sequenciados entre 1998 e 2001 representavam genes novos) (WANG *et al.*, 2000; CHEN *et al.*, 2002). Quando então as *tags* geradas pelo método SAGE passaram a ser utilizadas em estudos de expressão gênica em humanos, pode-se observar que para um grande número dessas *tags* não era possível encontrar as respectivas sequências de cDNAs ou ESTs de origem nos bancos de dados (DE JONGE *et al.*, 2003). Estas observações associadas à observação de que o método SAGE foi utilizado com sucesso na geração de novos genes ainda não identificados em humanos (CHEN *et al.*, 2002), e de que a metodologia SAGE é 26 vezes mais eficiente do que os ESTs para a detecção de transcritos de baixa abundância (SUN *et al.*, 2004), apontam para uma grande diferença na sensibilidade do método utilizado na construção dos bancos de dados (EST) em relação ao método SAGE, resultando, portanto, na dificuldade de identificação das SAGE *tags* em bancos de ESTs.

Este problema tem sido solucionado por alguns pesquisadores que utilizam bancos genômicos para permitir um sucesso maior na identificação das SAGE *tags*. A estratégia baseia-se na construção de uma lista de todas as *tags* possíveis a partir dos dados disponíveis do genoma (*tags* virtuais) da espécie em questão, e da comparação desta lista com a lista de *tags* produzidas experimentalmente, dando ênfase às *tags* virtuais localizadas na região 3' de cada gene. Porcentagens de identificação de 57% a partir de raízes de *arabidopsis* (FIZAMES *et al.*, 2004) e 79% a partir de folhas de *arabidopsis* (ROBINSON *et al.*, 2004) foram alcançadas com esta nova estratégia, no entanto, sua grande limitação é a ausência de dados genômicos para muitas das espécies de interesse, como é o caso para as espécies de *Eucalyptus*.

É também importante salientar que quando nós terminamos o sequenciamento das bibliotecas SAGE, não havia ainda informações de ESTs de *Eucalyptus* nos bancos de dados públicos e todo o trabalho de anotação foi primeiramente realizado apenas com base no banco de ESTs do consórcio Genolyptus cujas informações não são disponibilizadas ao público. Somente no ano de 2006 com a publicação de um grande número de sequências de ESTs de amostras de xilema de *Eucalyptus gunnii* no banco de dados públicos do NCBI nós pudemos então buscar as *tags* ainda não identificadas e as identificadas como genes importantes na formação da madeira e nos processos foliares, em sequências públicas. Assim, embora seja lamentável que os

pesquisadores que trabalham com eucalipto no Brasil, tenham que buscar informações geradas por grupos estrangeiros, mesmo possuindo informações tão relevantes e enriquecedoras, essa experiência foi muito importante para o nosso laboratório, pois pudemos confirmar as anotações previamente realizadas para um grande número de *tags* a partir de sequências de uma fonte diferente. Esta confirmação é muito relevante uma vez que demonstra a eficiência da técnica SAGE em acessar transcritos de forma única, mesmo a partir de sequências curtas de 14pb, quando é claro, o organismo a partir da qual as *tags* foram geradas é o mesmo que o utilizado na geração do banco de referência.

Outra importante observação realizada a partir dos nossos resultados foi a de que para uma mesma *tag* encontrada em ambas bibliotecas e anotada com sequências de ESTs originadas a partir de tecidos diferentes (madeira ou folhas), foi identificada uma mesma função, em todas as situações avaliadas. Conforme descrito no Material e Métodos, as *tags* independentemente de serem iguais ou não, foram anotadas em fichas separadas para a biblioteca de Madeira e para a biblioteca de Folhas e, como o banco de dados de ESTs do *genolyptus* conta tanto com sequências obtidas a partir de folhas quanto com sequências produzidas a partir de amostras do floema e xilema, esta abordagem de análise nos permitiu acessar esta importante informação. Uma alta especificidade na identificação de genes a partir de SAGE *tags* de 14pb foi também reportada por BAO *et al.* (2005) os quais alcançaram um percentual de 90% de *tags* que identificam um único gene, dentro de famílias gênicas. Por outro lado, GE *et al.* (2006) encontraram 13,2% de *tags* inespecíficas (que encontraram sequências com funções diferentes no banco de dados do UniGene) em bibliotecas de células normais e cancerígenas de humanos, e reportam ter conseguido 90% de identificação não-ambígua para essas *tags* quando sequências de tecidos específicos para cada biblioteca foram utilizadas na identificação. Duas hipóteses são possíveis para explicar esta discrepância: (i) as *tags* produzidas no experimento de GE *et al.* (2006) foram buscadas em sequências gênicas do UniGene, o que muito provavelmente pode ter levado a validação de *tags* internas que não necessariamente representam o transcrito acessado pela *tag*, (ii) como as comparações dos resultados de identificação de uma mesma *tag* nas bibliotecas de folhas e de madeira foi possível somente para 50% das *tags*/genes da biblioteca de folhas que apresentaram homologia a sequências de ESTs de folhas, e para 86% das *tags*/genes da biblioteca de madeira que apresentaram homologia a sequências de ESTs de madeira disponíveis no *genolyptus*, as *tags* que possivelmente poderiam gerar dados inespecíficos

estariam entre aquelas de baixa frequência não representadas no banco de referência e, portanto, excluídas da análise.

Após a análise das *tags lists* produzidas, as bibliotecas de madeira e folhas foram contrastadas, sendo o número de cópias das *tag/gene* encontradas em cada biblioteca comparada estatisticamente através do teste Z. A análise estatística revelou que das 9861 *tags* únicas/genes (produzidas a partir do número total de *tags* nas duas bibliotecas), 1163 apresentaram diferença significativa entre as bibliotecas com pelo menos 95% de confiança e 464 genes foram diferencialmente expressos com 99% de confiança (Tabela 4). Entre estes últimos genes, 47 foram exclusivos da biblioteca de madeira, 64 da biblioteca de folhas, 252 foram significativamente mais expressos na biblioteca de madeira e 212 na biblioteca de folhas (Tabela 4). Os genes preferencialmente expressos em cada tecido serão discutidos dentro das seções que caracterizam o perfil transcricional de folhas e madeira.

Entre os genes exclusivos muitos apresentaram um baixo número de cópias (7 – 13 cópias) o que impossibilitou a identificação de 80 genes ou 72% dos genes exclusivos, a partir dos ESTs. Por outro lado, foi possível identificar os outros 31 genes de expressão exclusiva na biblioteca de folhas ou de madeira. Estes genes representam um grande potencial para identificação de promotores de expressão tecido-específica a serem empregados em projetos de transformação genética de *Eucalyptus*. O estudo desses promotores candidatos já vem sendo desenvolvido no laboratório Max Feffer de Genética de Plantas da Esalq e estes resultados permanecerão em sigilo até que o estudo seja concluído.

Para que os genes encontrados fossem analisados em uma perspectiva dos processos biológicos de cada tecido, nós procuramos agrupá-los em categorias funcionais e também relacioná-los aos processos metabólicos mais importantes em cada tecido como demonstrado nas próximas seções.

5.2 Perfil transcricional da biblioteca de folhas

A distribuição dos genes em categorias funcionais revelou que a maior parte dos genes expressos na biblioteca de folhas está relacionada à categoria Metabolismo e Energia (21%) seguida pelas categorias Estrutura e Organização da Estrutura Celular (17%), Vias de Informação (15%), Processos Celulares (14%) e por fim, a categoria de Transporte (6%) (Figura 17). Na categoria Metabolismo e Energia estão genes que atuam na formação dos complexos proteicos

que formam os fotossistemas I e II, e que participam das reações de luz da fotossíntese bem como da síntese de ATP, e genes relacionados à assimilação de CO₂ em carboidratos através do ciclo de Benson-Calvin. Foram também representativos nessa categoria os genes relacionados ao processo de fotorrespiração, ao metabolismo de nitrogênio, aminoácidos e metabolismo secundário. Na categoria Estrutura e Organização da Estrutura Celular os genes de maior representatividade foram àqueles relacionados à RNA ribossomais sugerindo uma alta atividade de síntese proteica no material amostrado. De forma condizente com o resultado anterior, a categoria Vias de Informação foi também representada por um grande número de genes relacionados à síntese/degradação/modificação de proteínas. Claramente, os genes mais representativos na categoria Processos Celulares foram àqueles relacionados à proteção e destoxificação celular.

Figura 17 - Distribuição dos genes identificados a partir da biblioteca SAGE de folhas em categorias funcionais. A categoria Sem Função Definida foi representada por *tags* para as quais foram encontrados os prováveis ESTs de origem, mas, não foi possível identificar a função destas sequências através do NCBI

Na Tabela 5 é fornecida uma lista dos 50 genes mais expressos na biblioteca de folhas sendo possível observar que a maior parte dos transcritos identificados estão relacionados à fotossíntese (41,46%), proteção contra o estresse oxidativo (20,34%) e fotorrespiração (6,22%).

Tabela 5 - Lista dos 50 genes mais expressos na biblioteca de Folhas. Para cada *tag* é fornecido o número de cópias na biblioteca de Folhas, a identificação do gene bem como o valor de *E* para cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente, mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou cDNA de origem

<i>Tag</i>	Nº de cópias	Identificação	Valor <i>E</i>
1	518	Sem registro	-
2	338	gil2196772 chlorophyll a/b-binding protein	2e-136
3	274	gil77157637 ribulose-1,5-bisphosphate carboxylase	2e-82
4	273	gil1223652 metallothionein	2e-05
5	238	gil11095210 lipid transfer protein 3 precursor	1e-34
6	209	gil46850189 metallothionein 3a	2e-15
7	170	Sem registro	-
8	164	Sem registro	-
9	129	gil3023743 Ferredoxin-1	1e-43
10	124	gil3036945 light harvesting chlorophyll a/b-binding	6e-139
11	117	gil5608497 galactinol synthase	3e-142
12	103	gil1773330 glycolate oxidase	0.0
13	103	Sem registro	-
14	101	Sem registro	-
15	99	Sem registro	-
16	96	gil19911 photosystem II	2e-105
17	96	gil225895 photosystem I	2e-113
18	92	gil14422331 senescence cysteine protease	9e-130
19	88	gil10764491 allergenic isoflavone reductase	e-145
20	85	Sem registro	-
21	85	gil15810139 unknown protein	e-176
22	84	gil3694984 metallothionein-1-like protein	2e-28
23	80	Sem registro	-
24	78	gil5541885 galactinol synthase	1e-66
25	74	gil39653273 ADP-ribosylation factor	9e-100
26	74	Sem registro	-
27	73	Sem registro	-
28	66	gil49642980 unnamed protein product	3e-20
29	66	gil18399660 fructose-bisphosphate aldolase	0.0
30	65	gil34913256 unknown protein	1e-29
31	64	gil1729760 chitinase	1e-54
32	64	gil92876895 Glycine hydroxymethyltransferase	0.0
33	62	gil1545805 cytochrome P450 like	4e-82
34	60	gil2920666 Glutathione S-transferase/peroxidase	1e-96
35	58	gil92877555 photosystem II core complex proteins psby	9e-21
36	54	gil16930765 photosystem II reaction center	5e-11
37	52	gil8071628 aquaporin PIP2-2	7e-34
38	51	gil12231174 aspartic proteinase 2	0.0
39	51	gil18086 cysteine proteinase	7e-29
40	44	gil82080 chlorophyll a/b-binding protein type III	6e-124
41	43	gil18411555 unknown protein	2e-48
42	41	Sem registro	-
43	40	gil25992541 fiber protein Fb11	2e-23
44	40	gil98962139 unknown protein	3e-20
45	38	gil18405145 RCA (rubisco activase)	0.0
46	38	gil18654477 thioredoxin peroxidase	1e-77
47	37	gil18394220 unknown protein	2e-06
48	36	gil47176684 copper chaperone	9e-30
49	36	gil34420138 glycine decarboxylase complex H-protein	1e-74
50	36	gil1181599 subunit of photosystem I	3e-76

Esta análise de hierarquia de expressão para os 50 *hits* sugere que os genes relacionados à fotossíntese, proteção contra o estresse oxidativo e fotorrespiração parecem ter uma maior representatividade relativa no perfil transcricional da biblioteca de Folhas que a sugerida na Figura 17. Assim, os genes foram novamente agrupados nas mesmas categorias funcionais, porém agora, o agrupamento foi realizado levando-se em conta a abundância dos genes, ou seja, o número de transcritos em cada categoria (Figura 18). A nova distribuição das categorias funcionais por transcritos ou abundância dos genes revelou uma diferença de 14% entre a primeira categoria, Metabolismo e Energia, e a segunda, Processos Celulares, sendo que as duas categorias combinadas somam 64% dos transcritos presentes nas células foliares. Resultados semelhantes foram encontrados por BHALERAO *et al.* (2003) através da análise de ESTs de folhas de *Populus tremula* onde 42% dos transcritos estiveram relacionados a Metabolismo e Energia, sendo 16% relacionados às reações de luz da fotossíntese e 15% ao Ciclo de Benson-Calvin. Um outro trabalho realizado com SAGE em folhas de arábido também mostrou que a maior parte dos transcritos esteve relacionada à categoria metabolismo e energia (53% dos transcritos) (JUNG *et al.*, 2003).

Figura 18 - Distribuição dos genes por abundância ou número de transcritos identificados a partir da biblioteca SAGE de folhas em categorias funcionais

Embora a existência de uma correlação entre dados de transcritos e de proteínas seja um tema muito discutido, já que esta correlação tem se mostrado positiva para apenas 30% - 50% das

proteínas identificadas em alguns trabalhos (GION *et al.*, 2005; SOUZA *et al.*, 2006), recentemente, foi realizada a identificação de proteínas de folhas de *Pisum sativum* e 75% das proteínas identificadas estiveram relacionadas à fotossíntese (reações de luz e Ciclo de Benson-Calvin) (SCHILTZ *et al.*, 2004). Entre estas proteínas as mais abundantes foram as subunidades da rubisco, rubisco ativase, subunidades dos fotossistemas I e II, transquetolase, aldolase, gliceraldeído 3P desidrogenase e fosfoglicerato quinase, as quais com exceção da gliceraldeído 3P desidrogenase, foram todas encontradas entre os genes mais expressos na biblioteca de Folhas (Tabela 5). Outras proteínas também abundantes, mas de menor representatividade, identificadas neste trabalho foram proteínas relacionadas ao metabolismo de nitrogênio, a defesa da planta contra o estresse oxidativo e ao metabolismo secundário; entre elas uma glutamina sintetase, peroxidases e superóxido dismutase, quitinases e uma isoflavona redutase. Através do SAGE nós também encontramos um alto nível de expressão dos transcritos correspondentes às proteínas mencionadas acima, sendo que os transcritos para peroxidases, quitinases e isoflavona redutase estiveram entre os 50 genes mais expressos (Tabela 5). Os transcritos para glutamina sintetase e superóxido dismutase também foram abundantes somando, 27 e 36 cópias no tecido, respectivamente.

Possivelmente, essa forte correlação entre o padrão de expressão de transcritos e proteínas apresentada acima resulte da comparação entre os transcritos e proteínas mais abundantes no tecido, ou seja, aqueles cuja expressão caracterizam o tecido funcionalmente. Assim, a redução na correlação entre transcritos e proteínas apontada em outros trabalhos poderia estar relacionada à comparação dos padrões de expressão de transcritos e proteínas de menor abundância. Essa hipótese pode ser sugerida a partir dos próprios resultados de GION *et al.* (2005) que encontraram uma correlação positiva maior entre proteínas e transcritos de maior abundância ($r = 0,46$) quando comparados aos de menor representatividade ($r = 0,31$).

Além disso, a diferença de sensibilidade entre as metodologias utilizadas para acessar transcritos e proteínas pode representar um importante fator na redução da correlação entre os dois perfis. O trabalho realizado por FUKAO *et al.* (2002) é um bom exemplo para ilustrar a importância de se trabalhar com metodologias de sensibilidades diferentes para garantir uma maior amplitude nos resultados. Estes autores utilizaram a eletroforese em gel-2D (pH 3-10) e espectrometria de massa (MALDI-TOF MS) para identificar proteínas de preparados

especificamente constituídos de peroxissomos, e conseguiram identificar 29 spots com essa metodologia.

No entanto, a proteína Glicolato oxidase (pI 9,4 – 9,6), que é sabidamente importante na fotorrespiração e que, portanto, espera-se ser facilmente encontrada nos peroxissomos, não foi identificada. Os autores utilizaram então a metodologia de imuno-deteção por anticorpo e conseguiram identificar tanto a GO1 quanto a GO2 a partir dos preparados de peroxissomos. Se o padrão de expressão de muitas proteínas não pode ser predito a partir do padrão de expressão dos transcritos e vice-versa, todos esses resultados apenas demonstram que as duas formas de análise são igualmente importantes, pois quando combinadas possibilitam comparar o perfil de expressão de um transcrito com o de uma proteína identificada e assim compreender os mecanismos envolvidos na regulação da expressão do gene, e também permitem a identificação de novas proteínas quando os dados de transcritos são utilizados para apontar proteínas possivelmente importantes no material que podem ser identificadas com o emprego de outras metodologias mais específicas.

Os resultados do perfil transcricional da biblioteca de Folhas indicam que 3 processos metabólicos são fundamentais a este tecido: (i) fotossíntese (reações de luz e fixação de carbono), (ii) reciclagem de nitrogênio e fotorrespiração, (iii) reações de proteção ao estresse oxidativo.

Um grande número de transcritos referentes aos genes nucleares que codificam proteínas e/ou enzimas dos complexos proteicos de membrana responsáveis pela conversão da energia solar em energia química (PSI, PSII, complexo do citocromo *b₆f*, e ATP sintetase) foram encontrados na biblioteca de folhas (Tabela 6).

Ambos PSI e PSII são formados por muitas subunidades proteicas, codificadas tanto por genes nucleares quanto por genes cloroplastidiais, que se ligam a clorofilas do complexo coletor de luz LHCB. O PSI é formado pelas subunidades PsaA – PsaL, possivelmente também por subunidades PsaN e PsaO, clorofilas *a* do sistema coletor de luz LHCa1 – LHCa4, e outras proteínas que funcionam na manutenção estrutural do complexo (Revisado por NELSON & YOCUM, 2006). Já a composição do complexo PSII é muito mais complexa e ainda não foi totalmente elucidada em plantas superiores (NIELD & BARBER, 2006), mas atualmente as seguintes proteínas já foram identificadas como componentes do PSII: clorofilas LHCII, LHCb4 – LHCb6, proteínas PsbA – PsbF, feofitina *a*, plastoquinonas, proteínas extrínsecas PsbO, U e V as quais formam uma estrutura que facilita a retenção dos cofatores inorgânicos Ca^{2+} , Mn e Cl^-

utilizados pelo PSII para oxidar a H_2O e produzir elétrons para reduzir a plastoquinona (NIELD & BARBER, 2006).

No total, 1299 transcritos da biblioteca de folhas estão relacionados aos complexos proteicos da fotossíntese, sendo 828 relacionados ao PSII e citocromo *b₆f*, 450 ao PSI, e 21 ao complexo ATP sintetase. Um resultado interessante foi observar que a *tag* cujo transcrito é similar a uma proteína LHCII tipo I do fotossistema PSII apresentou, sozinha, 338 cópias sendo, no mínimo, 3 vezes mais expressa que a segunda *tag* de maior expressão (Tabela 6). Este resultado é condizente com a estrutura tridimensional proposta para o supercomplexo PSII onde o dímero do centro de reação do PSII se liga a dois trómeros de LHCII os quais estão diretamente relacionados à transferência de energia ao centro de reação (NIELD & BARBER, 2006). Da mesma forma, as *tags* referentes a uma clorofila e a um precursor cloroplastidial da ferredoxina I, ambas relacionadas ao fotossistema I, também apresentaram um nível de expressão até 3 vezes maior que as demais *tags* (96 e 129 cópias, respectivamente) (Tabela 6). A ferredoxina é o principal receptor final de elétrons do PSI promovendo a redução de $NADP^+$ através da ferredoxina NADP redutase. Além disso, a ferredoxina é importante na transferência cíclica de elétrons em torno do PSI (que ocorre quando a produção de açúcares através do ciclo de Benson-Calvin é desestimulada), redução do complexo citocromo *b₆f* ou plastocianina, gerando um gradiente de prótons e consequentemente, síntese de ATP (GOLDING & JOHNSON, 2003).

Tabela 6 - *Tags* relacionadas aos genes dos fotossistema I (PSI), fotossistema II (PSII), citocromo *b₆f* e ATP sintetase, encontradas nas bibliotecas SAGE. A sequência de cada *tag* é apresentada, seguida pelo seu número de cópias nas bibliotecas de folhas (F) e de madeira (M), identificação e valor *E*. A diferença estatística significativa entre o número de cópias de uma *tag* nas bibliotecas, quando presente, é representada por (**) para $P \leq 0,01$, e (*) para $P \leq 0,05$ (Teste Z). *Tags* precedidas de um mesmo número sobrescrito indicam isoformas para um mesmo gene

(continua)

Tag	Nº cópias		Identificação	Valor <i>E</i>
	F	M		
Fotossistema II				
1	54**	11	gil16930765 photosystem II reaction center PsbW	5e-11
2	96**	31	gil19911 photosystem II 23kDa polypeptide PsbP	2e-105
3	11**	0	gil400198 Photosystem II protein PsbT (11 kDa protein)	1e-07
4	58**	7	gil92877555 photosystem II core complex protein PsbY	1e-20
5	8*	2	gil108707991 Photosystem II reaction center Psb27	4e-44
6	338**	56	gil974850 LHCII type I protein	3e-124
7	7**	0	gil15235029 LHCB5 light harvesting complex of PSII	4e-126
8	8**	1	gil15231990 LHCB4.2 PSII	3e-126
9	6	4	gil115792 Chlorophyll a-b binding protein PSII (type I CAB-21)	3e-140
10	124**	21	gil115792 Chlorophyll a-b binding protein PSII (type I CAB-21)	5e-139
11	11**	1	gil15235029 LHCB5 light harvesting complex of PSII	7e-127

Tabela 6 - *Tags* relacionadas aos genes dos fotossistema I (PSI), fotossistema II (PSII), citocromo *b₆f* e ATP sintetase, encontradas nas bibliotecas SAGE. A sequência de cada *tag* é apresentada, seguida pelo seu número de cópias nas bibliotecas de folhas (F) e de madeira (M), identificação e valor *E*. A diferença estatística significativa entre o número de cópias de uma *tag* nas bibliotecas, quando presente, é representada por (**) para $P \leq 0,01$, e (*) para $P \leq 0,05$ (Teste Z). *Tags* precedidas de um mesmo número sobrescrito indicam isoformas para um mesmo gene

(conclusão)				
Tag	Nº cópias		Identificação	Valor E
	F	M		
Fotossistema II				
² 12	4	1	gil1709651 Plastocyanin A, chloroplast precursor	1e-41
12	18**	2	gil30683145 plastoquinol-plastocyanin reductase	7e-16
13	33**	6	gil68566175 Cytochrome b6-f complex petC	2e-88
14	19**	4	<u>gil115792 </u> Chlorophyll a-b binding protein 21 LHCII	5e-144
² 15	35**	8	gil929813 Plastocyanin a	2e-51
Fotossistema I				
¹ 1	4*	0	gil92876000 Photosystem I reaction center subunit PsaN	6e-55
¹ 2	5	1	gil92876000 Photosystem I reaction center subunit PsaN	9e-55
3	3	3	gblAAB31705.1 photosystem I subunit PsAE	2e-45
4	5	1	gil24030202 putative photosystem I subunit PSI PsAE	5e-24
5	5	1	gil30013661 photosystem I reaction center subunit PsAK	3e-41
6	36**	14	gil1181599 subunit of photosystem I PsAL	3e-76
7	18**	5	gil108802653 photosystem I subunit VIII PsAL	2e-12
8	11**	1	<u>gil6470348 </u> PSI light-harvesting antenna chlorophyll LHCB	2e-111
9	30**	5	<u>gil6470348 </u> PSI light-harvesting antenna chloro LHCB	3e-124
10	7	2	<u>gil2306981 </u> photosystem I antenna protein LHCB	2e-97
11	7*	0	<u>gil21554335 </u> PSI type III chlorophyll a/b-binding LHCB	2e-35
12	27**	6	<u>gil57335320 </u> photosystem I subunit O precursor	6e-13
13	44**	7	<u>gil110736814 </u> PSI type III chlorophyll a/b-binding protein	8e-119
14	96**	20	<u>gil493723 </u> photosystem I light-harvesting chlorophyll a/b binding	3e-112
² 15	129**	25	<u>gil3023743 </u> Ferredoxin-1, chloroplast precursor	1e-43
² 16	11*	1	gil3023743 Ferredoxin-1, chloroplast precursor	3e-44
17	5	3	<u>gil6899972 </u> chloroplast ferredoxin-NADP+ oxidoreductase precursor	3e-170
ATP sintetase				
1	21**	2	gil19787 chloroplast ATP synthase (delta subunit) AtpD	8e-71

A oxidação fotoquímica da água até a geração de O₂ que ocorre durante a transferência de elétrons através dos fotossistemas, culmina com a redução de NADP⁺ a NADPH e a geração de um gradiente de prótons através da membrana dos tilacóides que atua como força propulsora para a síntese de ATP. Ambos produtos das reações de luz da fotossíntese, NADPH e ATP são consumidos no estroma dos cloroplastos durante as reações de redução do CO₂ assimilado a moléculas de carboidrato através do Ciclo de Benson-Calvin. Foram encontradas *tags* para todas as enzimas do Ciclo de Benson-Calvin sendo que apenas para duas enzimas envolvidas na regeneração de ribulose -1,5 -bisfosfato, sedoheptulose 1,7-bisfosfato e fosforibuloquinase, não foi possível identificar transcritos (Figura 19). Quando foi realizado o (t)blastn sobre os bancos de dados de referência com a sequência das proteínas sedoheptulose 1,7-bisfosfato e

fosforibuloquinase (previamente depositadas no NCBI), nenhum dos ESTs similares nos bancos de dados possuía poli(A) 3', ou seja, provavelmente as *tags* para estas enzimas estejam representadas na biblioteca de folhas, mas sua expressão não pode ser caracterizada devido a impossibilidade de identificação.

