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

Análise serial da expressão gênica do caule de plantas de *Eucalyptus grandis* com 6 meses de idade

Raphael Tozelli Carneiro

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

Piracicaba
2007

Raphael Tozelli Carneiro
Engenheiro Agrônomo

**Análise serial da expressão gênica do caule de plantas de *Eucalyptus grandis*
com 6 meses de idade**

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
2007

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

Carneiro, Raphael Tozelli

Análise serial da expressão gênica do caule de plantas de *Eucalyptus grandis* com 6 meses de idade / Raphael Tozelli Carneiro. - - Piracicaba, 2007.
182 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 2007.
Bibliografia.

1. Caule 2. Eucalipto 3. Expressão gênica 4. Parede celular vegetal
5. Proteínas 6. Xilema I. Título

CDD 634.9734

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

*Aos meu pais, Ismael e Renata,
que jamais mediram esforços
para que eu pudesse chegar até aqui*

DEDICO

*À minha querida Dr^a Ana Letícia por
todo incentivo, amor, carinho, alegria e
ajuda, fundamentais para essa
conquista*

OFEREÇO

AGRADECIMENTOS

Em primeiro lugar a Deus, por me permitir acordar todos os dias e me dar forças para realizar minhas atividades;

Ao Prof. Dr. Carlos Alberto Labate pelo apoio e confiança desde a graduação;

Ao Dr. David Henry Moon, pelos ensinamentos, empenho e amizade, fundamentais para a realização desse trabalho;

Ao meu grupo de trabalho (SAGE), Daniele G. G. Caldas, Mayra C.C.G. Carvalho e Guillermo R. Salvatierra, pela constante troca de experiências, informações, apoio e alegria em cada passo desse trabalho;

À especialista em informática Livia M. Franceschini pela constante ajuda com o Banco de Dados;

A todos os colegas de trabalho do laboratório Max Feffer de Genética de Plantas, pessoas com as quais convivi e dividi experiências durante todos esses anos;

Aos meus grandes amigos Fernando “Gérsy” Andreote e Walter “Azeitona” Bernardi pelos momentos de alegria e por mostrar que a amizade verdadeira ainda pode existir;

A Fundação Rockefeller e ao Dr. Paulo T. Lacava, muito importantes na minha breve passagem pelos EUA;

Aos professores, funcionários e colegas do Dep. de Genética da ESALQ/USP;

A CAPES pela concessão da bolsa;

A Suzano Papel e Celulose pelo material vegetal e apoio financeiro;

Enfim, a todos aqueles que contribuíram de alguma forma para a realização desse trabalho, meu Grazzie Tanti !

Reach

*Some dreams live on in time forever
those dreams, you want with all your heart*

*and i'll do whatever it takes
follow through with the promise i made
put it all on the line
what i hoped for at last would be mine*

*if i could reach, higher
just for one moment touch the sky
from that one moment in my life
i'm gonna be stronger
know that i've tried my very best
i'd put my spirit to the test
if i could reach*

*some days are meant to be remembered
those days we rise above the stars*

*so i'll go the distance this time
seeing more the higher i climb
that the more i believe
all the more that this dream will be mine*

*if i could reach, higher
just for one moment touch the sky
from that one moment in my life
i'm gonna be stronger
know that i've tried my very best
i'd put my spirit to the test
if i could reach*