Figura 19 - Ciclo de Benson-Calvin. Assimilação de CO₂ e regeneração de ribulose-1,5-bisfosfato. As enzimas para as quais foram encontradas *tags* estão representadas na figura com setas pretas e todas as *tags* encontradas para uma enzima são apresentadas. *Tags* precedidas de um mesmo número sobrescrito representam isoformas de um mesmo gene. Para cada *tag* é fornecida uma ilustração colorida comparativa do nível de expressão entre as bibliotecas de folhas (F) e de madeira (M). A cor branca indica ausência e a cor cinza indica que não há diferença significativa entre as bibliotecas. As cores azul e vermelho foram utilizadas para indicar a ocorrência de diferença significativa, $P \leq 0,05$ (*) ou $P \leq 0,01$ (**), entre as bibliotecas, sendo que o vermelho indica maior expressão relativa e o azul indica menor expressão relativa. Setas cinza foram utilizadas para representar os passos enzimáticos para os quais nenhuma *tag* foi encontrada. As caixas sombreadas representam passos intermediários na regeneração de ribulose-1,5-bisfosfato

No total, 508 transcritos correspondendo a enzimas do ciclo de Benson-Calvin e regeneração de ribulose -1,5 -bisfosfato foram observados na biblioteca de folhas, sendo as

enzimas de maior expressão a subunidade menor da rubisco com 273 cópias, a gliceraldeído 3P desidrogenase com 67 cópias e a aldolase com 89 cópias.

Além de sua importância na assimilação de CO₂, a rubisco representa a maior fonte de reserva de N nas folhas (IRVING & ROBINSON, 2006) o que condiz com o elevado número de transcritos encontrados para esta proteína na biblioteca de folhas. Os transcritos encontrados representam um EST com alta similaridade ao gene que codifica a subunidade menor da rubisco (*rbcS*). A estrutura da holoenzima da Rubisco é composta pela associação de 8 monômeros da subunidade maior e 8 monômeros da subunidade menor (AKAZAWA *et al.*, 1894) e, no entanto, a expressão do gene que codifica para a subunidade maior (*rbcL*) não foi observada. Isso ocorre porque a subunidade maior da rubisco é codificada por um gene do genoma do cloroplasto, diferentemente do que ocorre para a subunidade menor, que é codificada por um gene nuclear. Nos cloroplastos, a presença de um terminal poli(A) 3' representa um sinal para degradação de moléculas de RNAm (Revisado por HAYES *et al.*, 1999) o que conseqüentemente dificultaria ou impossibilitaria o isolamento de transcritos a partir do terminal poli(A) 3' para geração das SAGE tags. Assim como o encontrado para o gene *rbcL*, para 33 genes de um total de 34 genes do cloroplasto de *Eucalyptus globulus* relacionados às reações de luz da fotossíntese, não foram encontradas tags (os genes cloroplastidiais de *Eucalyptus globulus* encontram-se listados em STEANE, 2005). Um fato interessante foi encontrar uma tag para o gene cloroplastidial *PsaI* do fotossistema I de *Eucalyptus globulus*, o que poderia sugerir que uma cópia deste gene está também presente no genoma nuclear, ou que o transcrito cloroplastidial de *PsaI* possui um terminal poli(A) 3' (Tabela 6).

A reação de carboxilação da ribulose-1,5-bisfosfato, catalisada pela Rubisco, ocorre a uma taxa extremamente baixa sendo necessária sua ativação para que a enzima se torne cataliticamente eficiente. Essa ativação é realizada pela rubisco ativase (RCA) através da remoção de inibidores ligados aos sítios catalíticos da enzima induzindo modificação conformacional destes sítios e conseqüente ativação (PORTIS, 2003). A RCA é uma importante proteína dos cloroplastos, mas é codificada pelo genoma nuclear (PORTIS, 1995). Duas tags referentes à RCA foram encontradas na biblioteca de folhas, uma com 38 cópias e a outra com 15 cópias. O alinhamento entre os dois ESTs foi possível para apenas uma região de 213 pb com somente 80% de identidade, o que sugere que os transcritos sejam originados a partir de genes diferentes ao invés de representarem isoformas de um mesmo gene. Entretanto, a possibilidade

das *tags* representarem isoformas de um único gene é fortemente sugerida pelo fato de um dos ESTs possuir as duas *tags* encontradas para RCA indicando, portanto, que uma das *tags* foi perdida nos transcritos devido ao *splicing* 3'. A segunda hipótese é também favorecida por dados anteriores que indicam a ocorrência de um gene cópia única para RCA, cujos transcritos sofrem *splicing* dando origem as duas isoformas diferentes traduzidas nos polipeptídeos de 41 KDa e 45 KDa os quais constituem a RCA em *arabidopsis* e espinafre (WERNEKE *et al.*, 1989).

A atividade da rubisco é também controlada pela luz através da tiorredoxina *f* que quando ativada pela ferredoxina reduzida (produto das reações de luz da fotossíntese) ativa a RCA (Revisado em GRACIET *et al.*, 2004). Foram observadas *tags* referentes as isoformas *m*, *f* e *h* da tiorredoxina na biblioteca de folhas sendo que para a tiorredoxina *f* uma única *tag* com 17 cópias foi encontrada. Além de controlar a atividade das enzimas envolvidas na assimilação de CO₂ (Revisado em GRACIET *et al.*, 2004), outros produtos da fotossíntese são também importantes controladores da expressão dos genes cloroplastidiais e nucleares, como por exemplo, a plastoquinona que quando oxidada durante o transporte de elétrons a partir do PSII estimula a expressão dos genes nucleares *Lhcb* (complexo coletor de luz). Os genomas dos cloroplastos e núcleo codificam diferentes subunidades dos complexos proteicos relacionados à fotossíntese (PSI, PSII, citocromo *b6f*, ATP sintetase, Rubisco e NADH desidrogenase), as quais devem ser sintetizadas e reunidas em complexos funcionais coordenadamente nos cloroplastos. Assim, a troca de informações entre cloroplastos-núcleo é fundamental para permitir respostas rápidas de aclimação da fotossíntese em diferentes situações de luz e/ou temperatura, bem como para possibilitar um controle preciso da expressão dos genes dos cloroplastos e do núcleo, uma vez que, enquanto cada célula contém apenas um genoma nuclear, em cada célula fotossintetizante estão presentes mais de 100 cloroplastos, cada qual com 100 ou mais cópias do genoma organelar (Revisado em FEY *et al.*, 2005).

As reações de redução que ocorrem durante a fotossíntese são também importantes na regulação do fluxo de carbono para a síntese de açúcares, amido, lipídeos e também para o metabolismo de nitrogênio (Revisado em GEIGENBERGER *et al.*, 2005). Por exemplo, as tiorredoxinas *f* e *m* quando reduzidas ativam a ADP-glicose pirofosforilase (AGPase) que é uma importante enzima para a síntese de amido nos cloroplastos. Da mesma forma, a enzima acetil-CoA carboxilase (ACCase), que catalisa o primeiro passo na síntese de lipídeos nos plastídeos, é também ativada através de modulação-redox. Outro importante exemplo é a ativação por

modulação-redox de enzimas envolvidas na assimilação de N e síntese de aminoácidos durante a fotorrespiração tais como a glutamina sintetase (GS) e glutamato sintase (GOGAT) (Revisado em GEIGENBERGER *et al.*, 2005).

A expressão de transcritos para a AGPase não pôde ser verificada nas bibliotecas provavelmente devido a ausência de sequências nos bancos de referência para a enzima, já que o (t)blastn de sequências de AGPase depositadas no NCBI não identifica nenhum EST de *Eucalyptus* com terminal poli(A) 3' nos bancos de referência. No entanto, uma outra enzima envolvida na síntese de amido, a amido sintetase, pode ser identificada na biblioteca de folhas por uma *tag* de somente 6 cópias; e duas enzimas envolvidas no direcionamento do fluxo de carbono para a glicólise, a invertase e sacarose sintetase, foram também identificadas por *tags* com 3 e 10 cópias no total, respectivamente (Figura 19 e 20).

O direcionamento de açúcares para a via glicolítica, ciclo dos ácidos tricarboxílicos (TCA) e cadeia respiratória, representa um processo de pouca expressão nas folhas somando no total, 268 transcritos, 2,5 vezes menos que o número de transcritos encontrados para o mesmo processo na biblioteca de madeira. A função primordial da glicólise nos cloroplastos ocorre durante a noite na quebra de amido e produção de intermediários para biossíntese de biomoléculas, geração de poder redutor para muitas reações enzimáticas, entre elas a assimilação de N pelo sistema GS/GOGAT (glutamina sintetase/glutamato sintase), e produção de ATP para construção de biomoléculas, por exemplo, os ácidos graxos (Revisado em PLAXTON, 1996; PLAXTON & PODESTÁ, 2006).

Figura 20. Glicólise e Ciclo dos Ácidos Tricarboxílicos. As enzimas para as quais foram encontradas *tags* estão representadas na figura com setas pretas e todas as *tags* encontradas para uma enzima são apresentadas. *Tags* precedidas de um mesmo número sobrescrito representam isoformas de um mesmo gene. Para cada *tag* é fornecida uma ilustração colorida comparativa do nível de expressão entre as bibliotecas de folhas (F) e de madeira (M). A cor branca indica ausência e a cor cinza indica que não há diferença significativa entre as bibliotecas. As cores azul e vermelho foram utilizadas para indicar a ocorrência de diferença significativa, $P \leq 0,05$ (*) ou $P \leq 0,01$ (**), entre as bibliotecas, sendo que o vermelho indica maior expressão relativa e o azul indica menor expressão relativa. Setas cinza foram utilizadas para representar os passos enzimáticos para os quais nenhuma *tag* foi encontrada

Grande parte dos açúcares sintetizados nas folhas deve ser preferencialmente transportada para os tecidos não fotossintetizantes como, por exemplo, o xilema, onde servirão como fonte de energia para a síntese de biomoléculas (KONISHI *et al.*, 2004). Duas enzimas catalisam a síntese de sacarose nas folhas, a sacarose sintetase (SUSY) e a sacarose fosfato sintetase (SPS), ambas apresentaram baixo nível de expressão nas folhas de *Populus* (GEISLER-LEE *et al.*, 2006) semelhante ao encontrado no presente trabalho.

Um resultado interessante foi a identificação de uma *tag* para galactinol sintase entre os 50 *hits* da biblioteca de folhas (Tabela 5). No total foram encontradas três *tags* para essa enzima com altíssimo nível de expressão somando 230 transcritos. A galactinol sintase é uma enzima

chave na via que leva à formação de rafinose e outros oligossacarídeos (RFO). Esta enzima catalisa a síntese de galactinol, a partir de UDP-galactose e mio-inositol, que serve como doador de grupos galactosil para a síntese de rafinose. A síntese de rafinose por sua vez é realizada a partir de sacarose e galactinol pela rafinose sintase, a qual esteve representada por uma *tag* com 8 cópias. Elevados níveis de galactinol sintase e RFOs estão relacionados à tolerância à desidratação (DOWNIE *et al.*, 2003), a reserva de carbono na forma de carboidratos em folhas e raízes, e ao transporte de açúcares no floema em algumas espécies (SPRENGER & KELLER, 2000). Nas folhas de *Lamium album* os oligossacarídeos da família ROF representam produtos da fotossíntese tão importantes quanto sacarose e amido (KOCHHAR *et al.*, 2003). Possivelmente em *E. grandis* estes oligossacarídeos sejam também importantes como produtos da fotossíntese, tanto na translocação de açúcares pelo floema, quanto como composto de reserva, já que a expressão de transcritos para enzimas envolvidas na síntese de amido foi realmente baixa na biblioteca de folhas.

Além da fotossíntese, as folhas são também caracterizadas por um outro processo que envolve a atividade oxigenase da rubisco, a fotorrespiração, e para o qual um grande número de transcritos foi observado na biblioteca de folhas (185 transcritos). A fotorrespiração inicia-se com a incorporação de O₂ à ribulose-1,5-bisfosfato pela rubisco dando origem a 2-fosfoglicolato e 3-fosfoglicerato no estroma do cloroplasto (Figura 21). A hidrolização do 2-fosfoglicolato leva a formação de glicolato que é então transportado para os peroxissomos onde é oxidado à glioxilato e H₂O₂ pela glicolato oxidase (Revisado em KEYS, 2006). Uma única *tag* com 103 cópias foi encontrada para a glicolato oxidase na biblioteca de folhas sugerindo que possivelmente a fotorrespiração seja um processo ativo nas folhas de *Eucalyptus grandis*. O glioxilato produzido nos peroxissomos é transaminado a glicina utilizando o grupamento amino doado pelo glutamato. A glicina por sua vez é transportada para o interior das mitocôndrias e convertida a serina pela glicina descarboxilase, reação que leva a liberação de amônia e CO₂ (Revisado em KEYS, 2006). Duas *tags* foram encontradas para a glicina descarboxilase, uma com 3 e a outra com 36 cópias. Quando o alinhamento dos ESTs referentes à estas *tags* foi realizado, nenhuma identidade foi encontrada, sugerindo, portanto, a ocorrência de pelo menos dois genes para a glicina descarboxilase em *E. grandis*. A enzima glicina descarboxilase é, entretanto, composta por 4 componentes proteicos (denominados P, T, H, e L) codificados por pequenas famílias multigênicas nucleares, e direcionados para os cloroplastos (Revisado em BAUWE &

KOLUKISAOGLU, 2003) sendo provável, portanto, que outros genes para a enzima estejam presentes em *E. grandis*.

Figura 21 - Oxigenação de ribulose-1,5-bisfosfato e Fotorrespiração. As reações estão divididas de acordo com sua localização celular em reações dos cloroplastos, peroxissomos e mitocôndrias. As enzimas para as quais foram encontradas tags estão representadas na figura com setas pretas e todas as tags encontradas para uma enzima são apresentadas. Tags precedidas de um mesmo número sobrescrito representam isoformas de um mesmo gene. Para cada tag é fornecida uma ilustração colorida comparativa do nível de expressão entre as bibliotecas de folhas (F) e de madeira (M). A cor branca indica ausência e a cor cinza indica que não há diferença significativa entre as bibliotecas. As cores azul e vermelho foram utilizadas para indicar a ocorrência de diferença significativa, $P \leq 0,05$ (*) ou $P \leq 0,01$ (**), entre as bibliotecas, sendo que o vermelho indica maior expressão relativa e o azul indica menor expressão relativa. Setas cinza foram utilizadas para representar os passos enzimáticos para os quais nenhuma tag foi encontrada.

A amônia liberada na descarboxilação da glicina é utilizada pela glutamina sintetase (GS)

e glutamato sintase (GOGAT) nos cloroplastos para a síntese de aminoácidos e consequente reincorporação de N. A glutamina sintetase é codificada por genes nucleares sendo que um gene com duas isoformas codificam a GS cloroplastidial ou GS2 (IRELAND & LEA, 1999). Através do SAGE foi possível também identificar e caracterizar a expressão de duas tags para a GS2, com 3 e 24 cópias (Figura 21), as quais foram identificadas como duas isoformas de um único gene a

partir do alinhamento dos ESTs que evidenciou a presença das duas *tags* em um dos ESTs sugerindo, portanto, a geração de isoformas através de *splicing* 3'. A expressão da GOGAT não foi observada devido à ausência de sequências de eucalipto nos bancos de referência para esta enzima. Assim, embora não tenha sido possível caracterizar a expressão da GOGAT, enzima importante na formação de glutamato a partir de glutamina, os níveis de expressão para as outras enzimas envolvidas no processo sugerem fortemente que a fotorrespiração e reciclagem de N pelo sistema GS/GOGAT são processos que ocorrem ativamente nas folhas de *Eucalyptus grandis*.

Tanto a fotossíntese, quanto a fotorrespiração, resulta na geração de uma grande quantidade de espécies reativas de oxigênio (ROS) as quais podem causar inativação temporária da fotossíntese, também conhecida como fotoinibição, danos fotooxidativos, e oxidação de proteínas, ácidos graxos insaturados e DNA, podendo levar a célula à morte. A produção de ROS durante a fotorrespiração ocorre como resultado da reação catalisada pela glicolato oxidase que libera grandes quantidades de H_2O_2 nos peroxissomos (Figura 21). Embora os peroxissomos e também as mitocôndrias produzam espécies reativas de oxigênio, os cloroplastos são a principal fonte de ROS nas plantas (ASADA, 2006). Nos cloroplastos ocorre a formação de uma grande quantidade de O_2 que pode ser reduzido durante o transporte de elétrons podendo levar a formação de superóxido O_2^- , radicais de hidroxila e peróxido de hidrogênio (MEHLER, 1951). Para evitar danos causados pelo estresse oxidativo, as plantas desenvolveram mecanismos de defesa que incluem a produção de antioxidantes enzimáticos, tais como superóxido dismutase (SOD), peroxidases, catalases e glutathione S-transferases/peroxidases, e não enzimáticos, como carotenóides, antocianinas e flavonóides (NIYOGI, 1999). Na Tabela 7 são apresentadas as *tags* e nível de expressão para todas as enzimas relacionadas ao estresse oxidativo que foram encontradas na biblioteca de folhas. Os resultados sugerem uma maior importância da glutathione S-transferase/peroxidase e catalases na eliminação das espécies reativas de oxigênio nas folhas de *Eucalyptus grandis*.

Tabela 7 - Lista das *Tags* que representam genes relacionados à proteção celular contra o estresse oxidativo encontradas nas bibliotecas SAGE. A sequência de cada *tag* é apresentada, seguida pelo seu número de cópias nas bibliotecas de folhas e de madeira, identificação e valor de *E*. A diferença estatística significativa entre o número de cópias de uma *tag* nas bibliotecas, quando presente, é representada por (**) para $P \leq 0,01$, e (*) para $P \leq 0,05$ (Teste Z)

<i>Tag</i>	Nº cópias		Identificação	Valor <i>E</i>
	F	M		
1	11	5	gil32526566l catalase	0.0
2	23	15	gil6006609l catalase	6e-160
3	26**	0	gil32526566l catalase	0.0
4	11	50**	gil38228697l superoxide dismutase	1e-71
5	3	0	gblAAK01931.1l superoxide dismutase	4e-60
6	9	6	gil38228697l superoxide dismutase	3e-72
7	3	0	gil33413303l superoxide dismutase	4e-78
8	5	25*	gil33340236l superoxide dismutase	3e-77
9	5	3	gil13274148l superoxide dismutase	4e-15
10	6	14	gil11385459l glutathione S-transferase	6e-89
11	60**	15	gil2920666l glutathione S-transferase/peroxidase	1e-96
12	5	7	gil2920666l glutathione	3e-97
13	13	30*	gil21217741l glutathione S-transferase	9e-42
14	3	0	gblABA42224.1l glutathione S-transferase	2e-87
15	8*	1	gil6652872l glutathione S-transferase 2	1e-79
16	10	7	gil33308408l glutathione peroxidase	5e-80
17	4	1	gil73647738l ascorbate peroxidase	5e-96
18	6	3	gil2754860l ascorbate peroxidase	3e-119

5.3 Perfil transcricional da biblioteca de madeira

A partir da distribuição dos genes encontrados na biblioteca de madeira em categorias funcionais foi possível observar que 20% dos genes pertencem à categoria Metabolismo e Energia, 17% à categoria Estrutura e Organização da Estrutura Celular, 15% à categoria Vias de Informação, 14% à categoria Processos Celulares, 7% à categoria Transporte, e 27% dos genes não puderam ter sua função definida (Figura 22). Foram mais representativos na categoria Metabolismo e Energia, genes relacionados ao metabolismo de aminoácidos, carboidratos, lipídeos e à respiração. A categoria Estrutura e Organização da Estrutura Celular teve a maior parte dos genes atribuída a formação da parede celular, ao citoesqueleto e a proteínas e RNA ribossomais. Coincidente com o grande número de genes para RNA e proteínas ribossomais, a maior parte dos genes na categoria Vias de Informação esteve também relacionada à síntese/modificação/degradação de proteínas. Na categoria Processos Celulares, um grande número de genes esteve associado à transdução de sinais, respostas de proteção e destoxificação e respostas a estímulos, enquanto a categoria Transporte apresentou a maior parte dos genes relacionados ao transporte interno de vesículas.

Figura 22 - Distribuição dos genes identificados a partir da biblioteca SAGE de madeira em categorias funcionais. A categoria “Sem Função Definida” foi representada por *tags* para as quais foram encontrados os prováveis ESTs de origem mas, não foi possível identificar a função destas sequências através do NCBI

A *tag* mais expressa na biblioteca de madeira identificou um EST muito similar a uma isoflavona redutase (1e-145) e a redutase éter benzílico fenilcoumaran ou PCBER (2e-135) (Tabela 8). Ambas enzimas são redutases dependentes de NADPH, responsáveis pela biossíntese de lignanas e isoflavonóides (SHOJI *et al.*, 2002). A expressão dessas NADPH-dependente redutases tem sido positivamente correlacionada à expressão de enzimas chave na síntese de lignina tais como a PAL, COMT e CCoAOMT (VINCENT *et al.*, 2005), e com o crescimento secundário em árvores transgênicas de *Aspen* (ISRAELSSON *et al.*, 2003). Recentemente, a expressão de PCBER foi localizada na zona de células do caule durante crescimento secundário ativo, e foi correlacionada a produção de compostos secundários adicionados à parede celular em lignificação, levando a formação do cerne e promovendo uma fonte adicional de defesa a patógenos em tabaco, *Pinus taeda* e *Forsythia intermedia* (SHOJI *et al.*, 2002). Assim, muito provavelmente a alta expressão de transcritos para NADPH-dependente redutases no xilema de *Eucalyptus grandis* esteja também relacionada ao crescimento secundário e deposição de compostos secundários nas células maduras do xilema.

Tabela 8 - Lista dos 50 genes mais expressos na biblioteca de Madeira. Para cada *tag* é fornecido o número de cópias na biblioteca de Madeira, a identificação do gene bem como o valor *E* de cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou o cDNA de origem

Tag	Nº de cópias	Blast(x) – NCBI	E value
1	472	gil10764491 allergenic isoflavone reductase-like protein	e-145
2	273	gil30841338 arabinogalactan protein	5e-53
3	231	gil15236267 Aquaporin PIP3	2e-148
4	195	gil56311394 ADP-ribosylation factor 1	9e-100
5	193	gil21536535 Unknown protein	9e-17
6	163	Sem registro	-
7	130	Sem registro	-
8	122	gil22448531 tubulin alpha-6 chain	2e-58
9	122	gil234523151 beta-tubulin 1	0.0
10	116	gil845791451 hypothetical protein	3e-09
11	115	gil928982461 Heavy metal transport/detoxification	1e-47
12	109	gil119901301 17.9 kDa heat-shock protein	6e-23
13	108	gil1206761 Glyceraldehyde-3-phosphate dehydrogenase	3e-21
14	104	gil509204391 membrane protein	1e-50
15	101	gil139244901 cytosolic glutamine synthetase	9e-20
16	97	Sem registro	-
17	97	gil215536141 blue copper protein	8e-35
18	97	gil15667001 metallothionein-like type 3	2e-24
19	96	gil928808841 Adenosine kinase	2e-113
20	93	Sem registro	-
21	92	gil152369891 translation initiation factor SUI1	4e-55
22	92	gil857012141 Profilin	2e-65
23	91	gil496429801 Unknown protein	3e-20
24	88	gil34829331 cdc2 protein kinase	3e-74
25	88	Sem registro	-
26	81	gil349132561 unknown protein	1e-29
27	80	gil380165211 fiber protein Fb34	9e-71
28	79	gil845791451 hypothetical protein	5e-13
29	78	gil293676051 adenosylhomocysteinase-like protein	1e-154
30	78	gil508446771 cytochrome b5 isoform Cb5-C	8e-62
31	75	Sem registro	-
32	73	gil26778281 cysteine protease	e-165
33	70	gil19348591 caffeoyl-CoA O-methyltransferase	3e-138
34	69	Sem registro	-
35	67	gil81345691 Methionine synthase	0.0
36	65	Sem registro	-
37	63	gil670039071 cellulose synthase EgCesA1	2e-144
38	62	gil771576371 ribulose-1,5-bisphosphate carboxylase small subunit	2e-82
39	61	gil324706351 Mob1/phocein family protein	1e-118
40	59	gil710406871 auxin-repressed protein	1e-31
41	56	gil34174511 light-harvesting chlorophyll a/b-binding protein	9e-80
42	56	gil74066631 ripening-related protein	5e-53
43	54	gil215535231 arabinogalactan protein	2e-42
44	54	Sem registro	-
45	53	gil4752161 translation initiation factor	1e-103
46	52	gil928815341 Ubiquitin	9e-75
47	52	gil197021 eukaryotic initiation factor 5A	8e-83
48	50	gil382286971 Superoxide dismutase	2e-71
49	50	gil482099681 mitochondrial ATP synthase	3e-57
50	48	gil781914421 S-adenosyl methionine synthase	6e-84

Foi observado um grande número de transcritos relacionados à destoxificação celular na biblioteca de madeira, sendo que um destes transcritos, representando uma superóxido dismutase, esteve entre os 50 mais expressos na biblioteca (Tabela 8). Na região cambial, há uma grande concentração de espécies reativas de oxigênio (H_2O_2) que são requeridas para a polimerização oxidativa dos monolignóis durante o processo de lignificação intenso. Evidências a partir de cultura de células de *Zinnia* induzidas a se transdiferenciarem em traqueídeos, indicam que os radicais H_2O_2 utilizados pelas peroxidases na polimerização de lignina são produzidos por células do parênquima, as quais não sofrem lignificação (Ros BARCELO, 2005, Ros GOMEZ *et al.*, 2006). Outra importante fonte de radicais de oxigênio na região cambial é o processo terminal de diferenciação dos elementos de vaso resultante da morte celular e da autólise (MOREAU *et al.*, 2005).

Foram também encontrados entre os 50 *hits*, genes relacionados à síntese, organização e modificação da parede celular, bem como ao citoesqueleto celular, genes associados à respiração, transporte de água e de metais pesados, degradação de proteínas, divisão celular, e genes relacionados ao metabolismo de metionina e de nitrogênio. A representatividade dos genes relacionados aos processos citados acima pôde ser melhor observada quando realizamos a distribuição dos transcritos, e não dos genes identificados, em categorias funcionais como mostrado na Figura 23. Como esperado, nesta nova divisão as categorias Metabolismo e Energia, e Estrutura e Organização da Estrutura Celular, passam a representar as categorias de maior representatividade na biblioteca de madeira somando 57% dos transcritos acessados.

Figura 23 - Distribuição dos genes por abundância ou número de transcritos identificados a partir da biblioteca SAGE de madeira em categorias funcionais

Durante a xilogênese, as células cambiais que darão origem ao xilema passam por uma série de etapas ao longo do processo de diferenciação que começa com a divisão celular seguida pela expansão das células, deposição das paredes celulares primárias e secundárias, lignificação e morte celular programada (CHAFFEY, 1999). A subcategoria divisão celular apresentou a maior relação transcritos/genes da categoria Processos celulares (Tabela 9), sendo a maior parte dos transcritos similares às ciclinas CDC2 e SKP1 (proteína do cinetócoro requerida para progressão do ciclo celular), e ao complexo promotor da anáfase APC10. Um elevado número de transcritos para genes relacionados ao ciclo celular foi também encontrado em células cambiais em divisão, tanto em *Populus* (HERTZBERG *et al.*, 2001) quanto em *Arabidopsis* (OH *et al.*, 2003), indicando que a amostra utilizada na construção da biblioteca de madeira incluiu também a zona de divisão celular e que, portanto, todas as zonas de diferenciação do xilema estão potencialmente representadas na biblioteca.

Após a divisão celular as células passam por um processo de expansão caracterizado pela coordenação do crescimento passivo mediado pelo turgor celular, com o afrouxamento da parede celular pré-existente e a deposição de novos polímeros na parede celular (CARPITA & GIBEAUT, 1993). Assim, embora apenas poucos trabalhos realizados sobre a formação da madeira tenham reportado a identificação de transcritos superexpressos para canais

transportadores de água (YANG *et al.*, 2003, RANIK *et al.*, 2006), certamente o grande número de transcritos encontrados para estas proteínas, principalmente para PIP2 e PIP3 (proteínas intrínsecas de membrana ou aquaporinas), na biblioteca de madeira, esteja relacionado à promoção da expansão celular mediada pela pressão de turgor.

A expansão celular, no entanto, só é possível com a modificação da estrutura da parede celular que é catalisada por enzimas como as xiloglucana endotransglicosilases (XET), expansinas, pectina metil esterases (PME), pectato liase e endopoligalacturonases (Revisado em MELLEROWICZ *et al.*, 2001). As XET foram representadas por 3 *tags* diferentes sendo uma similar a *XTH2* de *Eucalyptus globulus* com 8 cópias na biblioteca, uma *tag* similar a *XTH3* de arabidopsis com 3 cópias, e uma última *tag* similar a *XTH3* também de *Eucalyptus globulus* com 19 cópias na biblioteca. Quando as duas sequências de ESTs similares à *XTH3* de arabidopsis e eucalipto são alinhadas, nenhuma região de identidade é observada, sugerindo que possivelmente, *XTH3* de arabidopsis tenha similaridade com outro gene XTH de *E. globulus* que não *XTH3*. Somente duas *tags* para expansinas foram encontradas, uma similar a alfa-expansina de *Eucalyptus globulus* (*EXPA6*) com 2 cópias, e a outra similar a alfa-expansina 1 (*EXPA1*) de *Populus tremula x Populus tremuloides*, com 9 cópias na biblioteca.

Das enzimas relacionadas à modificação da parede celular, àquelas associadas à modificação dos componentes ricos em pectina da parede foram as mais abundantes. Somente para pectina metil esterases foram encontradas duas *tags* com alta expressão sendo uma similar a *PME2* de arabidopsis com 56 cópias e a outra similar a *PME1* também de arabidopsis com 35 cópias. Foi interessante notar que embora os ESTs para as duas PME encontradas sejam similares a genes distintos em arabidopsis, o alinhamento desses ESTs de eucalipto sugerem que as duas sequências representam isoformas de um único gene produzidas através de *splicing* 3'. As poligalacturonases, enzimas que despolimerizam pectinas desmetiladas pela ação das PME, foram representadas por uma única *tag* similar a *PG4* de *Eucalyptus globulus* também com alta expressão (31 cópias). Foram também encontradas duas *tags* referentes à isoformas do gene pectato liase 1 (*PL1*) de *Eucalyptus globulus* sendo uma com 4 e a outra com 3 cópias; e uma *tag* referente a um EST similar ao gene de pectato liase de arabidopsis com 18 cópias. A grande representatividade das PME e PG na biblioteca de madeira indica que a maior parte das células amostradas da região cambial está em crescimento secundário ativo e processo de aumento da rigidez da parede celular, uma vez que as PME catalisam a remoção dos grupamentos metil das

poligalacturonanas, aumentando sua capacidade de realizar ligações cruzadas com cátions como o cálcio, por exemplo, e consequentemente, tornando os componentes ricos em pectina viscosos (Revisado em DARLEY *et al.*, 2001). Por outro lado, as enzimas que atuam sobre as ligações entre xiloglucanas e xiloglucanas-celulose (expansinas), e que são responsáveis pela extensibilidade da parede celular, principalmente durante o crescimento primário, (Revisado em DARLEY *et al.*, 2001) encontram-se em menor abundância nessa região.

Embora a atividade das XETs na expansão celular esteja associada ao crescimento primário através da clivagem e transferência de cadeias de xiloglucanas promovendo um afrouxamento da parede primária, recentemente a atividade de XETs tem sido também correlacionada com o crescimento secundário em *Populus* e *Eucalyptus* (BOURQUIN *et al.*, 2002, PAUX *et al.*, 2004). Durante o crescimento secundário estas enzimas podem estar participando da reestruturação da parede primária enquanto a parede secundária é depositada através da conexão entre as cadeias de xiloglucanas pré-existentes com as novas cadeias que estão sendo depositadas (BOURQUIN *et al.*, 2002). Assim, o alto nível de expressão das *tags* para XETs e PMEs na região cambial das árvores jovens de *Eucalyptus grandis* são resultados coincidentes que apontam para o crescimento secundário ativo nessa região.

A expansão celular ocorre durante a síntese da parede primária terminando com o início da deposição da parede secundária e lignificação. A síntese dos polissacarídeos que compõem a parede celular é realizada a partir do precursor comum UDP-D-glicose que pode ser dirigido para a síntese de hemiceluloses e pectinas e para a síntese de celulose. Uma série de reações de desidrogenação, descarboxilação e epimerização que utilizam nucleosídeos 5'-açúcar-difosfato (NDPs) são responsáveis pela síntese de glicanas e pectinas a partir de UDP-D-glicose (REITER & VANZIN, 2001). Na Figura 24 é apresentado um resumo dessas reações com as *tags* encontradas e seus níveis de expressão nas bibliotecas.

Figura 24 - Reações de interconversões de açúcares-NDP e síntese de hemiceluloses e pectinas. As enzimas para as quais foram encontradas tags estão representadas na figura com setas pretas e todas as tags encontradas para uma enzima são apresentadas. Tags precedidas de um mesmo número sobrescrito representam isoformas de um mesmo gene. Para cada tag é fornecida uma ilustração colorida comparativa do nível de expressão entre as bibliotecas de folhas (F) e de madeira (M). A cor branca indica ausência e a cor cinza indica que não há diferença significativa entre as bibliotecas. As cores azul e vermelho foram utilizadas para indicar a ocorrência de diferença significativa, $P \leq 0,05$ (*) ou $P \leq 0,01$ (**), entre as bibliotecas, sendo que o vermelho indica maior expressão relativa e o azul indica menor expressão relativa. Setas cinza foram utilizadas para representar os passos enzimáticos para os quais nenhuma tag foi encontrada. GT8D, glicosil transferase da família 8D

As duas enzimas de maior representatividade nas reações de interconversões entre os açúcares-NDP foram a UDP-glicose desidrogenase (UGDH) e a UDP-glucuronato descarboxilase (UXS) que apresentaram no total 22 e 51 cópias na biblioteca de madeira, respectivamente. A UGDH é a principal enzima desta via, uma vez que regula o fluxo de UDP-D-glicose para a síntese de glicanas e pectinas funcionando no direcionamento da via como um passo limitante para a síntese dos polissacarídeos da parede celular (DALESSANDRO & NORTHCOTE, 1977).

Um total de quatro tags foi encontrado para a UXS na região cambial de *E. grandis*. Duas tags são similares a isoforma associada à membrana do Golgi, *UXS4* de arabis e que na

região cambial de *Eucalyptus grandis* estão presentes como isoformas de um único gene produzidas por *splicing* 3'. Já as outras duas *tags* de maior expressão são similares à isoforma solúvel, *UXS5* de *arabidopsis*, e provavelmente representam genes distintos em *Eucalyptus grandis* já que não há região de identidade entre os dois ESTs de *Eucalyptus*.

O nível de expressão das *tags* para as enzimas envolvidas na síntese de glicanas e pectinas foi de um modo geral, relativamente baixo quando comparado ao nível de expressão das *tags* para as enzimas relacionadas à síntese de celulose e lignina, sugerindo novamente que a maior parte das células da região cambial das árvores jovens amostradas estava em crescimento secundário ativo.

Além de ser utilizado como precursor na síntese de glicanas e pectinas, o UDP-D-glicose é, também, o precursor para a síntese das microfibrilas de celulose (Figura 25). Embora o UDP-D-glicose possa ser produzido tanto a partir de glicose 1P pela UDP-glicose pirofosforilase, quanto a partir de sacarose pela sacarose sintase (*SUSY*), o nível de expressão da *tag* referente à *SUSY* foi duas vezes maior que o encontrado para a UDP-glicose pirofosforilase. Este resultado certamente corrobora com o modelo atualmente proposto para a biossíntese de celulose onde complexos proteicos de membrana formados pela associação de proteínas celulose sintase, KORRIGAN (endo-1,4- β -glucanase) e *SUSY* (sacarose sintase) otimizam a síntese das microfibrilas de celulose (JOSHI *et al.*, 2004). Neste modelo a sacarose sintase possui a importante função de direcionar UDP-glicose diretamente para as proteínas celulose sintase do complexo, evitando, portanto, competição pelo *pool* celular de UDP-glicose e possibilitando, consequentemente, as elevadas taxas de síntese de celulose observadas durante o crescimento secundário (AMOR *et al.*, 1995). Uma única *tag* similar ao gene *SUSY1* de *Arabidopsis* foi identificada na biblioteca de madeira (Figura 25) para a qual foram observadas 45 cópias. Entretanto, não é possível afirmar com certeza se a *tag* identificada é referente à *SUSY1* ou *SUSY2* de *Arabidopsis*, pois as duas sequências diferem na região 5' e são muito similares na região 3', a partir da qual as SAGE *tags* são produzidas. Ambas *SUSY1* e *SUSY2* são representadas pelos transcritos relacionados ao metabolismo de carboidratos mais abundantes em *Populus* (GEISLER-LEE *et al.*, 2006), e são altamente expressos durante a formação de madeira de tensão nesta espécie (ANDERSON-GUNNERÅS *et al.*, 2006).

Figura 25 - Direcionamento de sacarose para síntese de celulose. As enzimas para as quais foram encontradas *tags* estão representadas na figura com setas pretas e todas as *tags* encontradas para uma enzima são apresentadas. *Tags* precedidas de um mesmo número sobrescrito representam isoformas de um mesmo gene. Para cada *tag* é fornecida uma ilustração colorida comparativa do nível de expressão entre as bibliotecas de folhas (F) e de madeira (M). A cor branca indica ausência e a cor cinza indica que não há diferença significativa entre as bibliotecas. As cores azul e vermelho foram utilizadas para indicar a ocorrência de diferença significativa, $P \leq 0,05$ (*) ou $P \leq 0,01$ (**), entre as bibliotecas, sendo que o vermelho indica maior expressão relativa e o azul indica menor expressão relativa. SUSY: sacarose sintase, KORRIGAN: endo-1,4- β -glucanase, *EgCesA1*, *EgCesA2*, *EgCesA3*: genes de *E. grandis* para celulose sintases relacionadas à síntese de parede secundária; *EgCesA4* e *EgCesA5*: genes de *E. grandis* para celulose sintases relacionadas à síntese de parede primária

A alta expressão de transcritos para sacarose sintase associada ao fato de que nenhuma *tag* referente às enzimas envolvidas na síntese e degradação de amido foi identificada na biblioteca de madeira é um forte indicativo de que a atividade da sacarose sintase é fundamental para produzir UDP-glicose e que o *pool* de UDP-glicose produzido está sendo direcionado para a síntese de celulose e não para a síntese de compostos de reserva. Este resultado corrobora com resultados anteriores que reportam a expressão de enzimas relacionadas à síntese de amido somente nas células cambiais em dormência, em gemas e órgãos de reserva, e não em células cambiais ativas (GEISLER-LEE *et al.*, 2006).

Recentemente foram identificados seis genes para celulose sintase em *Eucalyptus grandis* sendo três relacionados à síntese de parede primária (*EgCesA4*, *EgCesA5* e *EgCesA6*), e três mais expressos, relacionados à síntese de parede secundária (*EgCesA1*, *EgCesA2* e *EgCesA3*) (RANIK & MYBURG, 2006). Com exceção de *EgCesA6* para todos os outros genes foram encontradas

tags na biblioteca de madeira (Figura 25). No entanto, enquanto uma única *tag* foi identificada para *EgCesA2*, *EgCesA3* e *EgCesA5*, duas *tags* para *EgCesA1* e para *EgCesA4* foram encontradas. Através da análise do alinhamento produzido pelos ESTs de *Eucalyptus* referentes aos dois genes, foi possível observar que nos dois casos um único gene produz duas isoformas distintas por splicing 3'. A *tag* referente a uma das isoformas para *EgCesA1* foi a mais abundante entre todas as outras *tags* apresentando 63 cópias na biblioteca de madeira. As demais *tags* apresentaram número de cópias de no mínimo 4, encontrado para *EgCesA5*, à no máximo, 10-15 encontrados para *EgCesA2*, *EgCesA3* e *EgCesA4*. Dessa forma, os resultados observados são condizentes aos previamente reportados no sentido de que transcritos para celulose sintases relacionados à síntese de parede secundária são mais expressos na região cambial de *Eucalyptus grandis*. Mas são contrastantes ao revelar um nível de expressão bastante próximo para *EgCesA2* e *EgCesA3*, ambos relacionados à parede secundária, e *EgCesA4* que está relacionado à parede primária. Além disso, o gene de maior expressão no primeiro trabalho foi *EgCesA3* e não *EgCesA1* como encontrado no presente trabalho. Uma hipótese que explicaria esta discordância é a divergência na estrutura genética entre indivíduos de uma população natural, utilizados no primeiro trabalho, e os indivíduos da população utilizada no presente trabalho que representam um plantio comercial melhorado geneticamente para a produção de papel e celulose. Resultados anteriores de RANIK *et al.* (2006) também suportam esta hipótese. Os autores realizam um estudo do perfil transcricional de plantas altamente produtivas de *Eucalyptus* destinadas à produção de papel e celulose, e identificaram entre os genes de maior expressão, transcritos para subunidades de CESA similares a *PtrCesA1* e *AtCesA8*, os quais segundo a análise de alinhamento de sequências de aminoácidos, realizadas por RANIK & MYBURG (2006), são similares à *EgCesA1*.

Dois modelos para a síntese de microfibrilas de celulose são propostos, um mecanismo direto de síntese, e um mecanismo indireto. No mecanismo direto uma glicosiltransferase adiciona milhares de resíduos de glicose a longas cadeias sucessivamente. Já no mecanismo indireto, que tem sido suportado por evidências recentes, a síntese de celulose utiliza pequenos lipídeos ou proteínas como *primer* (DOBLIN *et al.*, 2002). Uma UDP-glicose:sterol glicosiltransferase possivelmente está envolvida na síntese de celulose através da síntese de um pequeno sitosterol- β -glicosídeo que serve como *primer* para a elongação β -1,4-glucana catalisada pelas Cesa. O produto é então clivado do *primer* por uma celulase endo-1,4- β -glucanase ou

Korrigan, e transferido para outra proteína CesA para dar continuidade a elongação da cadeia de celulose (DOBLIN *et al.*, 2002; PENG *et al.*, 2002). Três diferentes *tags* referentes à ESTs de *Eucalyptus* com alta similaridade à endo-1,4- β -glucanase foram identificadas na biblioteca de madeira. A *tag* de maior expressão apresentou 41 cópias na biblioteca de madeira e uma alta similaridade ao gene *EG2* de *Eucalyptus globulus*. As duas outras *tags* tiveram menor expressão (4-10 cópias) e ambos ESTs apresentaram similaridade ao gene *EG1* de *Eucalyptus globulus*, no entanto, o alinhamento dessas sequências não identificou nenhuma região de identidade sugerindo, portanto, que representam dois genes distintos e não isoformas de um único gene em *E. grandis*.

Aproximadamente 19 a 20% da parede celular de células maduras do xilema do *Populus* é formada por lignina (MacDOUGALL *et al.*, 1993). A lignificação inicia-se com a deposição de componentes da parede secundária e polímeros fenólicos de lignina formados a partir da polimerização de três diferentes álcoois hidroxicinamóis ou monolignóis, *p*-coumaril, coniferil e sinapil. Estes monolignóis dão origem às subunidades H, G e S, do polímero de lignina, respectivamente. Na Figura 26 é apresentado um resumo da via de síntese dos fenilpropanóides, que é comum a biossíntese dos precursores da lignina, com as *tags* encontradas para cada enzima nas bibliotecas.

Figura 26 - Síntese e polimerização dos monolignóis. As enzimas para as quais foram encontradas tags estão representadas na figura com setas pretas e todas as tags encontradas para uma enzima são apresentadas. Tags precedidas de um mesmo número sobrescrito representam isoformas de um mesmo gene. Para cada tag é fornecida uma ilustração colorida comparativa do nível de expressão entre as bibliotecas de folhas (F) e de madeira (M). A cor branca indica ausência e a cor cinza indica que não há diferença significativa entre as bibliotecas. As cores azul e vermelho foram utilizadas para indicar a ocorrência de diferença significativa, $P \leq 0,05$ (*) ou $P \leq 0,01$ (**), entre as bibliotecas, sendo que o vermelho indica maior expressão relativa e o azul indica menor expressão relativa. As setas tracejadas representam passos de conversão catalisados por F5H e COMT. PAL, fenilalanina amônia-liase; C4H, cinamato 4-hidroxilase; 4CL, 4-coumarato: coenzima A ligase; CCR, cinamoil-coenzima A redutase; CAD, cinamoil álcool desidrogenase; C3H, coumarato 3-hidroxilase; HCT, hidroxicinamoil-CoA shiquimato/quinato hidroxicinamoil-transferase; COMT, cafeato/5-hidroxi-coniferaldeído O-metiltransferase; CCoAOMT, cafeoil-coenzima A O-metiltransferase; F5H, ferulato/coniferaldeído 5-hidroxilase

A via dos fenilpropanóides inicia-se com a desaminação da fenilalanina a ácido cinâmico pela fenilalanina amônia liase (PAL). Um total de quatro tags para PAL foi identificado representando apenas dois genes distintos. As duas tags com menor número de cópias (4 e 3) apresentaram alta similaridade com os genes *PAL1* e *PAL2* de arabidopsis, e o alinhamento dos dois ESTs identificou a ocorrência de duas isoformas produzidas por *splicing* 3' a partir de um único gene. As outras duas tags possivelmente também representam isoformas de um único gene

com alta similaridade ao gene PAL de *Eucalyptus globulus*. Foram observadas 30 cópias para a *tag* de maior expressão e apenas 3 para a sua isoforma o que indica expressão preferencial de uma isoforma em relação à outra. Através da reação catalisada pela cinâmico ácido 4-hidrolase (C4H), o ácido cinâmico é convertido a ácido coumárico que é utilizado pela 4-coumarato:coenzima A ligase (4CL) para a produção dos tioésteres. Estes podem então ser utilizados na biossíntese das subunidades de lignina ou podem ser dirigidos à síntese de flavonóides (pigmentos), coumarinas (protetores contra UV), isoflavonóides e outros compostos relacionados à defesa em plantas (SUZUKI *et al.*, 2004).

Uma única *tag* com elevada expressão (36 cópias) foi identificada para C4H representando um EST com alta similaridade ao gene C4H2-1 de *Populus tremuloides*, o qual foi demonstrado estar mais especificamente expresso no xilema quando comparado a outros genes para C4H e em outros tecidos (LU *et al.*, 2006). Em *Populus tremuloides* foram identificados dois genes para C4H com duas isoformas cada, sendo a expressão destas isoformas relacionada a funções distintas no metabolismo de fenilpropanóides (LU *et al.*, 2006). Possivelmente, mais de um gene para esta enzima esteja também presente em *Eucalyptus grandis*, no entanto, os resultados do presente trabalho sugerem fortemente que um gene similar a *PtreC4H2-1* seja especificamente importante na região cambial e provavelmente, na lignificação de células do xilema de *Eucalyptus grandis*. Semelhantemente, uma *tag* com 24 cópias foi identificada para 4CL sendo seu EST de origem muito similar ao gene *4CL1* de *Eucalyptus camaldulensis* e de *Populus sp.* Dois genes para 4CL foram caracterizados em *Populus* com expressão diferencial sendo *4CL1* preferencialmente expresso no xilema quando comparado a *4CL2* (ALLINA *et al.*, 1998).

As enzimas PAL, C4H e 4CL estão presentes como pequenas famílias multigênicas em plantas com isoformas que exibem expressão tecido-específica e que vem sendo também correlacionadas a funções distintas no metabolismo dos fenilpropanóides em cada tecido (KAO *et al.*, 2002; RAES *et al.*, 2003; LU *et al.*, 2006). Embora estas enzimas representem as reações centrais do metabolismo dos fenilpropanóides, já que catalisam os passos iniciais da via, nenhuma alteração nos seus níveis de expressão foi observada durante a formação de madeira de tração (caracterizada pela redução no conteúdo de lignina) em *Eucalyptus* e *Populus*. Ao contrário, foram observadas reduções no número de transcritos para as enzimas que catalisam passos posteriores, específicos em cada braço da via que levam à formação das subunidades H, S

e G (PAUX *et al.*, 2005; ANDERSSON-GUNNERÅS *et al.*, 2006). Corroborando com estes resultados, foi observada na biblioteca de madeira uma maior expressão dos transcritos para enzimas nos passos mais específicos da via que para as enzimas das reações iniciais, principalmente para ácido cafeico O-metiltransferase (COMT), cafeoil-CoA O-metiltransferase (CCoAOMT), lacase e S-adenosil metionina sintase (SAMS).

O número de transcritos observados na biblioteca de madeira para COMT, CCoAOMT e SAMS foi também muito maior que o número de transcritos observados para cinamoil-CoA redutase (CCR) e cinamoil álcool desidrogenase (CAD). Nas angiospermas, a madeira é composta por grandes quantidades de unidades G e S, principalmente depositadas na parede secundária, enquanto pequenas quantidades de unidades H são predominantes apenas na lamela média e nas extremidades da parede celular (HOFFMANN *et al.*, 2004; LEWIS, 1999). Assim, seria realmente esperado que as enzimas direcionadas a síntese de S e G apresentem maior expressão que as direcionadas à síntese de unidades H em *Eucalyptus*. Além disso, foi demonstrado que genes para 4CL, CCR e CAD formam um *cluster* comum de controle transcricional enquanto genes para COMT e CCoAOMT formam outro *cluster* de co-regulação transcricional (PAUX *et al.*, 2005).

A única *tag* identificada para CCR na biblioteca de madeira apresentou uma baixa expressão (7 cópias) e alta similaridade ao gene CCR de *Eucalyptus gunnii*, único gene para a enzima identificado até então na espécie (LACOMBE *et al.*, 1997). Por outro lado, no genoma de *Populus*, recentemente disponibilizado no site <http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>, oito sequências homólogas a genes CCR estão presentes. Para a CAD, três *tags* diferentes foram encontradas, uma de alta expressão (37 cópias), e as outras duas de baixa expressão (3 e 8 cópias). As duas CAD *tags* mais expressas são similares aos genes de *Eucalyptus globulus* e *E. gunnii*, mas o alinhamento dos ESTs de *Eucalyptus* não produz nenhuma região de identidade indicando que provavelmente, as duas *tags* representam sequências transcritas a partir de genes diferentes em *E. grandis*. Foi interessante notar que o EST para a *tag* com menor número de transcritos não apresentou similaridade com os genes CAD previamente descritos para *Eucalyptus*, mas foi homólogo ao gene CAD de *Pinus taeda*.