be stronger

Gloria Estefan

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS.....	15
LISTA DE TABELAS	17
LISTA DE ABREVIATURAS	19
1 INTRODUÇÃO	23
2 REVISÃO BIBLIOGRÁFICA.....	27
2.1 O gênero <i>Eucalyptus</i> – origem e distribuição	27
2.2 O setor florestal brasileiro e a importância do eucalipto	28
2.3 Melhoramento genético e a biotecnologia aplicada ao eucalipto.....	31
2.4 Genômica.....	38
2.5 Transcritômica.....	39
2.5.1 A técnica de SAGE (Análise Serial da Expressão Gênica)	45
2.6 Proteômica.....	50
2.7 A formação da madeira.....	53
2.7.1 Parede celular vegetal	53
2.7.1.1 Parede celular primária.....	54
2.7.1.2 Parede celular secundária.....	55
2.7.2 Xilogêse.....	58
2.7.3 Análise da expressão gênica	59
3 MATERIAL E MÉTODOS	65
3.1 Material vegetal	65
3.2 Extração do RNA total	66
3.3 Purificação do mRNA a partir do RNA total	68
3.4 Reação de RT-PCR teste	69
3.5 Construção da biblioteca SAGE	70
3.5.1 Isolamento do RNA mensageiro e síntese de cDNA	70
3.5.2 Digestão do cDNA com a endonuclease <i>Nla</i> III e ligação dos adaptadores.....	72
3.5.3 Digestão com a endonuclease <i>Bsm</i> FI e formação dos dímeros de cDNA.....	73

3.5.4 Otimização das reações de PCR para produção de dímeros em larga escala	74
3.5.5 Amplificação em larga escala e purificação dos dímeros de cDNA.....	76
3.5.6 Digestão dos dímeros de cDNA com <i>Nla</i> III.....	77
3.5.7 Polimerização/concatenação e clonagem das <i>ditags</i>	79
3.6 Sequenciamento dos polímeros de <i>ditags</i>	82
3.7 Bioinformática e análise dos dados.....	82
3.7.1 Seleção de sequências para análise.....	82
3.7.2 Extração das SAGE <i>tags</i>	84
3.7.3 identificação e anotação das sequências de origem das <i>tags</i> e suas prováveis funções celulares.....	85
3.7.4 Classificação funcional	88
3.8 Correlação entre nível de transcritos e abundância de proteínas	91
4 RESULTADOS E DISCUSSÃO	93
4.1 Análise geral dos dados produzidos pela biblioteca SAGE	93
4.2 Perfil transcricional dos dados produzidos pela biblioteca SAGE	97
4.3 Processo de formação da madeira	115
4.3.1 Divisão celular	115
4.3.2 Expansão celular	115
4.3.3 Deposição da parede celular	118
4.3.3.1 Biossíntese de hemicelulose e pectina	118
4.3.3.2 Biossíntese de celulose	122
4.3.4 Lignificação.....	125
4.3.5 Morte celular programada (PCD).....	136
4.4 Correlação entre nível de transcritos e abundância de proteínas.....	137
5 CONSIDERAÇÕES FINAIS	143
REFERÊNCIAS	147
ANEXOS	171

RESUMO

Análise serial da expressão gênica do caule de plantas de *Eucalyptus grandis* com 6 meses de idade

De todas as adaptações que as plantas sofreram durante a evolução, a aquisição do sistema vascular à 400 milhões de anos atrás, foi sem dúvida um evento decisivo para sua bem sucedida existência na terra. A madeira é considerada o mais importante recurso natural de energia renovável e o setor econômico baseado na produção florestal cresce a cada ano. Inúmeros fatores como rápida taxa de crescimento, grande produção de biomassa, adaptabilidade a diversos ambientes e solos, boa qualidade de madeira para produção de uma ampla gama de produtos e presença de celulose de fibra curta, ideal para a produção de papel e celulose, contribuíram para o grande sucesso das espécies de *Eucalyptus* e tornaram o Brasil o maior produtor mundial de celulose de fibra curta utilizando o eucalipto como matéria prima. Devido à reconhecida importância econômica e também ambiental das árvores, o desenvolvimento do sistema vascular se tornou um importante e fascinante processo biológico para se estudar. No entanto, existe ainda pouco conhecimento sobre os processos celulares, moleculares e bioquímicos envolvidos na formação da madeira. Dessa forma, no presente trabalho foi utilizada a técnica de SAGE (*Serial Analysis of Gene Expression*) para caracterizar o perfil transcricional do caule de plantas de *E. grandis* com 6 meses de idade. A partir do sequenciamento de 826 clones, foi possível analisar 2.274 tags/genes, sendo que 989 (43,5%) genes puderam ser identificados e ter uma possível função atribuída. Genes que codificam enzimas e proteínas muito importantes durante o processo de formação da madeira, como aqueles relacionados à biossíntese e deposição da parede celular e organização do citoesqueleto, apresentaram elevada expressão, sendo possível ainda sugerir a ocorrência de possíveis mecanismos comuns de controle transcricional para grupos de genes funcionalmente relacionados. A posterior comparação com as proteínas identificadas por espectrometria de massas através do sistema LC-MS/MS a partir do mesmo material biológico mostrou que muitos desses genes representam também as proteínas mais abundantes. Juntamente com outros projetos que vêm sendo desenvolvidos no laboratório, o presente trabalho contribuiu para a construção de um banco de dados local com informações do transcritoma e do proteoma de diferentes idades e tecidos, fornecendo uma visão global sobre os genes envolvidos no processo de formação da madeira e possivelmente responsáveis pelo rápido crescimento nas espécies de *Eucalyptus*, indicando importantes alvos para futuros programas de melhoramento.