O passo determinante para o direcionamento de *p*-coumaril-CoA para a síntese de unidades G e S, é sua conversão ao ácido *p*-coumaril shiquimico/ácido quinico pela hidroxicinamoil-transferase (HCT), já que a C3H não é capaz de utilizar *p*-coumaril-CoA como

substrato (SCHÖCH *et al.*, 2001). Foi demonstrado que a enzima HCT catalisa a formação dos substratos *p*-coumaroil shiquimato/quinato e do precursor de cafeoil-CoA (HOFFMANN *et al.*, 2004). Uma única *tag* com 12 cópias foi identificada na biblioteca de madeira cujo EST apresentou similaridade ao gene HCT de tabaco (HOFFMANN *et al.*, 2004).

Durante a biossíntese dos monolignóis, reações que envolvem metilação são requeridas em duas posições do anel aromático da unidade de monolignol. Uma ocorre na posição 3 e a outra na posição 5. A metilação na posição 3 leva à formação de unidades G enquanto a dupla metilação nas posições 3 e 5 levam à formação de unidades S. A metilação da posição 3 ocorre sobre o intermediário CoA éster e é catalisada pela CCoAOMT, enquanto a metilação na posição 5 ocorre sobre o intermediário aldeído sendo catalisada pela COMT (Revisado em LI *et al.*, 2006). Embora tenha sido primeiramente acreditado que a COMT convertia ácido cafeico em ferulato (DIXON, 2001), evidências posteriores demonstraram que a COMT catalisa preferencialmente a conversão de 5-hidroxiferulato, 5-hidroxiciniferaldeído e 5-hidroxiciniferil álcool em ácido sinápico, sinapaldeído e álcool sinapil, respectivamente, e, portanto, sua atividade ocorre preferencialmente sobre os produtos da reação catalisada pela F5H (PARVATHI *et al.*, 2001). As *tags* para F5H e COMT, entretanto, apresentaram padrões de expressão bastante distintos. Foram encontrados 22 transcritos para F5H enquanto 52 transcritos foram observados para COMT na biblioteca de madeira. Estes resultados sugerem que provavelmente, a transcrição de F5H e COMT não seja co-regulada. Da mesma forma, proteínas F5H não estiveram presentes durante a regeneração cambial em *Populus*, enquanto isoformas proteicas de COMT foram detectadas em todos os estádios avaliados (JUAN *et al.*, 2006).

A *tag* para COMT com maior número de cópias (39), representou um EST com alta similaridade ao gene *OMT* de *Eucalyptus gunnii* (POEYDOMENGE *et al.*, 1994) enquanto o EST que deu origem à segunda *tag* (13 cópias na biblioteca de madeira) não apresentou similaridade com o gene de *Eucalyptus*, mas sim com o gene *AtOMT1*. O alinhamento das duas sequências de ESTs não identificou nenhuma região de identidade corroborando com a hipótese de que dois genes para COMT sejam preferencialmente expressos durante a formação do xilema em *Eucalyptus grandis*.

A CCoAOMT é uma enzima chave na produção dos monolignóis, que atua na metilação de cafeoil-CoA gerando feruloil-CoA. Altos níveis de expressão de transcritos para CCoAOMT foram encontrados em células de *Populus* durante o processo de lignificação (ZHONG *et al.*,

2000). Em um estudo recentemente realizado sobre o perfil transcricional de células do xilema de *Eucalyptus*, foi construída uma biblioteca subtrativa usando tecido de folhas como testador, e somente três genes relacionados à síntese de lignina foram demonstrados como altamente e especificamente expressos no xilema. Entre estes genes estavam dois genes de lacases e um gene CCoAOMT (PAUX *et al.*, 2004). Proteínas CCoAOMT foram também identificadas como uma das mais abundantes proteínas durante a diferenciação do xilema em *Populus* (VANDER MIJNSBRUGGE *et al.*, 2000). No presente trabalho, entre os transcritos identificados para as enzimas envolvidas na síntese de lignina, os transcritos para CCoAOMT foram também os mais abundantes com um total de 112 cópias na biblioteca de madeira. As duas *tags* encontradas para CCoAOMT representaram ESTs com alta similaridade aos genes CCoAOMT de *Eucalyptus gunnii* e *E. globulus* sendo a de menor expressão (42 cópias) similar a CCoAOMT2 de *E. globulus* e a de maior expressão (70 cópias) similar a *EgCCoAOMT1*.

A conversão do ácido cinâmico em unidades G e S requer adição de grupos metil em posições específicas dos monolignóis levando a um grande consumo desses grupamentos nas células do xilema durante a lignificação. Tanto na reação catalisada por COMT quanto na reação catalisada por CCoAOMT, a S-adenosilmetionina (SAM) é utilizada como doadora de grupos metil (CANTÓN *et al.*, 2005). A enzima que catalisa a síntese de SAM, S-adenosilmetionina sintase (SAMS) é ativa em muitos processos celulares sendo consequentemente considerada uma enzima *housekeeping* (HANSON & ROJE, 2001). Entretanto, transcritos para SAMS têm sido encontrados em elevada abundância no xilema durante o crescimento secundário em *Populus*, *arabidopsis* e outras espécies (STERKY *et al.*, 1998, VANDER MIJNSBRUGGE *et al.*, 2000; OH *et al.*, 2003; JUAN *et al.*, 2006). Foi encontrado, na biblioteca de madeira, um total de 96 transcritos para SAMS corroborando com os resultados anteriores que relacionam a expressão desta enzima ao crescimento secundário ativo. A elevada expressão de SAMS durante a lignificação funciona como um importante mecanismo local fornecendo precursores necessários ao processo, e assim garantindo a manutenção de altas taxas de lignificação que são observadas durante o crescimento secundário (CANTÓN *et al.*, 2005). O alinhamento dos ESTs para SAMS sugere fortemente que 2 genes para esta enzima sejam preferencialmente expressos durante a lignificação em *Eucalyptus grandis* sendo o gene de maior expressão representado por duas isoformas distintas, uma com 48 e a outra com 31 cópias, na biblioteca de madeira. É interessante notar que os níveis de expressão dos transcritos para SAMS e CCoAOMT foram bastante

próximos. Possivelmente, a correlação funcional das duas enzimas possa determinar um controle transcricional comum para ambas, embora nenhuma evidência neste sentido tenha sido ainda encontrada.

O último passo na síntese de lignina é a polimerização dos monolignóis que se acredita ser conduzida por peroxidases e lacases (BAUCHER *et al.*, 2003). A importância das lacases durante o crescimento secundário tem sido demonstrada em espécies de *Populus*, *Pinus* e *Eucalyptus* (ALLONA *et al.*, 1998; PAUX *et al.*, 2004; ANDERSSON-GUNNERÅS *et al.*, 2006). Enquanto seis *tags* com um total de 103 cópias, representando pelo menos cinco genes diferentes, foram identificadas para lacases na biblioteca de madeira, uma única *tag* de baixa expressão (2 cópias) foi identificada para uma peroxidase. Este resultado sugere que as lacases são relativamente muito mais importantes que as peroxidases durante o crescimento secundário em *Eucalyptus grandis*. Esta hipótese é também sugerida nos resultados obtidos por ANDERSSON-GUNNERÅS *et al.* (2006), onde nenhum transcrito para peroxidases teve sua expressão reprimida durante a formação de madeira de tração em *Populus*.

Recentemente foi também proposto que *dirigent proteins* atuam mediando a polimerização dos monômeros de lignina principalmente durante a lignificação em S1 e na lamela média de células do xilema (BURLAT *et al.*, 2001) embora nenhuma evidência deste processo tenha sido reportada até o momento. Corroborando com esta hipótese, foram encontradas *tags* para *dirigent proteins* na biblioteca de madeira com elevada expressão (49 cópias). Duas *tags* de menor expressão (3 e 8 cópias) representaram isoformas similares ao gene DIR de *Forsythia x intermédia*, e a *tag* de maior expressão (38 cópias), representou um EST similar ao gene DIR de *Picea glauca* o qual foi demonstrado estar relacionado à síntese de compostos fenólicos (lignanas e ligninas) de resistência ao ataque de insetos em coníferas (RALPH *et al.*, 2006). Como o alinhamento desses EST não identifica região de identidade, possivelmente dois genes DIR distintos estejam atuando na polimerização de monolignóis durante a lignificação em *E. grandis*.

A lignificação intensa da parede celular dos elementos de vaso e fibras que ocorre durante o crescimento secundário requer um alto consumo de N através da desaminação da fenilalanina pela PAL liberando amônia. Por si só, o conteúdo de N consumido nas células em intenso processo de lignificação já seria suficiente para causar um *défict* nutricional na planta caso todo N consumido fosse suprido pela assimilação direta através da redução de nitrato. Ao contrário disso,

nas células em crescimento secundário ocorre um importante mecanismo local de reciclagem de N através da amônia liberada na reação catalisada pela PAL, fornecendo o suprimento necessário de N para síntese de fenilalanina (KEYS, 2006; CÁNTON *et al.*, 2005). Este mecanismo local funcionaria através do sistema GS/GOGAT como sugerido por CÁNTON *et al.* (2005). Corroborando com esta hipótese, três *tags* diferentes foram encontradas com altíssimo nível de expressão na biblioteca de madeira (120 cópias) para a enzima glutamina sintetase (Figura 27).

Duas classes distintas de GS estão presentes em plantas, a GS citosólica ou GS1 e a GS cloroplastidial ou GS2 e, na maioria das plantas, várias isoformas de GS1 são codificadas por uma família de genes (CREN & HIREL, 1999). Consistente com resultados anteriores, as *tags* de ESTs similares à GS identificadas na biblioteca de madeira apresentaram homologia com genes GS1 e não foi identificada região de identidade entre estas sequências sugerindo que em *E. grandis*, ao menos três genes GS1 estão presentes.

Figura 27 - Reciclagem local de amônia durante a síntese e degradação de fenilalanina. As enzimas para as quais foram encontradas *tags* estão representadas na figura com setas pretas e todas as *tags* encontradas para uma enzima são apresentadas. *Tags* precedidas de um mesmo número sobrescrito representam isoformas de um mesmo gene. Para cada *tag* é fornecida uma ilustração colorida comparativa do nível de expressão entre as bibliotecas de folhas (F) e de madeira (M). A cor branca indica ausência e a cor cinza indica que não há diferença significativa entre as bibliotecas. As cores azul e vermelho foram utilizadas para indicar a ocorrência de diferença significativa, $P \leq 0,05$ (*) ou $P \leq 0,01$ (**), entre as bibliotecas, sendo que o vermelho indica maior expressão relativa e o azul indica menor expressão relativa. PAL, fenilalanina amônia-liase; GS1, glutamina sintetase citoplasmática

Quando a lignificação é completada, os elementos de vaso do xilema entram no último estágio de sua diferenciação, a morte celular programada (PCD) seguida pela autólise do conteúdo celular. Foram identificadas na biblioteca de madeira *tags* para muitas das proteínas que

vem sendo correlacionadas à morte celular programada e autólise celular em plantas tais como cisteínas protease, *SLP2* (subtilisin serina protease), *Endo93*, nodulina *MtN21*, reguladores de apoptose *Bcl-2* e *Bax inhibitor-1* (MOREAU *et al.*, 2005). Um grande número de transcritos foi também encontrado representando a via proteassomo-ubiquitina de degradação de proteínas a qual foi recentemente correlacionada a PCD (MOREAU *et al.*, 2005). Entre os genes relacionados à morte celular programada identificados na biblioteca de madeira, o maior número de transcritos foi encontrado para poliubiquitinas (299 cópias) e subunidades de proteassomo (67 cópias), cisteína proteases (168 cópias) e para a proteína reguladora de apoptose *Bcl-2* (74 cópias).

A via proteassomo-ubiquitina é o principal sistema proteolítico nas plantas. Nesta via, proteínas ubiquitina (Ub) funcionam como sinal de reconhecimento para o *turnover* seletivo de proteínas. As proteínas Ub se ligam às proteínas alvo que são então degradadas pelo complexo proteassomo 26S liberando as Ubs para reutilização (SMALLE & VIERSTRA, 2004). A via Ub/ proteassomo 26S está diretamente ou indiretamente relacionada à ação da maioria dos hormônios nas plantas (HELLMAN & ESTELLE, 2002). Vias de resposta às auxinas foram as primeiras para as quais identificou-se um importante papel da via Ub/ proteassomo 26S na sinalização hormonal (SMALLE & VIERSTRA, 2004). Dada a importância das auxinas na determinação da diferenciação em células do xilema em diferentes espécies (BERLETH & SACHS, 2001; KO *et al.*, 2004), sugere-se atualmente que a via Ub/ proteassomo 26S abundantemente encontrada durante o desenvolvimento do xilema secundário, participe da via de sinalização de resposta às auxinas no xilema (PAUX *et al.*, 2004). Assim, a elevada expressão de transcritos para a via Ub/ proteassomo 26S na biblioteca de madeira indica sua importância durante a formação de madeira em *E. grandis* que pode estar relacionada à morte celular programada dos elementos de vaso, bem como a sinalização de respostas às auxinas.

5.4 Perfil comparativo da expressão gênica entre folhas e madeira

Enquanto os órgãos fotossintetizantes são capazes de utilizar a energia solar para produzir energia química (ATP) e reduzir o carbono assimilado em pequenos carboidratos, estes carboidratos são utilizados nos órgãos não fotossintetizantes para, através da respiração, liberar a energia necessária para o metabolismo celular, e sintetizar os precursores envolvidos na biossíntese de muitas biomoléculas. Assim, a maior expressão de transcritos obtida para as

enzimas relacionadas à respiração na biblioteca de madeira quando comparada à biblioteca de folhas foi um resultado já esperado (Figura 20 e Tabela 9).

Tabela 9 - *Tags* relacionadas aos genes nucleares da cadeia respiratória mitocondrial encontradas nas bibliotecas SAGE. A sequência de cada *tag* é apresentada, seguida pelo seu número de cópias nas bibliotecas de folhas (F) e de madeira (M), identificação e valor *E*. A diferença estatística significativa entre o número de cópias de uma *tag* nas bibliotecas, quando presente, é representada por (**) para $P \leq 0,01$, e (*) para $P \leq 0,05$ (Teste Z). *Tags* precedidas de um mesmo número sobrescrito indicam isoformas para um mesmo gene

<i>Tag</i>	Nº cópias		Identificação	Valor <i>E</i>
	F	M		
1	4	4	gil21593880 NADH dehydrogenase/acetyl CoA-sintetase	5e-106
2	2	3	gil30693102 NADH dehydrogenase ubiquinone/electron transport	6e-15
3	0	4	gil18395290 NADH dehydrogenase ubiquinone/electron transport	3e-32
4	3	3	gil50509945 NADH-ubiquinone oxidoreductase	9e-35
5	7	18**	gil57226937 NADH-ubiquinone oxidoreductase	8e-06
6	4	13*	gil16930745 NADH-ubiquinone oxidoreductase	1e-66
7	22	16	gil21536569 NADH:ubiquinone oxidoreductase	4e-49
8	5	4	gil31433693 NADH:ubiquinone oxidoreductase 17.2 kD subunit	1e-65
¹ 9	12	78**	gil50844677 cytochrome b5 isoform Cb5-C	8e-62
¹ 10	4	11	gil50844677 cytochrome b5 isoform Cb5-C	2e-61
11	1	3	gil92872511 Cytochrome c C1 family)	4e-26
12	5	3	gil118018 Cytochrome c2	1e-48
² 13	2	3	gil15239430 ubiquinol-cytochrome-c reductase 14kD subunit	1e-42
² 14	1	4	gil15239430 ubiquinol-cytochrome-c reductase 14kD subunit	2e-42
³ 15	1	9*	gil50844677 cytochrome b5 isoform Cb5-C	8e-60
³ 16	0	3	gil2647949 cytochrome b5	3e-44
17	3	5	gil15218058 ubiquinol-cytochrome C reductase 7,8 kDa protein	2e-27
18	1	4	gil21592929 cytochrome c oxidase subunit Vb	2e-37
⁴ 19	5	7	gil15226582 cytochrome-c oxidase subunit Vc	2e-17
⁴ 20	2	4	gil15226582 cytochrome-c oxidase subunit Vc	1e-17
⁴ 21	2	4	gil2275216 cytochrome c oxidase Vc subunit	6e-19
⁴ 22	3	1	gil15226582 cytochrome c oxidase Vc subunit	1e-25
23	2	13**	gil31745230 cytochrome c oxidase subunit VIa precursor	7e-37
24	2	4	gil21592329 cytochrome c oxidase subunit VIb	2e-37
25	7	9	gil19685 ATP synthase beta subunit	7e-22
26	12	7	gil12597251 mitochondrial ATP synthase 6 KD subunit	1e-16
27	25	50**	gil48209968 mitochondrial ATP synthase D chain	3e-57

O uso das SAGE *tags* para identificação de genes foi eficiente para possibilitar a diferenciação entre isoformas citoplasmáticas e cloroplastidiais de enzimas que estão presentes em ambos locais, mas que assumem funções distintas no citoplasma e no cloroplasto. Este foi o caso, por exemplo, para a frutose-bisfosfato aldolase citoplasmática, que catalisa a formação de gliceraldeído 3P durante a glicólise, e que esteve muito mais representada na biblioteca de madeira, enquanto a aldolase cloroplastidial, que participa na regeneração de ribulose-1,5-bisfosfato, foi quase especificamente expressa na biblioteca de folhas (Tabela 10). Outros exemplos são a gliceraldeído 3P desidrogenase, cujos transcritos foram identificados como

citoplasmáticos corroborando com sua maior expressão na biblioteca de madeira, a fosfoglicerato quinase, que foi identificada por uma única *tag* referente a um EST similar a isoforma cloroplastidial e que foi especificamente expressa na biblioteca de folhas, e a transquetolase, que está relacionada ao ciclo de Benson-Calvin, e para a qual foi identificada uma *tag* especificamente expressa na biblioteca de folhas cujo EST de origem foi similar a isoforma cloroplastidial da enzima (Tabela 10).

Tabela 10 - Lista de genes diferencialmente expressos entre as bibliotecas de folhas e madeira ($P \leq 0,01$). Para cada *tag* é fornecido o número de cópias na biblioteca de Folhas e na biblioteca de Madeira, a identificação do gene fornecida pelo NCBI bem como o valor de *E* de cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou cDNA de origem

<i>Tag</i>	Nº de Cópias		Identificação	Valor <i>E</i>
	F	M		
1	518	122	Não identificado	-
2	338	56	gil3417451 light-harvesting chlorophyll a/b-binding protein	9e-80
3	273	28	gil1223652 metallothionein type I	2e-05
4	238	30	gil25992534 fiber lipid transfer protein	1e-34
5	274	62	gil77157637 Rubisco/small subunit	2e-82
6	170	16	Sem registro	-
7	124	21	gil3036945 light-harvesting chlorophyll a/b-binding protein	6e-139
8	129	25	gil2213869 ferredoxin I precursor	3e-44
9	84	4	gil3694984 metallothionein-I	2e-28
10	117	26	gil5608497 galactinol synthase	8e-153
11	101	20	Não identificado	-
12	66	4	gil18399660 Chloroplast Fructose-bisphosphate aldolase	0.0
13	96	20	gil225895 photosystem I protein CAB	2e-113
14	103	26	gil1773330 glycolate oxidase	0.0
15	209	97	gil1566700 metallothionein-like TYPE3	2e-24
16	75	13	Sem registro	-
17	74	13	Não identificado	-
18	85	20	gil1838961 acyl carrier protein	1e-33
19	58	7	gil92877555 photosystem II core complex proteins psby	9e-21
20	67	14	Sem registro	-
21	96	31	gil19911 photosystem II 23kDa polypeptide	e-106
22	64	14	gil92876895 Glycine hydroxymethyltransferase	0.0
23	78	22	gil4588376 galactinol synthase	e-143
24	38	2	Sem registro	-
25	38	2	Sem registro	-
26	99	36	Não identificado	-
27	38	3	gil18405145 RCA (RUBISCO ACTIVASE)	0.0
28	54	11	gil728716 protein of photosystem II	2e-20
29	37	3	Sem registro	-
30	32	1	Sem registro	-
31	85	30	gil15810139 unknown protein	e-176
32	44	7	gil82080 chlorophyll a/b-binding protein type III	6e-124
33	30	1	gil23308219 At5g64840	1e-44

(continua)

Tabela 10 - Lista de genes diferencialmente expressos entre as bibliotecas de folhas e madeira ($P \leq 0,01$). Para cada *tag* é fornecido o número de cópias na biblioteca de Folhas e na biblioteca de Madeira, a identificação do gene fornecida pelo NCBI bem como o valor de *E* de cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou cDNA de origem

(continuação)

<i>Tag</i>	Nº Cópias		Identificação	Valor <i>E</i>
	F	M		
34	37	4	Sem registro	-
35	64	19	gil3126965 basic chitinase	8e-40
36	40	6	gil98962139 unknown protein	3e-20
37	60	17	gil2920666 glutathione S-transferase	2e-97
38	43	8	gil18420175 threonine endopeptidase	1e-16
39	26	1	gil32526566 catalase	0.0
40	35	5	gil431957 thioredoxin m	1e-47
41	25	1	Sem registro	-
42	27	2	Sem registro	-
43	41	9	Não identificado	-
44	29	3	Sem registro	-
45	33	5	Sem registro	-
46	28	3	Sem registro	-
47	34	6	gil5170541 UDP-glucose glucosyltransferase (GT44)	0.0
48	43	11	Sem registro	-
49	146	86	Sem registro	-
50	33	6	gil37222949 Rieske Fe-S protein precursor	2e-88
51	30	5	gil33438474 chlorophyll a /b binding protein	9e-129
52	35	8	gil929813 plastocyanin a	2e-51
53	36	9	gil34420138 glycine decarboxylase complex H-protein	7e-75
54	19	1	Sem registro	-
55	19	1	Sem registro	-
56	61	26	Sem registro	-
57	21	2	gil19787 chloroplast ATP synthase	8e-71
58	18	1	Sem registro	-
59	127	78	Sem registro	-
60	20	2	Sem registro	-
61	17	1	Sem registro	-
62	17	1	Sem registro	-
63	17	1	Sem registro	-
64	29	7	Não identificado	-
65	27	6	gil6996562 putative 16kDa membrane protein	9e-56
66	30	8	gil248303 ribosomal protein L12	9e-26
67	28	7	Sem registro	-
68	51	22	gil18086 cysteine proteinase	7e-29
69	18	2	Sem registro	-
70	18	2	Sem registro	-
71	18	2	gil30683145 plastoquinol-plastocyanin reductase	7e-16
72	18	1	gil3738261 chloroplast phosphoglycerate kinase	2e-81
73	15	1	gil94549022 ribulose-1,5-bisphosphate carboxylase	1e-35
74	15	1	Sem registro	-
75	17	2	Sem registro	-
76	19	3	Sem registro	-
77	19	3	Não identificado	-
78	31	10	gil11994536 Unknown protein	2e-41
79	31	10	Sem registro	-
80	34	12	Sem registro	-

Tabela 10 - Lista de genes diferencialmente expressos entre as bibliotecas de folhas e madeira ($P \leq 0,01$). Para cada *tag* é fornecido o número de cópias na biblioteca de Folhas e na biblioteca de Madeira, a identificação do gene fornecida pelo NCBI bem como o valor de *E* de cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou cDNA de origem

(continuação)

<i>Tag</i>	Nº de Cópias		Identificação	Valor <i>E</i>
	F	M		
81	16	2	Sem registro	-
82	16	2	Sem registro	-
83	18	3	gil28411216 plastid ribosomal protein L	5e-28
84	18	3	Não identificado	-
85	36	14	gil1181599 subunit of photosystem I	7e-82
86	28	9	Não identificado	-
87	13	1	Sem registro	-
88	13	1	Sem registro	-
89	19	4	Sem registro	-
90	19	4	gil693920 chlorophyll A/B binding protein	8e-78
91	29	10	Sem registro	-
92	29	10	Sem registro	-
93	35	14	Sem registro	-
94	24	7	Sem registro	-
95	62	34	gil1545805 cytochrome P450 like	4e-82
96	12	1	Sem registro	-
97	12	1	Sem registro	-
98	12	1	Sem registro	-
99	16	3	Sem registro	-
100	16	3	Sem registro	-
101	14	2	Sem registro	-
102	19	5	Sem registro	-
103	19	5	Sem registro	-
104	44	22	Sem registro	-
105	27	10	gil33771374 zinc finger protein PIF1 (transcription factor)	6e-20
106	17	4	gil45356861 ultraviolet-B-repressible protein	5e-18
107	11	1	gil33438474 chlorophyll a /b binding protein	1e-115
108	11	1	Sem registro	-
109	11	1	Sem registro	-
110	11	1	Sem registro	-
111	11	1	Sem registro	-
112	11	1	Sem registro	-
113	11	1	gil18407178 TMP14 (thylakoid membrane phosphoprotein)	2e-30
114	11	1	gil15235029 LHCB5 (light harvesting complex)	7e-127
115	11	1	gil3023743 Ferredoxin-1, chloroplast precursor	3e-44
116	15	3	Sem registro	-
117	15	3	Sem registro	-
118	13	2	Sem registro	-
119	28	11	gil46798907 membrane related protein	3e-25
120	35	16	gil5608497 galactinol synthase	4e-43
121	26	10	unknown protein	-
122	18	5	gil18644694 susceptibility homeodomain transcription factor	6e-18
123	18	5	gil108802653 photosystem I subunit VIII	2e-12
124	18	5	Sem registro	-
125	21	7	Sem registro	-
126	16	4	Sem registro	-
127	16	4	Sem registro	-

Tabela 10 - Lista de genes diferencialmente expressos entre as bibliotecas de folhas e madeira ($P \leq 0,01$). Para cada *tag* é fornecido o número de cópias na biblioteca de Folhas e na biblioteca de Madeira, a identificação do gene fornecida pelo NCBI bem como o valor de *E* de cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou cDNA de origem

(continuação)

<i>Tag</i>	Nº de Cópias		Identificação	Valor <i>E</i>
	F	M		
128	10	1	Sem registro	-
129	10	1	gil66271031 Unknown protein	3e-11
130	10	1	Sem registro	-
131	10	1	Sem registro	-
132	10	1	Sem registro	-
133	10	1	Sem registro	-
134	10	1	Sem registro	-
135	10	1	Sem registro	-
136	10	1	Sem registro	-
137	12	2	gil3694984 metallothionein-1	3e-28
138	19	6	gil71676716 ATP-dependent Clp protease	8e-21
139	19	6	gil42572635 high mobility group protein alpha (HMGalpha)	3e-25
140	39	20	Sem registro	-
141	28	12	gil6561948 Nascent polypeptide-associated complex NAC	3e-32
142	22	8	gil50660327 chloroplast chaperonin 21	1e-112
143	60	37	Sem registro	-
144	29	13	gil74147576 acyl-CoA synthetase	2e-92
145	29	13	Sem registro	-
146	26	11	gil89257423 plastid ribosomal protein L19	4e-13
147	20	7	Sem registro	-
148	23	9	Sem registro	-
149	4	17	gil21555669 remorin	2e-16
150	4	17	Sem registro	-
151	3	15	Não identificado	-
152	3	15	gil50909971 serine/threonine-protein kinase	8e-10
153	3	15	gil21592628 nitrilase-associated protein	6e-20
154	10	28	gil21555203 nuclear transport factor 2	2e-50
155	2	13	Sem registro	-
156	2	13	Não identificado	-
157	2	13	Sem registro	-
158	2	13	Sem registro	-
159	2	13	gil31745230 putative cytochrome c oxidase	7e-37
160	7	23	Sem registro	-
161	8	25	Não identificado	-
162	12	32	gil42565401 ubiquitin-conjugating enzyme	1e-31
163	1	11	Sem registro	-
164	1	11	gil56797970 BCL-2 binding anthanogene-1	1e-11
165	1	11	gil92887677 Reticulon	6e-21
166	1	11	Sem registro	-
167	1	11	Sem registro	-
168	1	11	Sem registro	-
169	1	11	gil974782 cobalamine-independent methionine synthase	2e-21
170	1	11	gil4126473 adenylyl cyclase associated protein	4e-169
171	1	11	gil29691168 adenosylhomocysteinase	6e-92
172	13	34	gil21886603 peptidylprolyl isomerase	1e-81
173	10	29	gil9755608 phytochelatin synthetase	9e-34
174	4	18	gil39653273 ADP-ribosylation factor	e-100

Tabela 10 - Lista de genes diferencialmente expressos entre as bibliotecas de folhas e madeira ($P \leq 0,01$). Para cada *tag* é fornecido o número de cópias na biblioteca de Folhas e na biblioteca de Madeira, a identificação do gene fornecida pelo NCBI bem como o valor de *E* de cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou cDNA de origem

<i>Tag</i>	Nº de Cópias		Identificação	Valor <i>E</i>
	F	M		
175	4	18	Sem registro	-
176	5	20	gil6979167 Profilin	1e-54
177	6	22	gil15235599 cysteine-type endopeptidase	1e-22
178	3	16	gil1420887 non-specific lipid transfer-like protein	4e-15
179	7	24	gil56550695 Unknown protein	3e-130
180	2	14	gil47717927 fasciclin-like AGP 12	6e-66
181	2	14	gil1199774 extensin like protein	2e-24
182	2	14	Não identificado	-
183	12	33	gil483536 pyrophosphate-dependent phosphofructokinase PFK	0.0
184	1	12	Sem registro	-
185	1	12	gil14331109 beta-tubulin	5e-57
186	1	12	Sem registro	-
187	1	12	gil38194917 phytochelatin synthetase	3e-17
188	1	12	Não identificado	-
189	5	21	gil25809052 heat shock protein	4e-60
190	4	19	gil18377546 zinc finger-like protein	4e-08
191	3	17	gil62526583 auxin-repressed protein	3e-37
192	8	27	Não identificado	-
193	8	27	gil57232107 beta-galactosidase	0.0
194	23	52	gil19702 eukaryotic initiation factor 5A	8e-83
195	2	15	gil92877351 IQ calmodulin-binding region	3e-25
196	2	15	Não identificado	-
197	2	15	gil56202296 membrane protein-like	1e-54
198	2	15	gil21592366 clathrin assembly protein	2e-78
199	2	15	Sem registro	-
200	1	13	gil460991 beta tubulin	4e-38
201	1	13	Sem registro	-
202	12	35	gil6969575 phospholipase C1	3e-20
203	7	26	gil55297407 phosphatidylinositol 3- and 4-kinase	0.0
204	3	18	Sem registro	-
205	3	18	gil21555203 nuclear transport factor 2	2e-17
206	2	16	Sem registro	-
207	2	16	Sem registro	-
208	2	16	Sem registro	-
209	2	16	Sem registro	-
210	6	25	gil11602747 early nodulin ENOD18	1e-35
211	4	21	gil50900026 unknown protein	1e-33
212	1	14	gil52076690 cdc2 protein kinases	2e-32
213	1	14	Sem registro	-
214	1	14	Sem registro	-
215	1	14	gil11071902 laccase	7e-39
216	3	19	gil2459926 annexin	3e-38
217	14	40	gil3402279 beta-subunit of K ⁺ channels	4e-37
218	13	39	gil1938236 acyl-CoA-binding protein	1e-38
219	6	26	gil38198150 aquaporin	e-117
220	7	28	gil92874845 Targeting for Xklp2	1e-54
221	5	24	gil15234016 phospholipase C	5e-52

Tabela 10 - Lista de genes diferencialmente expressos entre as bibliotecas de folhas e madeira ($P \leq 0,01$). Para cada *tag* é fornecido o número de cópias na biblioteca de Folhas e na biblioteca de Madeira, a identificação do gene fornecida pelo NCBI bem como o valor de *E* de cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou cDNA de origem

(continuação)

<i>Tag</i>	Nº de Cópias		Identificação	Valor <i>E</i>
	F	M		
222	5	24	gil34925093 Remorin	4e-20
223	5	24	gil15232389 protein kinase family	e-119
224	4	22	gil37535352 unknown protein	1e-05
225	3	20	Não identificado	-
226	12	38	gil88771143 dirigent-like protein	3e-24
227	16	45	gil21902507 ribosomal protein small subunit 28	1e-19
228	2	18	gil22158 alpha-tubulin	0.0
229	2	18	gil14531296 pectate lyase	2e-88
230	2	18	Sem registro	-
231	2	18	gil2398679 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	1e-96
232	8	31	gil23397103 Methionine synthase	2e-157
233	8	31	gil13540318 S-adenosyl-L-methionine synthetase	5e-143
234	5	25	Sem registro	-
235	5	25	gil33340236 copper/zinc superoxide dismutase	3e-77
236	4	23	gil34851124 glycine-rich RNA-binding protein	2e-37
237	1	16	Sem registro	-
238	1	16	gil18418552 subtilase	1e-12
239	1	16	Não identificado	-
240	1	16	gil17979101 protein kinase	5e-132
241	7	30	gil5690433 phenylalanine ammonia lyase	3e-145
242	2	19	gil18420880 transporter	4e-15
243	2	19	gil60657598 Cell wall-related glycosyltransferase family GT47	7e-55
244	2	19	gil70779685 xyloglucan endotransglycosylase	8e-54
245	2	19	Não identificado	-
246	4	24	gil4539543 glyceraldehyde-3-phosphate dehydrogenase	e-164
247	1	17	Sem registro	-
248	1	17	Sem registro	-
249	1	17	Sem registro	-
250	5	27	Sem registro	-
251	4	25	Não identificado	-
252	54	108	gil120676 Glyceraldehyde 3-phosphate dehydrogenase	3e-21
253	1	18	Não identificado	-
254	5	28	gil730557 60S ribosomal protein L34	2e-47
255	13	44	gil20413 alpha-tubulin	0.0
256	9	37	gil4336893 cinnamyl alcohol dehydrogenase	3e-54
257	1	19	gil6118549 ubiquitin precursor	1e-42
258	6	31	gil56606534 aluminum-induced protein	8e-36
259	5	29	gil50251393 germin-like protein	5e-16
260	2	22	gil15222565 ATMYB103 DNA binding / transcription factor	2e-04
261	2	22	gil4640321 coniferylalcohol 5-hydroxylase	2e-33
262	4	27	gil48093463 UDP-glucuronate decarboxylase	1e-175
263	3	25	gil18417054 unknown protein	0.0
264	7	34	Não identificado	-
265	25	67	gil8134569 Methionine synthase	0.0
266	1	20	gil20413 alpha-tubulin	0.0
267	2	23	Sem registro	-
268	4	28	gil555974 14-3-3-like protein	e-124

Tabela 10 - Lista de genes diferencialmente expressos entre as bibliotecas de folhas e madeira ($P \leq 0,01$). Para cada *tag* é fornecido o número de cópias na biblioteca de Folhas e na biblioteca de Madeira, a identificação do gene fornecida pelo NCBI bem como o valor de *E* de cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou cDNA de origem

<i>Tag</i>	Nº de Cópias		Identificação	Valor <i>E</i>
	F	M		
269	7	35	Sem registro	-
270	3	26	gil18418552 subtilase	7e-21
271	2	24	gil73665529 4-coumarate:coenzyme A ligase	5e-40
272	11	44	gil22651521 p-coumarate 3-hydroxylase	0.0
273	7	36	gil15223027 protein carrier	2e-54
274	7	36	gil34016875 chitinase-like protein	4e-51
275	3	27	gil37536226 unknown protein	1e-15
276	5	32	gil9755608 phytochelatase synthetase	7e-174
277	1	22	gil19566 glyceraldehyde 3-phosphate dehydrogenase	6e-148
278	2	25	Não identificado	-
279	11	45	gil1621467 laccase	1e-173
280	11	45	gil2244749 hydroxymethyltransferase	0.0
281	19	60	Sem registro	-
282	6	35	gil15234547 pectinesterase	6e-20
283	4	31	gil30683247 unknown protein	2e-47
284	1	24	gil18661 heat shock protein	6e-24
285	8	41	gil47717929 fasciclin-like AGP 13	4e-26
286	2	27	Sem registro	-
287	3	30	gil36963062 Translationally-controlled tumor protein	6e-67
288	1	25	Sem registro	-
290	1	25	gil38194890 Annexin 1	1e-138
291	8	42	gil562779 blue copper protein	1e-39
292	6	38	Não identificado	-
293	5	36	gil3915089 Cinnamic acid 4-hydroxylase	0.0
294	16	59	Sem registro	-
295	7	41	Não identificado	-
296	11	50	gil38228697 superoxide dismutase	2e-71
297	5	37	gil35187687 zinc-finger protein/stress associated protein	6e-51
298	33	92	gil15236989 translation initiation factor	4e-55
299	2	31	gil15221958 polygalacturonase	9e-159
300	9	48	gil78191442 S-adenosyl methionine synthase	6e-84
301	3	34	gil15242651 arabinogalactan protein FLA11	3e-28
302	3	34	Sem registro	-
303	5	39	gil437777 Caffeic acid 3-O-methyltransferase	0.0
304	4	37	gil493725 beta tubulin	5e-61
305	7	45	gil15242073 sucrose synthase SUS1	6e-134
306	1	30	gil50509098 synaptobrevin-like protein	3e-99
307	2	33	gil24745959 lipid transfer protein	4e-19
308	5	41	gil46318067 membrane-anchored endo-1,4-beta-glucanase	0.0
309	4	39	gil18416160 unknown protein	3e-16
310	4	39	gil15236782 protein binding / ubiquitin-protein	1e-16
311	11	56	gil7406663 Plant invertase/ pectin methylesterase	5e-53
312	10	54	Não identificado	-
313	9	52	gil92881534 Ubiquitin/Apoptosis regulator	9e-75
314	2	35	gil53792206 Unknown protein	2e-10
315	2	35	gil60656567 endo-1,4-beta-xylanase	8e-107
316	34	101	gil62321345 glutamate-ammonia ligase/Gluatmine synthase 1	2e-20

Tabela 10 - Lista de genes diferencialmente expressos entre as bibliotecas de folhas e madeira ($P \leq 0,01$). Para cada *tag* é fornecido o número de cópias na biblioteca de Folhas e na biblioteca de Madeira, a identificação do gene fornecida pelo NCBI bem como o valor de *E* de cada identificação. As *tags* “Não identificadas” são aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente mas não há identificação da provável função do gene no NCBI, e as *tags* “Sem registro” representam aquelas para as quais não foi possível encontrar o EST ou cDNA de origem

<i>Tag</i>	Nº de Cópias		Identificação	(conclusão)
	F	M		Valor <i>E</i>
317	4	42	gil47717929 fasciclin-like AGP 13	1e-45
318	4	42	gil5739373 caffeoyl-CoA O-methyltransferase	6e-35
319	1	34	gil28395420 plasma intrinsic protein 2,2	6e-145
320	1	34	gil11071902 laccase	7e-98
321	30	96	gil92880884 Adenosine kinase	2e-113
322	15	70	gil1934859 caffeoyl-CoA O-methyltransferase	3e-138
323	4	45	Não identificado	-
324	8	56	Sem registro	-
325	11	65	Não identificado	-
326	6	54	gil21553523 arabinogalactan protein	2e-42
327	1	41	Não identificado	-
328	7	59	gil71040687 auxin-repressed protein	1e-31
329	8	63	gil67003907 cellulose synthase 1 (EgCesA1)	2e-144
330	7	61	gil32470635 Mob1/phocein family protein	1e-118
331	1	46	gil4586049 hypothetical protein	7e-46
332	19	92	gil85701214 Profilin	2e-65
333	13	80	gil38016521 fiber protein Fb34	9e-71
334	12	78	gil50844677 cytochrome b5	9e-61
335	74	195	gil56311394 ADP-ribosylation factor 1	9e-100
336	11	78	gil29367605 adenosylhomocysteinase	1e-154
337	7	69	Não identificado	-
338	13	88	gil3482933 cdc2 protein kinase	3e-74
339	8	81	Sem registro	-
340	22	122	gil23452315 beta-tubulin 1	0.0
341	11	97	gil21553614 blue copper protein	8e-35
342	19	122	gil2244853 tubulin alpha-6 chain	2e-58
343	15	116	gil84579145 Unknown protein	3e-09
344	21	130	Não identificado	-
345	7	97	Não identificado	-
346	5	93	Não identificado	-
347	13	115	gil92898246 Heavy metal transport/detoxification	1e-47
348	7	104	gil50920439 Putative membrane protein	1e-50
349	7	109	gil11990130 17.9 kDa heat-shock protein	6e-23
350	52	231	gil15236267 PIP3/Canal de transporte de água	2e-148
351	16	193	gil21536535 unknown	9e-17
352	50	309	Sem registro	-
353	22	273	gil30841338 arabinogalactan protein	5e-53
354	88	472	gil10764491 allergenic isoflavone reductase-like	e-145

O crescimento secundário ativo na região cambial requer um grande consumo de energia e consequentemente uma grande quantidade de O₂ livre para geração de energia através da respiração. Entretanto, a quantidade de O₂ livre na região cambial é limitada pela barreira física criada pela casca e pelo próprio consumo de O₂ na respiração, de forma que ao menos parte do

processo pode ser dirigida à produção de energia através da fermentação alcoólica (KIMMERER & STRINGER, 1988).

Em trabalhos recentes, transcritos e proteínas para enzimas relacionadas à respiração anaeróbia têm sido identificadas durante a formação do xilema e crescimento secundário (GION *et al.*, 2005, JUAN *et al.*, 2006, RANIK *et al.*, 2006). Consistentemente com estes resultados, para a maior parte das enzimas do ciclo dos ácidos tricarboxílicos não foi possível verificar a expressão de transcritos, enquanto *tags* para duas enzimas relacionadas à fermentação alcoólica, a álcool desidrogenase (ADH) e a piruvato descarboxilase (PCD), foram observadas nas bibliotecas de folha e de madeira (Figura 20). KIMMERER & STRINGER (1988) demonstraram que ambas as enzimas são ativas na região cambial e nas folhas das espécies arbóreas consideradas. Entretanto, a concentração de etanol e de ADH, foi inversamente proporcional na região cambial das árvores sugerindo que a ADH possa estar atuando na degradação do etanol produzido e que outra enzima estaria atuando na produção de etanol.

Segundo KIMMERER & STRINGER (1988), a possibilidade de a ADH estar relacionada à degradação de etanol na região cambial tem o importante papel de possibilitar a geração de uma quantidade maior de energia a partir do processo fermentativo (como ocorre nas células animais onde o lactato produzido na fermentação é degradado aumentando a geração de ATP), tornando possível atribuir a este processo uma grande importância relativa no tecido cambial em crescimento. Os autores demonstraram também que parte do etanol produzido na região cambial é transportado pelo xilema até as folhas e não é liberada na transpiração, ao contrário disso, a alta atividade estimada da ADH nas folhas estaria relacionada à degradação desse etanol. Assim, os transcritos encontrados para ADH nas bibliotecas de folhas e de madeira de *E. grandis* devem estar similarmente relacionados à degradação de etanol, mas possuem funções distintas no metabolismo da planta, ora atuando na otimização da produção de energia pela respiração, ora atuando na eliminação de produtos do metabolismo.

Para revelar as diferenças entre os perfis de expressão gênica das bibliotecas e caracterizar funcionalmente grupos de genes de expressão tecido-preferencial, os genes foram analisados em cada categoria e subcategoria de acordo com o número de genes identificados, o número de transcritos observados (abundância) e com a relação abundância/genes (Tabela 11). Fica claro através da análise apresentada na Tabela 11 que o número de genes identificados em cada categoria ou subcategoria difere entre as bibliotecas de forma que a correção abundância/genes

fornece o melhor ponto de partida para comparação dos perfis transcricionais das bibliotecas. Como já esperado, as bibliotecas apresentaram valores muito próximos para a categoria Metabolismo e Energia, mas enquanto na biblioteca de folhas o maior número de transcritos esteve relacionado ao metabolismo autotrófico e assimilação de carbono (fotossíntese), na biblioteca de madeira a maior relação abundância/genes foi observada para a respiração, metabolismo de carboidratos, nitrogênio e metabolismo secundário.

Foi interessante notar que, ao contrário do esperado, um grande número de transcritos para a subunidade menor da rubisco foi observado na biblioteca de madeira (62 transcritos, Tabela 8). Esse número foi inclusive maior que o número de transcritos observados para a maioria das celulose sintases e outras enzimas importantes na síntese da parede celular. Além disso, um grande número de transcritos de genes nucleares que codificam proteínas dos fotossistemas foi também observado na biblioteca de madeira (Tabela 6 e 8).

Proteínas do complexo coletor de luz e rubisco vêm sendo identificadas por outros pesquisadores que estudam a formação do xilema, mas nenhuma hipótese para explicar este fato foi apresentada até o momento (JUAN *et al.*, 2006). Em um trabalho de tese conduzido no mesmo laboratório onde foram produzidos perfis proteicos para amostras cambiais de *E. grandis*, foram também identificados *spots* para essas mesmas proteínas (ANDRADE, 2006). A presença do transcrito não necessariamente está correlacionada à presença da proteína e à sua funcionalidade na célula, assim, outras evidências da funcionalidade dessas proteínas na região cambial precisam ser buscadas para propor novas funções a essas proteínas e/ou elucidar os mecanismos de regulação da expressão desses genes nos diferentes tecidos. A rubisco, por exemplo, provavelmente não tenha sua função catalítica ativa nas células da região cambial já que somente 3 transcritos foram encontrados para a rubisco ativase na biblioteca de madeira. Além disso, a *tag* para RCA da biblioteca de madeira é encontrada somente no sentido complementar invertido em ESTs anti-sense de *Eucalyptus* o que sugere a possibilidade de regulação anti-sense do gene RCA nas células da região cambial, embora esta hipótese deva ainda ser confirmada.

Tabela 11 - Distribuição dos genes identificados nas bibliotecas de Madeira e Folhas em categorias e subcategorias funcionais. Para cada categoria/subcategoria são fornecidos: o número de genes identificados, o número total de transcritos e a relação abundância/genês; onde o número total de transcritos é corrigido pelo número de genes identificados. Biblioteca de Madeira (M), biblioteca de Folhas (F)

Categorias Funcionais	Genes		Abundância		Ab/Genes	
	M	F	M	F	M	F
1 Metabolismo e Energia	250	207	3564	3385	14,2	16,3
Metabolismo autotófico/energia	106	100	1221	1920	11,5	19,2
Metabolismo de carbono	3	5	69	348	23	69,6
Metabolismo de carboidratos	57	43	706	355	12,4	8,25
Metabolismo de nitrogênio	8	6	165	83	20,6	13,8
Metabolismo de aminoácidos	30	20	455	235	15,2	11,7
Metabolismo de lipídeos	23	12	213	100	9,3	8,3
Metabolismo de pequenas moléculas	12	7	146	79	12,2	11,3
Metabolismo secundário	11	14	589	265	53,5	18,9
2 Processos Celulares	179	136	2038	2205	11,4	16,2
Divisão celular	19	12	311	84	16,4	7
Tradução de sinais	50	26	429	135	8,6	5,2
Regulação celular	27	19	248	147	9,2	7,7
Respostas de proteção/detoxificação	46	54	663	1546	14,4	28,6
Respostas a estímulos	37	25	387	293	10,5	11,7
3 Transporte	88	57	1277	473	14,5	8,3
Transporte de grandes moléculas	13	12	99	62	7,6	5,2
Transporte de pequenas moléculas	25	13	434	98	17,4	7,5
Transporte de vesículas	43	28	670	280	15,6	10
Kinesinas/fatores transportadores	7	4	74	33	10,6	8,2
4 Estrutura e Organização da Estrutura Celular	221	165	3563	1326	16,1	8
Membrana celular	12	11	120	87	10	7,9
Parede celular	79	38	1779	253	22,5	6,6
Citoesqueleto	32	13	732	145	22,9	11,1
RNAr e proteínas ribossomais	91	97	897	805	9,8	8,3
Organização dos cromossomos	7	6	35	36	5	6
5 Vias de Informação	189	146	2121	1332	11,2	9,1
Síntese/degradação/modificação de DNA	11	10	72	101	6,5	10,1
Síntese/degradação/modificação de RNA	44	35	371	279	8,4	8
Síntese/degradação/modificação de proteínas	134	101	1678	952	12,5	9,4
6 Sem Função definida	345	264	-	-	-	-
Total	1272	975	-	-	-	-

A observação de que transcritos para enzimas relacionadas ao metabolismo de nitrogênio estão relativamente mais representados na biblioteca de madeira que na de folhas foi um resultado, a princípio, inesperado. Isso porque a rubisco é uma das proteínas mais abundantes, principalmente nas folhas, representando a maior fonte de reserva de N assimilado que pode ser reciclado durante o desenvolvimento para uso na construção de compostos proteicos nas próprias folhas ou em outros órgãos (IRVING & ROBINSON, 2006). A maior representatividade dos transcritos relacionados ao metabolismo de nitrogênio na biblioteca de madeira, entretanto, pôde ser atribuída quase que exclusivamente à elevada expressão de transcritos para a glutamina sintetase citosólica (GS1) (Tabelas 8 e 11). Assim, o metabolismo de nitrogênio na formação da

madeira está relacionado à reciclagem de N direcionada à síntese de fenilalanina e manutenção das altas taxas de lignificação que são observadas durante o crescimento secundário (KEYS, 2006; CÁNTON *et al.*, 2005).

Outros importantes resultados foram a baixíssima representatividade dos transcritos relacionados à síntese de amido na biblioteca de folhas, a ausência de transcritos para enzimas envolvidas na degradação de amido na biblioteca de madeira, e o grande número de transcritos identificados para galactinol sintetase na biblioteca de folhas. A ausência de transcritos relacionados à síntese e degradação de amido na região cambial em crescimento ativo, e baixa expressão de transcritos associados à síntese de amido em folhas, têm sido também reportada para *Populus*, indicando, portanto, que outro composto de reserva esta presente em *Populus*, ou ainda que a sacarose produzida pela fotossíntese é transportada até as células da região cambial onde é utilizada como fonte de precursores para a biossíntese de parede celular (GEISLER-LEE *et al.*, 2006). Por outro lado, o açúcar rafinose representa um importante produto da fotossíntese em algumas plantas funcionando tanto no transporte através do floema quanto como composto de reserva de energia (SPRENGER & KELLER, 2000, NORWOOD *et al.*, 2000, KOCHHAR *et al.*, 2003). A enzima galactinol sintetase, que catalisa o primeiro passo na formação de rafinose, foi representada por um grande número de transcritos preferencialmente expressos na biblioteca de folhas (Tabela 10), o que é um indicativo de que a rafinose pode estar sendo produzida em altas taxas pela fotossíntese nas folhas. Outras evidências, entretanto, são necessárias para atribuir a rafinose algum papel no armazenamento e/ou transporte de açúcares bem como na síntese dos polissacarídeos de parede celular em *E. grandis*.