Palavras-chave: *Eucalyptus*; SAGE; Caule; Xilema; Parede Celular; Proteínas

ABSTRACT

Serial analysis of gene expression in stems of 6 months old *Eucalyptus grandis*

From the numerous adaptations that plants have developed during evolution, the acquisition of the vascular system some 400 million years ago was been a decisive event for their successful existence on earth. Wood is considered the most important natural resource of renewable energy and the Forest-based economical sector grows every year. Several factors like the fast growth rate, large biomass production, adaptability to a wide range of environments and soils, good wood quality for the production of a wide range of products and the presence of short cellulose fiber, suitable for pulp and paper production, have contributed to the great success of *Eucalyptus* species making Brazil the main producer of short cellulose fiber using eucalypt as raw material. Due to the recognized economical and also environmental importance of trees, the development of the vascular system became an important and fascinating biological process to study. However, little is known about the cellular, molecular and biochemical processes involved in wood formation. In this way, in the current work SAGE (Serial Analysis of Gene Expression) technique was used to characterize the transcriptional profile in stems of 6 months old *Eucalyptus grandis*. From the sequencing of 826 clones, it was possible to analyse 2,274 tags/genes, and 989 (43,5%) genes could be identified and to have a possible function attributed. Genes that code for enzymes and proteins important for wood formation process, like those related to cell wall biosynthesis and deposition, and cytoskeleton organization had high expression, making it possible to suggest the occurrence of a common transcriptional control for a few functionally related genes. The posterior comparison with the set of proteins identified by LC ESI-MS/MS from the same biological material showed that some of these genes also represent the most abundant proteins. Taken together with other projects that are being developed in the laboratory, the present work contributed for the construction of a local data-base with transcriptome and proteome information from different ages and tissues, giving a global vision of the genes involved in the wood formation process and potentially responsible for the fast growth in the *Eucalyptus* species, indicating important targets for future breedings programs.

Key-words: *Eucalyptus*; SAGE; Stem; Xylem; Cell Wall; Proteins

LISTA DE FIGURAS

Figura 1 -	Representação esquemática da construção da biblioteca de SAGE.....	46
Figura 2 -	Representação esquemática da parede celular primária, mostrando a interação dos seus constituintes (celulose, hemicelulose e pectina).....	54
Figura 3 -	Representação esquemática da parede celular secundária, mostrando a presença da parede primária; das camadas S1, S2 E S3 da parede secundária; lúmen e lamela média.....	56
Figura 4 -	Material vegetal utilizado no experimento.....	65
Figura 5 -	Gel de eletroforese do produto do teste de RT-PCR realizado com a amostra de RNA total das plantas de 6 meses de idade usando primers para a UDP-glucose pyrophosphorylase de <i>E. grandis</i>	70
Figura 6 -	Gel de eletroforese dos produtos das reações de otimização da concentração de DNA para PCR.....	75
Figura 7 -	Gel de eletroforese dos produtos da amplificação em larga escala dos dímeros de cDNA.....	77
Figura 8 -	Gel de eletroforese dos produtos da digestão com <i>Nla</i> III.....	78
Figura 9 -	Gel de eletroforese dos produtos da ligação das ditags.....	79
Figura 10 -	Perfil eletroforético dos clones testados.....	81
Figura 11 -	Cromatogramas gerados por dois eventos de sequenciamento de um mesmo clone.....	83