Assim como esperado, os transcritos relacionados a genes envolvidos no metabolismo de carboidratos, transporte de vesículas, síntese e organização da parede celular, e organização do citoesqueleto celular, foram identificados em grande abundância na biblioteca de madeira (Tabelas 8 e 11). As enzimas ativas no metabolismo de carboidratos (CAZymes) inclui famílias de glicosiltransferases, glicosil hidrolases, polissacarídeo liases e esterases, que são importantes na síntese de açúcares e amido durante a fotossíntese, mas que são especialmente importantes na síntese dos polímeros de celulose, pectina, xilanas e compostos secundários na região cambial (HENRISSAT, 1991; COUTINHO *et al.*, 2003; GEISLER-LEE *et al.*, 2006). Entre estas enzimas, foram identificados transcritos para CAZymes cuja função já é bem conhecida, como por exemplo celulose sintases, mas foram também identificados transcritos para enzimas de função

ainda não conhecida cuja expressão têm sido relacionada a madeira em formação, tais como membros da família de glicosil transferases GT8, GT43 e GT47, e membros da família de glicosil hidrolases GH1, GH2, GH42, GH55, GH 64 e GH81.

Enquanto na biblioteca de madeira a síntese de celulose parece estar sendo principalmente conduzida por subunidades de *EgCesA1* a qual está relacionada à síntese de parede secundária, na biblioteca de folhas *EgCesA1* e *EgCesA4* apresentaram níveis de expressão apenas pouco maiores que os encontrados para *EgCesA2* e *EgCesA3*. Além disso, para *EgCesA5* não foi observado nenhum transcrito na biblioteca de folhas. Os transcritos encontrados na biblioteca de folhas para KORRIGAN e sacarose sintase também apresentaram baixos níveis de expressão acompanhando os transcritos para as celulose sintases. Segundo RANIK & MYBURG (2006) os genes *EgCesA1*, *EgCesA2* e *EgCesA3* são muito menos expressos em folhas do que no xilema, o que coincide com os resultados obtidos neste trabalho, mas *EgCesA4* e *EgCesA5* segundo os autores, são duas vezes mais expressos em folhas do que no xilema, contrastando com os nossos resultados.

Foram observadas *tags* para quase a totalidade das enzimas relacionadas à síntese de lignina na biblioteca de folhas. No entanto, estas *tags* apresentam níveis similarmente baixos de expressão. Somente para a enzima SAMS foi observado um número maior de transcritos (23 cópias), o que pode estar relacionado à função desta enzima em diversos outros processos celulares (HANSON & ROJE, 2001). Uma observação interessante foi a presença de uma *tag* referente a uma UDP-glicose glicosil transferase da família GT44, identificada como preferencialmente expressa na biblioteca de folhas (Tabela 10). Esta enzima foi relacionada à síntese de lignina em frutos de morango atuando na formação de cinamoil glicose e p-comaroil glicose (LUNKENBEIN *et al.*, 2006). Caso esta enzima apresente a mesma função nas folhas de *E. grandis*, sua elevada expressão é um indicativo de que enzimas diferentes podem estar relacionadas a processos bem conhecidos da síntese de lignina em outros tecidos.

O grande número de transcritos relacionados ao transporte de vesículas e ao citoesqueleto na biblioteca de madeira (Tabela 11) está estritamente relacionado à síntese e organização dos polímeros de parede celular. Os polissacarídeos da parede celular são sintetizados no Golgi a partir de açúcares-NDP e são exportados para a matrix celular em vesículas derivadas do Golgi (DELMER & STONE, 1988). Além disso, os complexos proteicos de membrana envolvidos na síntese de celulose são também incorporados à membrana celular a

partir da fusão de vesículas (ODA *et al.*, 2005). O transporte de vesículas no interior das células é direcionado por microtúbulos do citoesqueleto os quais determinam também a orientação da deposição das microfibrilas de celulose na parede celular (ODA *et al.*, 2005).

Foi importante notar que a subcategoria “degradação de proteínas” foi similarmente representativa nas bibliotecas de folhas e de madeira somando 338 e 366 transcritos, respectivamente. Conforme já mencionado na seção anterior, os transcritos relacionados à degradação de proteínas na biblioteca de madeira representaram membros da via Ub/proteassomo 26S de degradação que estão relacionados à sinalização hormonal e diferenciação do xilema. Na biblioteca de folhas, entretanto, dos 338 transcritos, 187 estiveram relacionados a esta mesma via, mas 151 representam ATP-dependentes proteases tais como metalopeptidases, serina, cisteína e aspartato proteases. Os plastídeos não possuem o complexo proteassomo 26S, ao contrário disso, eles apresentam proteases procarióticas análogas ao sistema Ub/proteassomo 26S de degradação denominadas conjuntamente de ATP-dependente proteases. Nos plastídeos, as proteases assumem importantes funções como a degradação dos peptídeos de sinalização das proteínas importadas do citoplasma (catalisada pelas metalopeptidases), maturação de proteínas dos complexos fotossintéticos (catalisada pelas serina proteases), *turnover* constante de proteínas regulatórias permitindo rápida aclimação a diferentes condições, e também degradação dos complexos fotossintéticos danificados pela fotoinibição (ATP-dependente proteases Clp) (SAKAMOTO, 2006).

De um modo geral, a comparação dos perfis transcricionais de folhas e madeira forneceu uma visão mais global do metabolismo das árvores em crescimento possibilitando, por exemplo, a identificação de processos que são importantes em folhas e madeira e que representam possíveis candidatos para modificação. Como se sabe, cada destino final no uso da madeira requer determinadas características anatômicas, químicas e físicas que conjuntamente determinam a qualidade da madeira para um determinado uso. No caso da indústria de papel e celulose, a qualidade da madeira é influenciada profundamente pela quantidade de lignina que representa um grande obstáculo à produção eficiente de polpa e papel uma vez que deve ser removida da massa de celulose em um processo que requer uma grande quantidade de energia química. A transgenia tem sido muito utilizada para buscar reduções no conteúdo de lignina nos plantios comerciais através principalmente do silenciamento ou supressão de genes diretamente

relacionados à síntese de lignina como 4CL (HU *et al.*, 1999), COMT e CAD (PILATE *et al.*, 2002) e muitos outros (Revisado em CAMPBELL *et al.*, 2003).

Alternativamente, alterações no conteúdo final de lignina poderiam também ser alcançadas reduzindo a disponibilidade de precursores da síntese dos monolignóis. De acordo com os resultados obtidos no presente trabalho, há uma forte evidência que tanto a GS1 quanto a SAMS são fundamentais na reciclagem de N e grupos metil, os quais são consumidos em altas taxas durante a lignificação, representando, portanto, importantes candidatos para transgenia em *E. grandis*. Por outro lado, a produtividade dos plantios comerciais depende num primeiro momento da biomassa investida em crescimento secundário. Assim, a otimização da fotossíntese nas folhas deve também ser um objetivo buscado no aumento da produtividade das florestas plantadas. Em um trabalho recente, por exemplo, foi demonstrado que híbridos mais produtivos de arroz apresentavam um menor nível de expressão para genes relacionados à fotorrespiração quando comparados aos seus parentais menos produtivos (BAO *et al.*, 2005). Na biblioteca de folhas de *E. grandis* a glicolato oxidase, importante enzima relacionada à fotorrespiração, foi identificada por um grande número de transcritos (Tabela 5). Caso a comparação entre o nível de expressão dessa enzima entre plantios (clones) de diferentes produtividades sugira uma correlação entre a fotorrespiração e produtividade, alterações no nível de expressão da glicolato oxidase podem representar um importante alvo alternativo ao melhoramento do eucalipto.

As diferenças observadas nos níveis de expressão dos genes entre as bibliotecas de madeira e de folhas são também úteis na identificação de genes de expressão tecido-específica e promotores tecido-específicos a serem utilizados em eventos de transgenia. Neste trabalho, um total de 111 genes foi exclusivamente expresso nas bibliotecas, sendo 47 exclusivos da biblioteca de madeira e 64 da biblioteca de folhas. Neste momento já vem sendo realizado no laboratório um estudo acerca do potencial de desenvolvimento de promotores tecido-específicos a partir destes genes.

6 CONSIDERAÇÕES FINAIS

Considerando-se a atual busca da genômica funcional pelo esclarecimento dos mecanismos regulatórios da expressão gênica e elucidação das funções de genes em diferentes tecidos e/ou situações, uma das principais contribuições deste trabalho foi demonstrar o grande potencial da técnica SAGE em gerar perfis transcricionais e caracterizar funcionalmente diferentes tecidos e processos celulares. Além de possibilitar o acesso a transcritos de baixa abundância nas células, as SAGE *tags* de 14pb foram suficientemente específicas em identificar genes únicos, mesmo quando ESTs originados de diferentes tecidos foram utilizados na identificação dos genes. No entanto, o *tag mapping* através do uso de bancos de dados de referência produzidos a partir dos mesmos tipos de tecidos utilizados na construção das bibliotecas possibilitou a diferenciação de isoformas tecido-específica ou de localização celular específica, como mostrado em algumas situações. Dois importantes obstáculos podem limitar os dados produzidos através do SAGE: (i) a ausência de cauda poli(A) em alguns transcritos de genes nucleares e nos transcritos codificados por genes organelares, (ii) a limitação da identificação direta dos genes expressos (38%) pela indisponibilidade de sequências públicas para o organismo em questão. A última limitação pode ser bem contornada em espécies onde o genoma é disponibilizado em bancos públicos e, felizmente no caso do eucalipto, há a previsão de em breve ser disponibilizado o sequenciamento do genoma de espécies de *Eucalyptus* (*E. globulus*, *E. camaldulensis* e *E. grandis*) pelo consórcio internacional IEuGC.

Através do uso do SAGE foi possível ainda a construção de um banco de dados local de transcritos contendo 975 *tags* identificadas de folhas e 1272 *tags* identificadas de madeira de *E. grandis*. Este banco reúne informações das sequências de ESTs de provável origem de cada *tag*, sua identificação e *score*, ocorrência de domínios funcionais, categoria funcional, e nível de expressão, em fichas eletrônicas criadas para cada *tag*. São disponibilizadas no banco ferramentas de busca que permitem ao consultor acessar a informação de interesse por tecido a partir de diferentes campos de busca como, por exemplo, uma categoria funcional específica. O banco permite também o processamento rápido de informações para estabelecimento de perfis transcricionais como, por exemplo, o número de genes ou de transcritos encontrados em cada categoria funcional em um determinado tecido.

A análise dos transcritos identificados na biblioteca de folhas revelou que a maior parte dos transcritos (39%) esteve relacionada à categoria Metabolismo e Energia. Entre estes foram

representativos transcritos para proteínas dos complexos fotossintéticos PSI, PSII, citocromo *b₆f* e ATP sintase, para enzimas relacionadas ao ciclo de Benson-Calvin e para enzimas relacionadas à fotorrespiração. Foi também observados na biblioteca de folhas um grande número de transcritos para enzimas envolvidas na destoxificação de espécies reativas de oxigênio (ROS) que são principalmente produzidas durante a fotossíntese e a fotorrespiração. Provavelmente, as principais enzimas que atuam prevenindo os danos causados pelo estresse oxidativo nas folhas de *E. grandis* são glutathione S-transferase/peroxidase e catalases. Foi importante observar ainda que o grande número de transcritos para a galactinol sintase na biblioteca de folhas é um indicativo de que a rafinose é um dos açúcares produzidos em quantidade pela fotossíntese em *E. grandis*. Alternativamente, o baixo número de transcritos para enzimas relacionadas à síntese de amido nas folhas sugere ainda que o armazenamento de energia na forma de amido parece não ser favorecido nas folhas de *E. grandis*.

A biblioteca de Madeira foi caracterizada por um número similar de transcritos relacionados à categoria Metabolismo e Energia (29%) e à categoria Estrutura e Organização da Estrutura Celular (28%). A produção de energia para o crescimento secundário nas células cambiais esteve principalmente relacionada à glicólise e fermentação alcoólica a partir da sacarose produzida pela fotossíntese nas folhas. A clivagem da sacarose pela SUSY nas células cambiais foi também demonstrada ser a principal fonte de UDP-glicose para a síntese dos polissacarídeos da parede celular em *E. grandis*. O metabolismo nas células cambiais foi principalmente representado por transcritos relacionados aos processos de respiração, síntese de aminoácidos e reciclagem de nitrogênio e ao metabolismo secundário. A alta expressão dos transcritos para glutamina sintetase na biblioteca de madeira é um indicativo da importância desta enzima na reciclagem local do N liberado na desaminação da fenilalanina e consequente suporte ao alto investimento em lignificação na madeira em formação de *E. grandis*.

A categoria Estrutura e organização da Estrutura Celular foi principalmente representada na biblioteca de Madeira, por transcritos relacionados à síntese da parede celular, principalmente à síntese de lignina, e por proteínas do citoesqueleto celular. Entre os genes mais expressos estiveram *EgCesA1*, *COMT*, *CCoAOMT*, *SAMS* e lacases evidenciando, portanto, um grande investimento em crescimento secundário. Uma importante informação foi a de que os níveis de expressão dos genes *COMT*, *CCoAOMT* e *SAMS* foram bastante similares na biblioteca de madeira sugerindo que possivelmente a *SAMS* participe também do *cluster* de co-regulação

transcricional formado por COMT e CCoAOMT. Embora nenhuma evidência direta do papel da SAMS na lignificação em *E. grandis* tenha sido alcançada no presente trabalho, os resultados sugerem fortemente um papel importante para esta enzima durante o crescimento secundário em *E. grandis*, provavelmente, funcionando em um mecanismo local de reciclagem de grupos metil consumidos na lignificação. Foi também observado na biblioteca de madeira um grande número de transcritos similares á *dirigent proteins*, o que sugere que possivelmente estas enzimas sejam importantes na polimerização de monolignóis em *E. grandis*.

A expressão de SAMS foi também observada através de um número relativamente grande de transcritos na biblioteca de folhas. Entretanto, outros genes relacionados à síntese de parede celular foram pouco representativos nas folhas indicando que neste tecido, a SAMS seja principalmente importante em outros processos celulares como uma enzima de função *housekeeping* na síntese de metionina. Foi identificado como preferencialmente expressos na biblioteca de folhas transcritos para uma UDP-glicose glicosiltransferase, relacionada à síntese de monômeros de lignina em morango. Embora seja necessário elucidar a função desta enzima nos tecidos foliares, sua expressão em *E. grandis* pode ser um importante exemplo da ocorrência de enzimas alternativas tecido-específicas em vias importantes já bem caracterizadas em outros tecidos.

O uso do método SAGE possibilitou a identificação de genes exclusivamente expressos em folhas e em madeira revelando uma lista de genes potenciais para busca de promotores tecido-específicos. Além disso, a comparação dos perfis transcricionais de folhas e madeira foi útil em permitir a identificação de processos candidatos para modificação em programas de melhoramento da qualidade da madeira de eucalipto tais como redução da fotorrespiração, bloqueando ou minimizando a expressão da glicolato oxidase, e redução do conteúdo de lignina, reduzindo-se os processos de reciclagem de N e grupos metil localmente catalisados pela glutamina sintetase e S-adenosil metionina sintetase. Além disso, enzimas diretamente relacionadas à síntese de celulose tais como a sacarose sintase, endo-1,4- β -glucanase (principalmente o gene similar à *EG2*), e celulose sintase (principalmente *EgCesA1*), representam também genes alvo potenciais para modificação da síntese de celulose como sugerido pelo alto nível de expressão encontrado para essas enzimas.

REFERÊNCIAS

- ADAMS, M.D.; KELLEY, J.M.; GOCAYNE, J.D.; DUBNICK, M.; POLYMEROPOULOS, M.H.; XIAO, H.; MERRIL, C.R.; WU, A.; OLDE, B.; MORENO, R.F.; KERLAVAGE, A.K.; McCOMBIE, W.R.; VENTER, J.C. Complementary DNA sequencing: expressed sequence tags and human genome project. **Science**, Washington, v. 252, p. 1651–1656, 1991.
- AHARONI, A., VORST, O. DNA microarrays for functional plant genomics. **Plant Molecular Biology**, New York, v. 48, p. 99–118, 2001.
- AKAZAWA, T.; TETSUCO, T.; KOBAYASHI, H. Molecular evolution of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). **Trends in biochemical sciences**, Amsterdam, v. 9, p. 380–384, 1984.
- ALBA, R.; FEI, Z.; PAYTON, P.; LIU, Y.; MOORE, S.L.; DEBBIE, P.; COHN, J.; D'ASCENZO, M.; GORDON, J.S.; ROSE, J.K.C.; MARTIN, G.; TANKSLEY, S.D.; BOUZAYEN, M.; JAHN, M.M.; GIOVANNONI, G. ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. **The Plant Journal**, Oxford, v. 39, n. 5, p. 697-714, 2004.
- ALLEN, J. F., FORSBERG, J. Molecular recognition in thylakoid structure and function. **Trends in Plant Science**, Kidlington, v. 6, n. 7, p. 317-326, 2001.
- ALLINA, S.M.; PRI-HADASH, A.; THEILMANN, D. A.; ELLIS, B. E.; DOUGLAS, C. J. 4-Coumarate:coenzyme A ligase in hybrid poplar. Properties of native enzymes, cDNA cloning, and analysis of recombinant enzymes. **Plant Physiology**, Lancaster, v. 116, n. 2, p. 743-754, 1998.
- ALLONA, I.; QUINN, M.; SHOOP, E.; SWOPE, K.; CYR, S. S.; CARLIS, J.; RIEDL, J.; RETZEL, E.; CAMPBELL, M. M.; SEDEROFF, R.; WHETTEN, R. W. Analysis of xylem formation in pine by cDNA sequencing. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 95, p. 9693-9698, 1998.
- ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFFER, A. A.; ZHANG, J. H.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, London, v. 25, n. 17, p. 3389-3402, 1997.
- AMOR, Y.; HAIGLER, C. H.; JOHNSON, S.; WAINSCOTT, M.; DELMER, D. P. A membrane-associated form of sucrose 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, p. 9353-9357, 1995.
- ANDERSSON-GUNNERÅS, S.; MELLEROWICZ, E. J.; LOVE, J.; SEGERMAN, B.; OHMIYA, Y.; COUTINHO, P. M.; NILSSON, P.; HENRISSAT, B.; MORITZ, T.; SUNDBERG, B. Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. **The Plant Journal**, Oxford, v. 45, p. 144-165, 2006.

ANDRADE, A. **Sequenciamento, identificação e análise de proteínas do caule de mudas de *Eucalyptus grandis***. 2006. 120p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Querioz”, Universidade de São Paulo, Piracicaba, 2006.

ANDRADE, E. N. **O Eucalipto**. 2.ed. Jundiaí: Companhia Paulista de Estradas de Ferro, 1961. 667 p.

ASADA, K. Production and scavenging of reactive oxygen species in chloroplasts and their functions. **Plant Physiology**, Lancaster, v. 141, n. 2, p. 391-396, June, 2006.

BACHEM, C.; VAN DER HOEVEN, R.; DE BRUIJN, S.; VREUGDENHIL, D.; ZABEAU, M.; VISSER, R. Visualisation of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. **The Plant Journal**, Oxford, v. 9, p. 745–753, 1996.

BAO, J.; LEE, S.; CHEN, C.; ZHANG, X.; ZHANG, Y.; LIU, S.; CLARK, T.; WANG, J.; CAO, M.; YANG, H.; WANG, S.M.; YU, J. Serial analysis of gene expression study of a hybrid rice strain (LYP9) and its parental cultivars. **Plant Physiology**, Lancaster, v.138, p. 1216-1231, 2005.

BAUCHER, M.; HALPIN, C.; PETIT-CONIL, M.; BOERJAN, W. Lignin: Genetic engineering and impact on pulping. **Critical reviews in biochemistry and molecular biology**, Boca Raton, v. 38, p. 305-350, 2003.

BAUWE, H.; KOLUKISAOGLU, U. Genetic manipulation of glycine decarboxylation. **Journal of Experimental Botany**, Lancaster, v. 54, n. 387, p. 1523-1535, June, 2003.

BERLETH, T., SACHS, T. Plant morphogenesis: long-distance coordination and local patterning. **Current Opinion in Plant Biology**, London, v. 4, n. 1, p. 57-62, 2001.

BHALLA, R.; NARASIMHAN, K.; SWARUP, S. Metabolomics and its role in understanding cellular responses in plants. **Plant Cell Reports**, Berlin, v. 24, n. 10, p. 562-71, 2005.

BHALERAO, R.; KESKITALO, J.; STERKY, F.; ERLANDSSON, R.; BJORKBACKA, H.; BIRVE, S.J.; KARLSSON, J.; GARDESTROM, P.; GUSTAFSSON, P.; LUNDEBERG, J.; JANSSON, S. Gene expression in autumn leaves. **Plant Physiology**, Lancaster, v.131, n. 2, p. 430-442, 2003.

BOURQUIN, V.; NISHIKUBO, N.; ABE, H.; BRUMER, H.; DENMAN, S.; EKLUND, M.; CHRISTIERNIN, M.; TEERI, T. T.; SUNDBERG, B.; MELLEROWICZ, E. J. Xyloglucan endotransglycosylases have a function during the formation of secondary cell walls of vascular tissues. **Plant Cell**, Rockville, v. 14, p. 3073-3088, 2002.

BRENNER, S.; JOHNSON, M.; BRIDGHAM, J.; GOLDA, G.; LLOYD, D.H.; JOHNSON, D. LUO, S.; McCURDY, S.; FOY, M.; EWAN, M.; ROTH, R.; GEORGE, D.; ELETR, S.; ALBRECHT, G.; VERMAAS, E.; WILLIAMS, S.R.; MOON, K.; BURCHAM, T.; PALLAS, M.; DUBRIDGE, R.B.; KIRCHNER, J.; FEARON, K.; MAO, J.; CORCORAN, K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. **Nature Biotechnology**, New York, v.18, n. 6, p. 630-4, 2000.

BRONDANI, R. P. V.; BRONDANI, C.; TARCHINI, R.; GRATTAPAGLIA, D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical and Applied Genetics**, Berlin, v. 97, p. 816-827, 1998.

BRONDANI, R. P.; WILLIAMS, E. R.; BRONDANI, C.; GRATTAPAGLIA, D. A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. **BMC Plant Biology**, London, v. 22, p. 6-20, 2006.

BURLAT, V.; KWON, M.; DAVIN, L. B.; LEWIS, N. G. Dirigent proteins and dirigent sites in lignifying tissues. **Phytochemistry**, Oxford, v. 57, p. 883-897, 2001.

BYRNE, M.; MURRELL, J. C.; ALLEN, B.; MORAN, G. F. An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. **Theoretical and Applied Genetics**, Berlin, v. 91, p. 869-875, 1995.

CAMPBELL, M. M.; BRUNNER, A. M.; JONES, H. M.; STRAUSS, S. H. Forestry's fertile crescent: the application of biotechnology to forest trees. **Plant Biotechnology Journal**, Oxford, v. 1, p. 141-154, 2003.

CANTÓN, F. R; SUÁREZ, M. F.; CÁNOVAS, F. M. Molecular aspects of nitrogen mobilization and recycling in trees. **Photosynthesis Research**, Hague, v. 83, p. 265-278, 2005.

CARPITA, N. C.; GIBEAUT, D. M. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. **The Plant Journal**, Oxford, v.3, p. 1-30, 1993.

CHAFFEY, N. Wood formation in forest trees: from Arabidopsis to Zinnia. **Trends in Plant Science**, Kidlington, v. 4, n. 6, p. 203-204, June, 1999.

CHAFFEY, N.; CHOLEWA, E.; REGAN, S.; SUNDBERG, B. Secondary xylem development in Arabidopsis: a model for wood formation. **Physiologia Plantarum**, Lund, v. 114, n. 4, p. 594-600, 2002.

CHEN, S., HARMON, A.C. Advances in plant proteomics. **Proteomics**, Weinheim, v. 6, n. 20, p. 5504-16, 2006.

CHEN, J.; SUN, M.; LEE, S.; ZHOU, G.; ROWLEY, J.; WANG, S. Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 99, p. 12257-12262, 2002.

COEMANS, B.; MATSUMURA, H.; TERAUCHI, R.; REMY, S.; SWENNEN, R.; SAGI, L. SuperSAGE combined with PCR walking allows global gene expression profiling of banana (*Musa acuminata*), a non-model organism. **Theoretical and Applied Genetics**, Berlin, v. 111, n. 6, p. 1118-26, 2005.

COUTINHO, P. M.; DELEURY, E.; DAVIES, G. J.; HENRISSAT, B. An evolving hierarchical family classification for glycosyltransferases. **Journal of Molecular Biology**, New York, v. 328, n. 2, p. 307-327, 2003.

CREN, M.; HIREL, B. Glutamine synthetase in higher plants. Regulation of gene and protein expression from the organ to the cell. **Plant Cell Physiology**, Kyoto, v. 40, p. 1187-1193, 1999.

DALESSANDRO, G.; NORTHCOTE, D. H. Possible control sites of polysaccharide synthesis during cell growth and wall expansion of pea seedlings (*Pisum sativum* L.). **Planta**, Berlin, v. 134, p. 39-44, 1977.

DARLEY, C. P.; FORRESTER, A. M.; MACQUEN-MASON, J. The molecular basis of plant cell wall extension. **Plant Molecular Biology**, New York, v. 47, p. 179-195, 2001.

DE JONGE, R. R.; VREIJLING, J. P.; MEINTJES, A.; KWA, M. S.; VAN KAMPEN, A. H.; SCHAIK, I. N. van; BAAS, F. Transcriptional profile of the human peripheral nervous system by serial analysis of gene expression. **Genomics**, San Diego, v. 82, n. 2, p. 97-108, 2003.

DELMER, D. P.; STONE, B. A. Biosynthesis of plant cell walls. In: STUMPF, P. K.; CONN E. E. (Eds). **The biochemistry of plants: a comprehensive treatise**. New York: Academic Press, 1988. chap. 14, p. 373-420.

DEMURA, T; TASHIRO, G.; HORIGUCHI, G.; KISHIMOTO, N.; KUBO, M.; MATSUOKA, N.; MINAMI, A.; NAGATA-HIWATASHI, M.; NAKAMURA, K.; OKAMURA, Y.; SASSA, N.; SUZUKI, S.; YAZAKI, J.; KIKUCHI, S.; FUKUDA, H. Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 93, p. 12637-12642, 2002.

DIATCHENKO, L.; LAU, Y.F.; CAMPBELL, A.P.; CHENCHIK, A.; MOQADAM, F.; HUANG, B.; LUKYANOV, S.; LUKYANOV, K.; GURSKAYA, N.; SVERDLOV, E.D.; SIEBERT, P.D. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 93, n. 12, p. 6025-30, 1996.

DIXON, R. A. Natural products and plant disease resistance. **Nature**, London , v. 411, p. 843-847, 2001.

DOBLIN, M. S.; KUREK, I.; JACOB-WILK, D.; DELMER, D. P. Cellulose biosynthesis in plants: from genes to rosettes. **Plant Cell Physiology**, Kyoto, v. 43, n. 12, p.1407-1420, 2002.

DOWNIE, B.; GURUSINGHE, S.; DAHAL, P.; THACKER, R.R.; SNYDER, J.C.; NONOGAKI, H.; YIM, K.; FUKANAGA, K.; ALVARADO, V.; BRADFORD, K.J. Expression of a galactinol synthase gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. **Plant Physiology**, Lancaster, v. 131, n. 3, p. 1347-59, 2003.

DRMANAC, R., DRMANAC, S. cDNA screening by array hybridization. **Methods in Enzymology**, New York, v. 303, p. 165–178, 1999.

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.

FERREIRA, M. E., GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 2 ed. Brasília: EMBRAPA-CENARGEN, 1995. p. 220.

FERREIRA, M. Melhoramento florestal e silvicultura intensiva com eucalipto. **Silvicultura**, São Paulo, v. 31, p. 5-11, 1983.

FERREIRA, S. **Produção e qualidade da madeira serrada de clones de híbridos de *Eucalyptus ssp.*** 2003. 71 p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal de Lavras, Lavras, 2003.

FEY, V.; WAGNER, R.; BRAUTIGAM, K.; PFANNSCHMIDT, T. Photosynthetic redox control of nuclear gene expression. **Journal of Experimental Botany**, Lancaster, v. 56, n. 416, p. 1491-1498, 2005.

FILLATTI, J.J.; SELLMER, J.; McCOWN, B.; HAISSIG, B.; COMAI, L. *Agrobacterium*-mediated transformation and regeneration of *Populus*. **Molecular & Genreal Genetics**, Berlin, v. 206, p. 192-199, 1987.

FIZAMES, C.; MUÑOS, S.; CAZETTES, C.; NACRY, P.; BOUCHEREZ, B.; GAYMARD, F.; PIQUEMAL, D.; DELORME, V.; COMMES, T.; DOUMAS, P.; COOKE, R.; MARTI, J.; SENTENAC, H.; GOJON, A. The Arabidopsis root transcriptome by Serial Analysis of Gene Expression. Gene identification using the genome sequence. **Plant Physiology**, London, v. 134, p. 67-80, 2004.

FLEMING, A. Metabolic aspects of organogenesis in the shoot apical meristem. **Journal of Experimental Botany**, Oxford, v. 57, n. 9, p. 1863-1870, 2006.

FOOD AND AGRICULTURE ORGANIZATION. **Global Forest resources assessment 2000 – main report**. FAO Forestry paper. Disponível em < <http://www.fao.org/forestry/fo/fra/main/index.jsp> > Acesso em 31 out. 2006.

FOUCART, C.; PAUX, E.; LADOUCE, N.; SAN-CLEMENTE, H.; GRIMA-PETTENATI, J.; SIVADON, P. Transcript profiling of a xylem vs phloem cDNA subtractive library identifies new genes expressed during xylogenesis in *Eucalyptus*. **New Phytologist**, London, v. 170, n. 4, p. 739-752, 2006.

FUKAO, Y.; HAYASHI, M.; NISHIMURA, M. Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of *Arabidopsis thaliana*. **Plant Cell Physiology**, Kyoto, v. 43, n. 7, p. 689-696, 2002.

FUKUDA, H. Xylogenesis: initiation, progresion, and cell death. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 47, p. 299-325, 1996.

GE, X.; WU, Q.; JUNG, Y.C.; CHEN, J.; WANG, S.M. A large quantity of novel human antisense transcripts detected by LongSAGE. **Bioinformatics**, Oxford, v. 22, n. 20, p. 2475-2479, 2006.

GEIGENBERGER, P.; KOLBE, A.; TIESSEN, A. Redox regulation of carbon storage and partitioning in response to light and sugars. **Journal of Experimental Botany**, Lancaster, v. 56, n. 416, p. 14691-1479, 2005.

GEISLER-LEE, J.; GEISLER, M.; COUTINHO, P. M.; SEGERMAN, B.; NISHIKUBO, N.; TAKAHASHI, J.; ASPEBORG, H.; DJERBI, S.; MASTER, E.; ANDERSSON-GUNNERÅS, S.; SUNDBERG, B.; KARPINSKI, S.; TEERI, T. T.; KLECZKOWSKI, L. A.; HENRISSAT, B.; MELLEROWICZ, E. J.. Poplar carbohydrate-active enzymes. Gene identification and expression analysis. **Plant Physiology**, Lancaster, v. 140, p. 946-962, 2006.

GIBBINGS, J. G.; COOK, B. P.; DUFAULT, M. R.; MADDEN, S. L.; JHURI, S.; TURNBULL, C. J.; DUNWELL, J. M. Global transcript of rice leaf and seed using SAGE technology. **Plant Biotechnology Journal**, Oxford, v. 1, p. 271-285, 2003.

GION, J.; LALANNE, C.; LE PROVOST, G.; FERRY-DUMAZET, H.; PAIVA, J.; CHAUMEIL, P.; FRIGERIO, J.; BRACH, J.; BARRÉ, A.; DE DARUVAR, A.; CLAVEROL, S.; BONNEU, M.; SOMMERER, N.; NEGRONI, L.; PLOMION, C. The proteome of maritime pine wood forming tissue. **Proteomics**, Weinheim, v. 5, p. 3731-3751, 2005.

GOLDING, A. J.; JOHNSON, G. N. Down-regulation of linear and activation of cyclic electron transport during drought. **Planta**, Berlin, v. 218, n. 1, p. 107-114, 2003.

GONZALEZ, E. R.; DE ANDRADE, A.; BERTOLO, A. L.; LACERDA, G. C.; CARNEIRO, R. T.; DEFAVARI, V. A. P.; LABATE, M. T. V.; LABATE, C. A. Production of transgenic *Eucalyptus grandis* x *E urophylla* using the sonication-assisted *Agrobacterium* transformation (SAAT) system. **Functional Plant Biology**, Collingwood, v. 29, n. 1, p. 97-102, 2002.

GOWDA, M.; JANTASURIYARAT, C.; DEAN, R.A.; WANG, G.L. Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. **Plant Physiology**, Lancaster, v. 134, n. 3, p. 890-7, 2004.

GRACIET, E.; LEBRETON, S.; GONTERO, B. Emergence of new regulatory mechanisms in the Benson-Calvin pathway via protein-protein interactions: a glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase complex. **Journal of Experimental Botany**, Oxford, v. 55, n. 400, p. 1245-1254, 2004.

GRATTAPAGLIA, D., SEDEROFF, R. R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. **Genetics**, Austin, v. 137, p. 1121-1137, 1994.

GROOVER, A., JONES, A.M. Tracheary element differentiation uses a novel mechanism coordinating programmed cell death and secondary cell wall synthesis. **Plant Physiology**, Lancaster, v.119, p. 375-384, 1999.

HANSON, A.D., ROJE, S. 2001. ONE-CARBON METABOLISM IN HIGHER PLANTS. **Annual Review of Plant Physiology and Plant Molecular Biology**. Palo Alto, v. 52, p. 119-137, 2001.

HAYES, R; KUDLA, J; GRUISSEM, W. Degrading chloroplast mRNA: the role of polyadenylation. **Trends in Biochemical Sciences**, Amsterdam, v. 25, n. 5, p. 199-202, 1999.

HENRISSAT, B. A classification of glycosyl hydrolases base don amino acid sequence similarities. **The Biochemical journal**, London, v. 280, p. 309-316, 1991.

HELLMANN, H., ESTELLE, M. Plant development: regulation by protein degradation. **Science**, Washington, v. 297, p. 793-97, 2002.

HERTZBERG, M.; ASPEBORG, H.; SCHRADER, J.; ANDERSSON, A.; ERLANDSSON, R.; BLOMQVIST, K.; BHALERAO, R.; UHLÉN, M.; TEERI, T. T.; LUNDEBERG, J.; SUNDBERG, B.; NILSSON, P.; SANDBERG, G. 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.

HOFFMANN, L.; BESSEAU, S.; GEOFFROY, P.; RITZENTHALER, C.; MEYER, D.; LAPIERRE, C.; POLLET, B.; LEGRAND, M. Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. **Plant Cell**, Rockville, v. 16, p. 1446-1465, 2004.

HU, W.-J.; HARDING, S.A.; LUNG, J.; POPKO, J.L.; RALPH, J.; STOKKE, D.D.; TSAI, C.-J.; CHIANG, V.L. Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. **Nature Biotechnology**, New York , v. 17, p. 808-812, 1999.

IRELAND, R. J.; LEA, P. J. **The enzymes of glutamine, glutamate, asparagine, and aspartate metabolism**. In: SINGH, B. K. (Ed.). **Plant Amino Acids: Biochemistry and Biotechnology**. New York: Marcel Dekker, p. 49-109. 1999.

IRVING, L.; ROBINSON, D. On modelling Rubisco turnover: dynamics and applicability. **New Phytologist**, London, v. 170, n. 2, p. 204-205, 2006.

ISRAELSSON, M.; ERIKSSON, M. E.; HERTZBERG, M.; ASPEBORG, H.; NILSSON, P.; MORITZ, T. Changes in gene expression in the wood-forming tissue of transgenic hybrid aspen with increased secondary growth. **Plant Molecular Biology**, New York, v. 52, n. 4, p. 893-903, 2003.

JOSHI, C. P.; BHANDARI, S.; RANJAN, P.; KALLURI, U. C.; LIANG, X.; FUJINO, T.; SAMUGA, A. Genomics of cellulose biosynthesis in poplars. **The New phytologist**, London, v. 164, p. 53-61, 2004.

JUAN, D.; HONG-LI, X.; DE-QIANG, Z.; MIN-JIE, W.; YING-ZHANG, L.; KE-MING, C.; MENG-ZHU, L. Regeneration of the secondary vascular system in poplar as a novel system to investigate gene expression by a proteomic approach. **Proteomics**, Weinheim, v. 6, p. 881-895, 2006.

JUNG, S. H.; LEE, J. Y.; LEE, D. H. Use of SAGE technology to reveal changes in gene expression in *Arabidopsis* leaves undergoing cold stress. **Plant Molecular Biology**, The Hague, v. 52, p. 553-567, 2003.

KAL, A. J.; ZONNEVELD, A. J. VAN; BENES, V.; BERG, M. VAN DEN; KOERKAMP, M. G.; ALBERMANN, K.; STRACK, N.; RUIJTER, J. M.; RICHTER, A.; DUJON, B.; ANSORGE, W.; TABAK, H. F. Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. **Molecular Biology of The Cell**, Bethesda, v. 10, p. 1859-1872, 1999.

KAO, Y.Y.; HARDING, S.A.; TSAI, C.J. Differential expression of two distinct phenylalanine ammonia-lyase genes in condensed tannin-accumulating and lignifying cells of quaking aspen. **Plant Physiology**, Lancaster, v. 130, p. 796-807, 2002.

KEYS, A. J. The re-assimilation of ammonia produced by photorespiration and the nitrogen economy of C₃ higher plants. **Photosynthesis Research**, Hague, v.87, p. 165-175, 2006.

KIMMERER, T. W.; STRINGER, M. A. Alcohol Dehydrogenase and Ethanol in the Stems of Trees : Evidence for Anaerobic Metabolism in the Vascular Cambium. **Plant Physiology**, Lancaster, v. 87, n. 3, p. 693-697, July, 1988.

KO, J.H.; HAN, K.H.; PARK, S.; YANG, J. Plant body-induced secondary growth in *Arabidopsis* and its transcription phenotype revealed by whole-transcriptome profiling. **Plant Physiology**, Lancaster, v. 135, n. 2, p. 1069-1083, 2004.

KOCHHAR, A.; REES, T.A. P.; POLLOCK, C. J. Assimilate partitioning in leaves of the raffinose-storing herb *Lamium album* L: the effects of altering source-sink balance. **Revista Brasileira de Botânica**, São Paulo, v. 26, n. 4, p. 533-540, 2003.

KONISHI, T.; OHMIYA, Y.; HAYASHI, T. Evidence that sucrose loaded into the phloem of a poplar leaf is used directly by sucrose synthase associated with various β -glucan synthases in the stem. **Plant Physiology**, Lancaster, v. 134, p. 1146-1152, 2004.

KREUZWIESER, J.; PAPADOPOULOU, E.; RENNENBERG, H. Interaction of flooding with carbon metabolism of forest trees. **Plant biology (Stuttgart, Germany)**, Stuttgart, v. 6, n. 3, p. 299-306, 2004.

LACOMBE, E.; HAWKINS, S.; DOORSSELAERE, J. van; PIQUEMAL, J.; GOFFNER, D.; POEYDOMENGE, O.; BOUDET, A. M.; GRIMA-PETTENATI, J. Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. **The Plant Journal**, Oxford, v. 11, n. 3, p. 429-441, 1997.

LARSON, P. R. **The vascular cambium: development and structure**. Berlin: Springer-Verlag, 1994. 351p.

LASH, A. E.; TOLSTOSHEV, C. M.; WAGNER, L.; SCHULER, G. D.; STRAUSBERG, R. L.; RIGGINS, G. J.; ALTSCHUL, S. F. SAGEmap: A public gene expression resource. **Genome Research**, Cold Spring Harbor, v. 10, p. 1051-1060, 2000.

LEE, J., LEE, D. Use serial analysis of gene expression technology to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. **Plant Physiology**, Lancaster, v. 132, p. 517-529, 2003.

LEE, J.; EOM, E.; KIM, D.; LEE, Y. M.; LEE, D. Analysis of gene expression profiles of gastric normal and cancer tissues by SAGE. **Genomics**, San Diego, v. 82, p. 78-85, 2003.

LEE, D.; MEYER, K.; CHAPPLE, C.; DOUGLAS, C.J. Antisense suppression of 4-coumarate:coenzyme A ligase activity in *Arabidopsis* leads to altered lignin subunit composition. **The Plant Cell**, Rockville, v. 9, p. 1985-1998, 1998.

LEITE, N. B. Avanços da silvicultura brasileira são significativos. **Visão Agrícola**, Piracicaba, v. 4, p. 58-61, 2005.

LEWIS, N. G. A 20th century roller coaster ride: a short account of lignification. **Current Opinion in Plant Biology**, London, v. 2, n. 2, p. 153-162, 1999.

LI, L.; LU, S.; CHIANG, V. A genomic and molecular view of wood formation. **Critical Reviews in Plant Sciences**, Boca Raton, v. 25, p. 215-233, 2006.

LIANG, P., PARDEE, A. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. **Science**, Washington, v. 257, p. 967-971, 1992.

LINCOLN, T., ZEIGER, E. **Fisiologia Vegetal**. 3 ed. Porto Alegre: Artmed, 2004. 719p.

LORENZ, W. W., DEAN, J. F. D. SAGE Profiling and demonstration of differential gene expression along the axial developmental gradient xylem in loblolly pine (*Pinus taeda*). **Tree Physiology**, Victoria, v. 22, n. 5, p. 301-310, 2002.

LU, S.; ZHOU, Y.; LI, L.; CHIANG, V. L. Distinct roles of cinnamate 4-hydroxylase genes in *Populus*. **Plant & Cell Physiology**, Kyoto, v. 47, n. 7, p. 905-914, 2006.

LUNKENBEIN, S.; BELLIDO, M.; AHARONI, A.; SALENTIJN, E. M.; KALDENHOFF, R.; COINER, H. A.; MUNOZ-BLANCO, J.; SCHWAB, W. Cinnamate metabolism in ripening fruit. Characterization of a UDP-glucose:cinnamate glucosyltransferase from strawberry. **Plant Physiology**, Lancaster, v. 140, n. 3, p. 1047-1058, 2006.

MacDONALD, R.C., KIMMERER, T.W. Ethanol in the stems of trees. **Physiologia Plantarum**, Lund, v. 82, p. 582-588, 1991.

MacDOUGALL, G. J.; MORRISON, I. M.; STEWART, C.; WEYERS, J. D. B.; HILLMAN, J.R. Plant fibers: botany, chemistry and processing for industrial use. **Journal of the Science of Food and Agriculture**, London, v. 62, p. 1-20, 1993.

MARTIN, C., SMITH, A.M. Starch biosynthesis. **The Plant Cell**, Rockville, v. 7, n. 7, p. :971-85, 1995.

MATSUMURA, H.; NIRASAWA, S.; TERAUCHI, R. Transcript profiling in rice (*Oryza sativa* L) seedlings using the serial analysis of gene expression (SAGE). **The Plant Journal**, Oxford, v. 20, n. 6, p. 719-726, 1999.

MATSUMURA, H.; REICH, S.; ITO, A.; SAITOH, H.; KAMOUN, S.; WINTER, P.; KALS, G.; REUTER, M.; KRÜGER, D.H., TERAUCHI, R. Gene expression analysis of plant host-patogen interactions by SuperSAGE. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 100, n. 26, p. 15718-15723, 2003.

MEHLER, A.H. Studies on reactivities of illuminated chloroplasts. I. Mechanism of the reduction of oxygen and other Hill reagents. **Archives of Biochemistry. Biophysics**, London, v. 33, p. 65–77, 1951.

MELLEROWICZ, E.J.; BAUCHER, M.; SUNDBERG, B.; BOERJAN, W. Unravelling cell wall formation in the woody dicot stem. **Plant Molecular Biology**, New York, v. 47, p. 239-274, 2001.

MORA, A., GARCIA, C. H. **A cultura do eucalipto no Brasil**. São Paulo: Sociedade Brasileira de Silvicultura, 2000. 112 p.

MORALEJO, M.; ROCHANGE, F.; BOUDET, A. M.; TEULIERES, C. Generation of transgenic *Eucalyptus globulus* plantlets through *Agrobacterium tumefaciens* mediated transformation. **Australian Journal of Plant Physiology**, Melbourne, v. 25, n. 2, p. 207-212, 1998.

MOREAU, C.; AKSENOV, N.; LORENZO, M. G.; SEGERMAN, B.; FUNK, C.; NILSSON, P.; JANSSEN, S.; TUOMINEN, H. A genomic approach to investigate developmental cell death in woody tissues of *Populus* trees. **Genome Biology**, London, v. 6, n. 4, p. 1-14, 2005.

NACHT, M.; FERGUSON, A.T.; ZHANG, W.; PETROZIELLO, J.M.; COOK, B.P.; GAO, Y.H.; MAGUIRE, S.; RILEY, D.; COPPOLA, G.; LANDES, G.M.; MADDEN, S.L.; SUKUMAR, S. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. **Cancer Research**, Chicago, v. 59, p. 5464-5470, 1999.

NELSON, N., YOCUM, C. F. Structure and Function of Photosystems I and II. **Annual Review of Plant Biology**, Palo Alto, v. 57, p. 521-565, 2006.

NIELD, J; BARBER, J. Refinement of the structural model for the Photosystem II supercomplex of higher plants. **Biochimica et Biophysica Acta**, Amsterdam, v. 1757, n. 5-6, p. 353-361, 2006.

NIYOGI, K. K. Photoprotection Revisited: Genetic and Molecular Approaches. **Annual Review of Plant Physiology and Plant Molecular Biology**. Palo Alto, v. 50, p. 333-359, 1999.

NORWOOD, M.; TRUESDALE, M. R.; RICHTER, A.; SCOTT, P. Photosynthetic carbohydrate metabolism in the resurrection plant *Craterostigma plantagineum*. **Journal of Experimental Botany**, Oxford, v.51, n. 343, p. 159-165, 2000.

ODA, Y.; MIMURA, T.; HASEZAWA, S. Regulation of secondary cell wall development by cortical microtubules during tracheary element differentiation in *Arabidopsis* cell suspensions. **Plant Physiology**, Lancaster, n. 137, p. 1027-1036, 2005.

OH, S.; PARK, S.; HAN, K. Transcriptional regulation of secondary growth in *Arabidopsis thaliana*. **Journal of Experimental Botany**, Oxford, v. 54, n. 393, p. 2709-2722, 2003.

PARVATHI, K.; CHEN, F.; GUO, D.; BLOUNT, J. W.; DIXON, R. A. Substrate preferences of O-methyltransferases in alfalfa suggest new pathways for 3-O-methylation of monolignols. **The Plant Journal**, Oxford, v. 25, n. 2, p. 193-202, 2001.

PAUX, E.; CAROCHA, V.; MARQUES, C.; DE SOUSA, A. M.; BORRALHO, N.; SIVADON, P.; GRIMA-PETTENATI, J. Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. **New Phytologist**, London, v. 167, p. 89-100, 2005.

PAUX, E.; TAMASLOUKHT, M.; LADOUCE, N.; SIVADON, P.; GRIMA-PETTENATI, J. Identification of genes preferentially expressed during wood formation in *Eucalyptus*. **Plant Molecular Biology**, The Hague, v. 55, p. 263-280, 2004.

PENG, L.; KAWAGOE, Y.; HOGAN, P.; DELMER, D. Sitosterol-beta-glucoside as primer for cellulose synthesis in plants. **Science**, Washington, v. 295, n. 5552, p. 147-150, 2002.

PILATE, G.; GUINEY, E.; HOLT, K.; PETIT-CONIL, M.; LAPIERRE, C.; LEPLE, J. C.; POLLET, B.; MILA, I.; WEBSTER, E. A.; MARSTORP, H. G.; HOPKINS, D. W.; JOUANIN, L.; BOERJAN, W.; SCHUCH, W.; CORNU, D.; HALPIN, C. Field and pulping performances of transgenic trees with altered lignification. **Nature Biotechnology**, New York, v. 20, p. 607-612, 2002.

PLAXTON, W.C. The organization and regulation of plant glycolysis. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 47, p. 185–214, 1996.

PLAXTON, W.C., PODESTÁ, F.E. The Functional Organization and Control of Plant Respiration. **Critical Reviews in Plant Sciences**, Boca Raton, v. 25, p. 159–198, 2006.

PLEASANCE, E.D.; MARRA, M.A.; JONES, S.J.M. Assessment of SAGE in transcript identification. **Genome Research**, Cold Spring Harbor, v. 13, p. 1203-1215, 2003.

PLOMION, C.; LEPROVOST, C.; STOKES, A. Wood formation in trees. **Plant Physiology**, Lancaster, v. 127, p. 1513-1523, 2001.

POEYDOMENGE, O.; BOUDET, A. M., GRIMA-PETTENATI, J. A cDNA encoding S-adenosyl-L-methionine:caffeic acid 3-O-methyltransferase from Eucalyptus. **Plant Physiology**, Lancaster, v.105, n. 2, p. 749-750, 1994.

POROYKO, V.; HEJLEK, L.G.; SPOLLEN, W.G.; SPRINGER, G.K.; NGUYEN, H.T.; SHARP, R.E.; BOHNERT, H.J. The maize root transcriptome by serial analysis of gene expression. **Plant Physiology**, Lancaster, v. 138, n. 3, p. 1700-10, 2005.

PORTIS Jr, A. R. Rubisco activase - Rubisco's catalytic chaperone. **Photosynthesis Research**, Hague, v. 75, n. 1, p. 11-27, 2003.

PORTIS Jr, A. R.. The regulation of Rubisco by Rubisco activase. **Journal of Experimental Botany**, Oxford, v. 46, p. 1285–1291, 1995.

POT, D.; CHANTRE, G.; ROZENBERG, P.; RODRIGUES, J. C.; JONES, G. L.; PEREIRA, H.; HANINRUP, B.; CAHALAN, C.; PLOMION, C. Genetic control of pulp and timber properties in maritime pine (*Pinus pinaster* Ait.). **Annals of Forest Science**, Paris, v. 59, n. 5/6, p. 563-575, 2002.

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

RAES, J.; ROHDE, A.; CHRISTENSEN, J. H.; PEER, Y. van der; BOERJAN, W. Genome-wide characterization of the lignification toolbox in *Arabidopsis*. **Plant Physiology**, Lancaster, v. 133, p. 1051-1071, 2003.

RALPH, S.; PARK, J. Y.; BOHLMANN, J.; MANSFIELD, S. D. Dirigent proteins in conifer defense: gene discovery, phylogeny, and differential wound- and insect-induced expression of a family of DIR and DIR-like genes in spruce (*Picea* spp.). **Plant Molecular Biology**, The Hague, v. 60, n. 1, p. 21-40, 2006.

RANIK, M.; CREUX, N. M.; MYBURG, A. A. Within-tree transcriptome profiling in wood-forming tissues of a fast-growing *Eucalyptus* tree. **Tree Physiology**, Victoria, v. 26, p. 265-375, 2006.

- RANIK, M.; MYBURG, A. A. Six new cellulose synthase genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. **Tree Physiology**, Victoria, v. 26, p. 545-556, 2006.
- RAYMOND, C. Genetics of *Eucalyptus* wood properties. **Annals of Forest Science**, Paris, v. 59, n. 5/6, p. 525-531, 2002.
- REITER, W. Biosynthesis and properties of the plant cell wall. **Current Opinion in Plant Biology**, London, v. 5, p. 536-542, 2002.
- REITER, W.; VANZIN, G. F. Molecular genetics of nucleotide sugar interconversions pathways in plants. **Plant Molecular Biology**, New York, v. 47, p. 95-113, 2001.
- REYNOLDS, M. P.; VAN GINKEL, M.; RIBAUT, J. M. Avenues for genetic modification of radiation use efficiency in wheat. **Journal of Experimental Botany**, Oxford, v. 51, p. 459-473, 2000.
- RISON, S. C.; HODGMAN, T. C.; THORNTON, J. M. Comparison of functional annotation schemes for genomes. **Functional & Integrative Genomics**, Berlin, v.1, n. 1, p. 56-69, 2000.
- RITTER, E.; GEBHARDT, C.; SALAMINI, F. Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. **Genetics**, Austin, v. 125, p. 645-654, 1990.
- ROBINSON, S. J.; CRAM, D. J.; LEWIS, C. T.; PARKIN, I. A. P. Maximizing the efficacy of SAGE analysis identifies novel transcripts in *Arabidopsis*. **Bioinformatics**, Oxford, v. 136, p. 3223-3233, 2004.
- RONNING, C.M.; STEGALKINA, S.S.; ASCENZI, R.A.; BOUGRI, O.; HART, A.L.; UTTERBACH, T.R.; VANAKEN, S.E.; RIEDMULLER, S.B.; WHITE, J.A.; CHO, J.; PERTEA, G.M.; LEE, Y.; KARAMYCHEVA, S.; SULTANA, R.; TSAI, J.; QUACKENBUSH, J.; GRIFFITHS, H.M.; RESTREPO, S.; SMART, C.D.; FRY, W.E.; VAN DER HOEVEN, R.; TANKSLEY, S.; ZHANG, P.; JIN, H.; YAMAMOTO, M.L.; BAKER, B.J.; BUELL, C.R. Comparative analyses of potato expressed sequence tag libraries. **Plant Physiology**, Lancaster, v. 131, p. 419-429, 2003.
- ROS BARCELO, A. Xylem parenchyma cells deliver the H₂O₂ necessary for lignification in differentiating xylem vessels. **Planta**, Berlin, v. 220, n. 5, p.747-56, 2005.
- ROS GOMEZ, L.V.; PARADISO, A.; GABALDON, C.; PEDRENO, M.A.; DE GARA, L.; ROS BARCELO, A. Two distinct cell sources of H₂O₂ in the lignifying *Zinnia elegans* cell culture system. **Protoplasma**, Leipzig, v. 227, n. 2-4, p.175-83, 2006.
- SAKAMOTO, W. Protein Degradation Machineries in Plastids. **Annual Review of Plant Biology**, Palo Alto, 57:599-621, 2006.

SALZMAN, R. A.; FUJITA, T.; ZHU-SALZMAN, K.; HASEGAWA, P. M.; BRESSAN, R. A. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. **Plant molecular biology reporter / ISPMB**, New York, v. 17, p. 11-17, 1999.

SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Preparation of reagents and buffers used in molecular cloning**: molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press, 1989. chap. Appendix E, P. E.1-E.39.

SHARKEY, T.D. Estimating the rating of photorespiration in leaves. **Physiologia Plantarum**, Lund, v. 73, n. 1, p. 147-152, 1988.

SCHILTZ, S.; GALLARDO, K.; HUART, M.; NEGRONI, L.; SOMMERER, N.; BURSTIN, J. Proteome reference maps of vegetative tissues in pea. An investigation of nitrogen mobilization from leaves during seed filling. **Plant Physiology**, Lancaster, v. 135, n. 4, p. 2241-1160, 2004.

SCHOCH, G.; GOEPFERT, S.; MORANT, M.; HEHN, A.; MEYER, D.; ULLMANN, P.; WERCK-REICHHART, D. CYP98A3 from *Arabidopsis thaliana* is a 3-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. **The Journal of Biological Chemistry**, Baltimore, v. 276, n. 39, p. 36566–36574, 2001.

SCHRADER, J.; NILSSON, J.; MELLEROWICZ, E.; BERGLUND, A.; NILSSON, P.; HERTZBERG, M.; SANDBERG, G. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. **The Plant Cell**, Rockville, v. 16, n. 9, p. 2278-92, 2004.

SERRANO, L.; ROCHANGE, F.; SEMBLAT, J. P.; MARQUE, C.; TEULIERES, C.; BOUDET, A. M. Genetic transformation of *Eucalyptus globulus* through biolistics: Complementary development of procedures for organogenesis from zygotic embryos and stable transformation of corresponding proliferating tissue. **Journal of Experimental Botany**, Oxford, v. 47, n. 295, p. 285-290, 1996.

SHOJI, T.; WINZ, R.; IWASE, T.; NAKAJIMA, K.; YAMADA Y, HASHIMOTO T. Expression patterns of two tobacco isoflavone reductase-like genes and their possible roles in secondary metabolism in tobacco. **Plant Molecular Biology**, New York, v. 50, n. 3, p. 427-440, 2002.

SILVA, J. C. Cresce a presença do eucalipto no Brasil. **Revista da Madeira**, Curitiba, n. 92, p. 61-66, out. 2005.

SMALLE, J., VIERSTRA, R.D. The ubiquitin 26s proteasome proteolytic pathway. **Annual Review of Plant Biology**, Palo Alto, 55:555–90, 2004.

SOCIEDADE BRASILEIRA DE SILVICULTURA. **Fatos e números do setor florestal brasileiro**. São Paulo, 2005. 3 p.

SOUZA, G.A. de; GODOY, L.M.; TEIXEIRA, V.R.; OTAKE, A.H.; SABINO, A.; ROSA, J.C.; DINARTE, A.R.; PINHEIRO, D.G.; SILVA, W.A.; EBERLIN, M.N.; CHAMMAS, R.; GREENE, L.J. Proteomic and SAGE profiling of murine melanoma progression indicates the reduction of proteins responsible for ROS degradation. **Proteomics**, Weinheim, v. 6, n. 5, p. 1460-1470, 2006.

SPRENGER, N.; KELLER, F. Allocation of a raffinose family oligosaccharide to transport and storage pools in *Ajuga reptans*: the roles of two distinct galactinol synthases. **The Plant Journal**, Oxford, v. 21, n. 3, p. 249-258, 2000.

STEANE, D. A. Complete Nucleotide Sequence of the Chloroplast Genome from the Tasmanian Blue Gum, *Eucalyptus globulus* (Myrtaceae). **DNA research : an international journal for rapid publication of reports on genes and genomes**, Tokyo, v. 12, n. 3, p. 215-220, 2005.

STERKY, F.; REGAN, S.; KARLSSON, J.; HERTZBERG, M.; ROHDE, A.; HOLMBERG, A.; AMINI, B.; BHALERAO, R.; LARSSON, M.; VILLARROEL, R.; VAN MONTAGU, M.; SANDBERG, G.; OLSSON, O.; TEERI, T.T.; BOERJAN, W.; GUSTAFSSON, P.; UHLÉN, M.; SUNDBERG, B.; LUNDERBERG, J. Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 95, p. 1330-1335, 1998.

STOLLBERG, J.; URSCHITZ, J.; URBAN, Z.; BOYD, C.D. A Quantitative Evaluation of SAGE. **Genome Research**, Cold Spring Harbor, v. 10, p. 1241-1248, 2000.

SUN, M.; ZHOU, G.; LEE, S.; CHEN, J.; SHI, R.Z.; WANG, S.M. SAGE is far more sensitive than EST for detecting low-abundance transcripts. **BMC Genomics**, London, v. 5, n. 1, p.1-4, 2004.

SUZUKI, S.; YAMAMURA, M.; SHIMADA, M.; UMEZAWA, T. A heartwood norlignan, (E)-hinokiresinol, is formed from 4-coumaryl 4-coumarate by a *Cryptomeria japonica* enzyme preparation. **Chemical communications (Cambridge, England)**, Cambridge, v. 21, n. 24, p. 2838-2839, 2004.

TIMELL, T. E. **Compression wood in Gymnosperms**. ed. Springer-Verlag: Heidelberg. 1986.

TSAI, C.J.; POPKO, J.L.; MIELKE, M.R.; HU, W.J.; PODILA, G.K.; CHIANG, V.L. Supresion of O-methyltransferase gene by homologous sense transgene in quaking aspen causes red-brown Word phenotypes. **Plant Physiology**, Lancaster, v. 117, p. 101-112, 1998.

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, p. 689-701, 1997.

TUSKAN, A.; DIFAZIO, S.; JANSSON, S.; BOHLMANN, J.; GRIGORIEV, I.; HELLSTEN, U.; PUTNAM, N.; RALPH, S.; ROMBAUTS, S.; SALAMOV, A.; SCHEIN, J.; STERCK, L.; AERTS, A.; BHALERAO, R. R.; BHALERAO, R. P.; BLAUDEZ, D.; BOERJAN, W.; BRUN, A.; BRUNNER, A.; BUSOV, V.; CAMPBELL, M.; CARLSON, J.; CHALOT, M.; CHAPMAN, J.; CHEN, G. L.; COOPER, D.; COUTINHO, P. M.; COUTURIER, J.; COVERT, S.; CRONK, Q.; CUNNINGHAM, R.; DAVIS, J.; DEGROEVE, S.; DE JARDIN, A.; DEPAMPHILIS, C.; DETTER, J.; DIRKS, B.; DUBCHAK, I.; DUPLESSIS, S.; EHLTING, J.; ELLIS, B.; GENDLER, K.; GOODSTEIN, D.; GRIBSKOV, M.; GRIMWOOD, J.; GROOVER, A.; GUNTER, L.; HAMBERGER, B.; HEINZE, B.; HELARIUTTA, Y.; HENRISSAT, B.; HOLLIGAN, D.; HOLT, R.; HUANG, W.; ISLAM-FARIDI, N.; JONES, S.; JONES-RHOADES, M.; JORGENSEN, R.; JOSHI, C.; KANGASJA, J.; KARLSSON, J.; KELLEHER, C.; KIRKPATRICK, R.; KIRST, M.; KOHLER, A.; KALLURI, U.; LARIMER, F.; LEEBENS-MACK, J.; LEPLE, J. C.; LOCASCIO, P.; LOU, Y.; LUCAS, S.; MARTIN, F.; MONTANINI, B.; NAPOLI, C.; NELSON, D.R.; NELSON, C.; NIEMINEN, K.; NILSSON, O.; PEREDA, V.; PETER, G.; PHILIPPE, R.; PILATE, G.; POLIAKOV, A.; RAZUMOVSKAYA, J.; RICHARDSON, P.; RINALDI, C.; RITLAND, K.; ROUZE', P.; RYABOY, D.; SCHMUTZ, J.; SCHRADER, J.; SEGERMAN, B.; SHIN, H.; SIDDIQUI, A.; STERKY, F.; TERRY, A.; TSAI, C. J.; UBERBACHER, E.; UNNEBERG, P.; VAHALA, J.; WALL, K.; WESSLER, S.; YANG, G.; YIN, T.; DOUGLAS, C.; MARRA, M.; SANDBERG, G.; VAN DE PEER, Y.; ROKHSA, D. The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). **Science**, Washington, v. 313, p. 1596-1604, 2006.

TZVI, T.; ZUKER, A.; ALTMAN, A. Forest-tree biotechnology: genetic transformation and its application to future forests. **Trends in Biotechnology**, Amsterdam, v. 16, p. 439-446, 1998.

VANDER-MIJNSBRUGGE, K.; MEYERMANS, H.; MONTAGU, M. VAN; BAUW, G.; BOERJAN, W. Wood formation in poplar: identification, characterization, and seasonal variation of xylem proteins. **Planta**, Berlin, v. 210, n. 4, p. 589-598, 2000.

VELCULESCU, V.E.; ZHANG, L.; VOGELSTEIN, B.; KINZLER, K.W. Serial analysis of gene expression. **Science**, Washington, v. 270, p. 484-487, 1995.

VELCULESCU, V.E.; ZHANG, L.; ZHOU, W.; VOGELSTEIN, J.; BASRAI, M.A.; BASSET, D.E.; HIETER, P.; VOGELSTEIN, B.; KINZLER, K.W. Characterization of the yeast transcriptome. **Cell**, Cambridge, v. 88, p. 243-251, 1997.

VINCENT, D.; LAPIERRE, C.; POLLET, B.; CORNIC, G.; NEGRONI, L.; ZIVY, M. Water deficits affect caffeate O-methyltransferase, lignification, and related enzymes in maize leaves. A proteomic investigation. **Plant Physiology**, Lancaster, v. 137, n. 3, p. 949-960, 2005.

VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; VAN DE LEE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Research**, London, v. 23, n. 21, p. 4407-14, 1995.

WANG, S.; FEARS, S.; ZHANG, L.; CHEN, J.; ROWLEY, J. Screening poly(dA/dT)-cDNAs for gene identification. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 97, p. 4162-4167, 2000.

WEI, C.L.; PATRICK, N.G.; CHIU, K.P.; WONG, C.H.; ANG, C.C.; LIPOVICH, L.; LIU, E.T.; RUAN, Y. 5' Long serial analysis of gene expression (LongSAGE) and 3' LongSAGE for transcriptome characterization and genome annotation. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 101, n. 32, p. 11701-6, 2004.

WERNEKE, J. J. M.; CHATFIELD, J. M.; OGREN, W. L. Alternative mRNA splicing generates the two Ribulosebisphosphate Carboxylase/Oxygenase Activase Polypeptides in spinach and arabidopsis. **The Plant Cell**, Rockville, v. 1, n. 8, p. 815-825, 1989.

YANG, J.; PARK, S.; KAMDEM, D. P.; KEATHLEY, D. E.; RETZEL, E.; PAULE, C.; KAPUR, V.; HAN, K. H. Novel gene expression profiles define the metabolic and physiological processes characteristic of wood and its extractive formation in a hardwood tree species, *Robinia pseudoacacia*. **Plant Molecular Biology**, New York, v. 52, n. 5, p. 935-956, 2003.

ZHAO, C.; JOHNSON, B.J.; KOSITSUP, B.; BEERS, E.P. Exploiting secondary growth in *Arabidopsis*. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. **Plant Physiology**, Lancaster, v. 123, p. 1185-1196, 2000.

ZHONG, R.; MORRISON III, W. H.; HIMMELSBACH, D. S.; POOLE II, F. L.; YE, Z. Essential role of caffeoyl coenzyme A *O*-methyltransferase in lignin biosynthesis in woody poplar plants. **Plant Physiology**, Lancaster, v. 124, p. 563-577, 2000.

ZHONG, R.; BURK, D.H.; YE, Z. Fibers. A model for studying cell differentiation, cell elongation, and cell wall biosynthesis. **Plant Physiology**, Lancaster, v. 126, p. 477-479, 2001.

ANEXOS

ANEXO A

Tratamento do material para extração de RNA

1 Tratamento dos tubos

- (a) Lavagem dos tubos com extran;
- (b) Lavagem com SDS 2% em forno a 70°C por 2 horas;
- (c) Lavagem com álcool comercial em agitador por 20 minutos;
- (d) Lavagem com água oxigenada 3% em agitador por 20 minutos;
- (e) Lavagem com água DEPC 0,01% em agitador por 20 minutos;
- (f) Autoclavar os tubos por 15 minutos.

2 Preparo de reagentes e vidrarias

Todos reagentes devem ser preparados com água DEPC 0,1% e autoclavados

As vidrarias devem ser tratadas em forno à 180°C por 8 horas

ANEXO B

Protocolo de extração de RNA total (SALZMAN *et al.*, 1999)

- (a) Macerar 2g de tecido e adicionar 10ml do tampão de extração;
- (b) Agitar vigorosamente por 1 minuto;
- (c) Adicionar 1 volume de clorofórmio:álcool isoamílico (24:1) e agitar por 20';
- (d) Centrifugar à 16000 x g por 10 minutos à 4°C;
- (e) Transferir a fase superior para outro tubo;
- (f) Repetir os passos (c) e (d) até a interfase estar limpa;
- (g) Remover o sobrenadante para um novo tubo; e;
- (h) Adicionar 2X vol. de etanol 100% e 0,1X vol. de NaCl 5M;
- (i) Precipitar *overnight* à -2-°C;
- (j) Centrifugar à 16000 x g por 10 minutos à 4°C;
- (k) Ressuspender o *pellet* em 10ml de água DEPC 0,1%..
- (l) Se necessário, centrifugar por 5' à 20000 x g e recuperar o sobrenadante;
- (m) Adicionar 1X fenol saturado em TE (pH 8,0): clorofórmio : álcool isoamílico (25:24:1) e agitar por 10 minutos à temperatura ambiente;
- (n) Centrifugar à 13000 x g por 10 minutos à temperatura ambiente;
- (o) Transferir a fase superior a um novo tubo;
- (p) Repetir os passos (i) e (j) até a interfase estar limpa;
- (q) Adicionar 2X o vol. de álcool 100% e 0,1X o vol. de NaCl 5M;
- (r) Precipitar *overnight*;
- (s) Centrifugar à 16000 x g por 15 minutos à 4°C;
- (t) Ressuspender o *pellet* em 500 µl de água DEPC 0,1%.
- (u) Completar o volume para 1 ml com água DEPC e adicionar 333 µl de LiCl 8M;
- (v) Precipitar por 3 horas;
- (x) Centrifugar 20 minutos à 12000 x g à 4°C;
- (w) Lavar o *pellet* com álcool 80%;
- (y) Ressuspender o *pellet* em um volume apropriado de água DEPC 0,1% (50 µl).

ANEXO C

Tampão de extração (SALZMAN *et al.*, 1999)

Tiocinato de guanidina	4M
Tris-HCl pH 8,0	100mM
Citrato de sódio pH 8,0	25mM
Sarcosina N-lauril	0,5%
PVP solúvel	1%
2-mercaptoetanol	2%
Água DEPC 0,1%	Completar o volume

ANEXO D*Lysis/Binding buffer (I-SAGETM)*

Tris-HCl pH 7,5	100 mM
LiCl	500 mM
EDTA	10 mM
Duodecil sulfato de lítio	1%
DTT (DI- Tiotreitol)	5 mM

ANEXO E*Washing buffer A (I-SAGETM)*

Tris-HCl pH 7,5	10 mM
LiCl	0,15 M
EDTA	1 mM
Duodecil sulfato de lítio	0,1%
Glicogênio	10 µg/ml

ANEXO F*Washing buffer B (I-SAGETM)*

Tris-HCl pH 7,5	10 mM
LiCl	150 mM
EDTA	1 mM
Glicogênio	10 µg/ml

ANEXO G*1X First Strand buffer (I-SAGETM)*

Tris-HCl pH 8,3	50 mM
KCl	75 mM
MgCl ₂	3 mM
Duodecil sulfato de lítio	0,1%
Mussel glycogen	10 µg/ml

ANEXO H*Reação de síntese da primeira fita de cDNA (I-SAGETM)*

5X First Strand buffer	18 µl
RnaseOUT TM	1 µl
Água DEPC	54,5 µl
0,1 M DTT (DI - Tiotreitol)	9 µl
dNTP Mix (10 mM de cada)	4,5 µl

ANEXO I*RT SuperScriptTM II (Transcriptase reversa) (I-SAGETM) 200 U/µl em:*

Tris-HCl pH 7,5	20 mM
NaCl	100 mM
EDTA	0,1 mM
DTT	1mM
Nonidet – 40 (v/v)	0,01%
glicerol (p/v)	50%

ANEXO J*5X Second Strand buffer (I-SAGE™)*

Tris-HCl pH 6,9	100 mM
KCl	450 mM
MgCl ₂	23 mM
β – NAD ⁺	0,075 mM
Sulfato de amônia	50 mM

ANEXO K*dNTP Mix (I-SAGE™)*

dATP	10 mM
dGTP	10 mM
dCTP	10 mM
dTTP	10 mM
Em água DEPC	

ANEXO L*E. coli DNA ligase (I-SAGE™) 10 U/μl em:*

Tris-HCl pH 7,4	10 mM
KCl	50 mM
EDTA	0,1 mM
DTT	1 mM
BSA	200 μg/ml
Glicerol (p/v)	50%
Triton® X-100	0,1%

ANEXO M*E. coli DNA polimerase (I-SAGE™) 10 U/μl em:*

Fostato de potássio pH 7,4	50 mM
KCl	100 mM
DTT	1 mM
Glicerol (p/v)	50%

ANEXO N*E. coli RNase H (I-SAGE™) 2 U/μl em:*

Tris-HCl pH 7,5	20 mM
KCl	100 mM
EDTA	0,1 mM
DTT	0,1 mM
BSA	50 μg/ml
Glicerol (p/v)	50%
MgCl ₂	10 mM

ANEXO O*Washing buffer C (I-SAGE™)*

Tris-HCl pH 7,5	50 mM
NaCl	1 M
EDTA	0,5 mM
SDS (Duodecil Sulfato de Sódio)	1%
Mussel glycogen	10 μg/ml

ANEXO P*Washing buffer D (I-SAGETM)*

Tris-HCl pH 7,5	5 mM
NaCl	1 M
EDTA	0,5 mM
BSA	200 µg/ml

ANEXO Q*IX Buffer 4 (I-SAGETM)*

Tris-acetato pH 7,9	20 mM
Acetato de magnésio	10 mM
Acetato de potássio	50 mM
DTT	1 mM

ANEXO R*LoTE (I-SAGETM)*

Tris-HCl pH 7,5	3 mM
EDTA pH 7,5	0,2 mM

ANEXO S

100X BSA

BSA	10 mg/ml
Fosfato de potássio pH 7	20 mM
NaCl	50 mM
EDTA	0,1 mM
Glicerol (p/v)	5%

ANEXO T*Nla III (I-SAGETM) 10 U/µl em:*

KCl	200 mM
Tris-HCl pH 7,5	10 mM
EDTA	0,1 mM
DTT	1 mM
100X BSA	500 µg/ml
Glicerol (p/v)	50%

ANEXO U*IX Ligase buffer (I-SAGETM)*

Tris-HCl pH 7,5	6 mM
NaCl	5 mM
MgCl ₂	6 mM
BSA	0,1 mg/ml
β - mercaptoetanol	7 mM
ATP	0,1 mM
DTT	2 mM
Espermidina	1 mM

ANEXO V*T4 DNA ligase (I-SAGETM) 4 U/µl em:*

Tris-HCl pH 7,5	10 mM
Cloreto de potássio	50 mM
DTT	1 mM
Glicerol (p/v)	50%

ANEXO W

<i>BsmF I (I-SAGE™)</i> 2 U/μl em:	
NaCl	50 mM
Tris-HCl pH 7,4	10 mM
EDTA	0,1 mM
DTT	1 mM
BSA	200 μg/ml
Glicerol (p/v)	50%

ANEXO X

PCR teste – Verificação da digestão com <i>BsmF I</i>	1	2	3	4
10X BV <i>buffer</i>	5 μl	5 μl	5 μl	5 μl
DMSO	3 μl	3 μl	3 μl	3 μl
dNTP <i>mix</i>	5 μl	5 μl	5 μl	5 μl
DTP-1 ou DTP-2	2μl	2μl	2μl	2μl
UDGH anti-sense <i>primer</i>	2μl	2μl	2μl	2μl
<i>Platinum® Taq</i> DNA polimerase	0,5 μl	0,5 μl	0,5 μl	0,5 μl
Água DEPC	32 μl	32 μl	32 μl	32 μl
Amostra do tubo A	-	-	0,5 μl	-
Amostra do tubo B	-	-	-	0,5 μl
Controle positivo A (antes da digestão)	0,5 μl	-	-	-
Controle positivo B (antes da digestão)	-	0,5 μl	-	-

ANEXO Y

<i>10X Klenow buffer (I-SAGE™)</i>	
Tris-HCl pH 8,0	500 mM
NaCl	500 mM
MgCl ₂	100 mM

ANEXO Z

<i>Klenow polimerase (I-SAGE™)</i> 3-9 U/μl em:	
Fosfato de potássio pH 7,0	50 mM
KCl	100 mM
DTT	1 mM
Glicerol (p/v)	50%

ANEXO AA

Gel de poliacrilamida 12%	
40% Acrilamida/bisacrilamida	12,8 ml
Tampão de corrida TBE 5X	8 ml
TEMED	16 μl
Água deionizada	18,8 ml
Persulfato de amônio 10%	0,4 ml

ANEXO AB

Tampão BV 10X concentrado	
Sulfato de amônio	166 mM
Tris-HCl pH 8,8	670 mM
MgCl ₂	67 mM
β - mercaptoetanol	100 mM

ANEXO AC

Tampão de amostra 5X TBE

Tris base	18 mM
Ácido Bórico	18 mM
EDTA (livre de ácidos)	0,4 mM
Ficoll TM tipo 400	3%
Azul de bromofenol	0,02%
Xileno cianol	0,02%

ANEXO AD*pZErO[®] - I (I-SAGETM)*

<i>pZErO[®] - I</i>	1 µg/µl em tampão TE pH 7,5
------------------------------	-----------------------------

ANEXO AE

10X Tampão 2

Tris-HCl pH 7,9	100 mM
MgCl ₂	10 mM
NaCl	50 mM
DTT	1 mM

ANEXO AF*Sph I (I-SAGETM)* 5 U/µl em:

NaCl	100 mM
Tris-HCl pH 7,4	10 mM
EDTA	0,1 mM
DTT	1 mM
BSA	100 µg/ml
Glicerol (p/v)	50%

ANEXO AG

Meio SOC

Triptona	2%
Extrato de levedura	0,5%
MgCl ₂	10 mM
NaCl	10 mM
KCl	2,5 mM
MgSO ₄	10 mM
Glicose	20 mM

ANEXO AH

Meio LB sólido pH 7,5

Triptona	1%
Extrato de levedura	0,5%
NaCl	0,5%
Agar	1,5%

ANEXO AI*ZeocinTM (I-SAGETM)* (antibiótico)

<i>ZeocinTM</i>	100 mg/ml em água
----------------------------	-------------------