Figura 12 -	Visualização do arquivo de saída do software SAGE2000™ (versão 4.5) ou <i>tag</i> list.....	84
Figura 13 -	Sequência ilustrativa de um EST utilizado na anotação de uma <i>tag</i>	86
Figura 14 -	Resultado ilustrativo gerado pelo Blast-X.....	87
Figura 15 -	Resultado ilustrativo para análise do EST associado a uma tag, apresentando a sequência completa da proteína gerada pelo Blast-X, e o domínio funcional obtido pelo programa CDD.....	88
Figura 16 -	Fichas “eletrônicas” do banco de dados de genômica funcional de <i>e. grandis</i> do Laboratório Max Feffer de Genética de Plantas.....	90
Figura 17 -	Distribuição do número de cópias para os 2.274 genes com número de cópias ≥ 2	94
Figura 18 -	Distribuição dos genes identificados a partir da biblioteca SAGE, de plantas de eucalipto com 6 meses de idade, em categorias funcionais.....	97
Figura 19 -	Distribuição dos genes pelo número total de transcritos identificados a partir da biblioteca SAGE em categorias funcionais.....	104
Figura 20 -	Reações de interconversões de açúcares nucleotídeos e biossíntese de hemicelulose, pectina e celulose.....	119
Figura 21 -	Direcionamento de sacarose para síntese de celulose.....	123
Figura 22 -	Biossíntese e polimerização dos monolignóis.....	126
Figura 23 -	Correlação entre o nível de expressão dos transcritos (número de cópias) e volume total de proteínas (%) para os 102 spots (48 proteínas) identificados por espectrometria de massas.....	141

LISTA DE TABELAS

Tabela 1 -	Otimização das reações de PCR a partir de diferentes concentrações da amostra.....	75
Tabela 2 -	Reação de ligação do vetor aos polímeros de <i>ditags</i>	80
Tabela 3 -	Valores de qualidade relacionados à probabilidade de erro de acordo com o programa <i>Phred</i> (EWING et al., 1998).....	83
Tabela 4 -	Classificação funcional adaptada a partir de Rison et al. (2000).....	89
Tabela 5 -	Resumo dos dados gerais produzidos pela biblioteca SAGE.....	94
Tabela 6 -	Lista dos 50 genes (<i>tags</i>) mais expressos na biblioteca SAGE de plantas de eucalipto com 6 meses de idade.....	99
Tabela 7 -	Distribuição dos genes identificados na biblioteca SAGE em categorias e subcategorias funcionais.....	105
Tabela 8 -	Proteínas identificadas a partir do caule de plantas de <i>E. grandis</i> com 6 meses de idade.....	138

ABREVIATURAS E SÍMBOLOS

λ - DNA do fago lambda (marcador de peso molecular)

μg - Micrograma

μL - Microlitro

μM - Micromolar

2D-PAGE - Sistema de eletroforese bidimensional

4CL - *4-coumarate:coenzyme A ligase*

A₂₆₀ ou A₂₈₀ - Absorbância a 260 ou 280 nm de comprimento de onda

AGP - *Arabinogalactan protein*

atm - Atmosfera

ATP - Adenosina tri-fosfato, molécula orgânica responsável pelo armazenamento de energia

BLAST - *Basic Local Alignment Search Tool*

BSA - Albumina sérica bovina

bulk - Uma porção de material que é representativo de todo o lote

β -(1-4) D-glicose - Cadeia linear de monômeros D-glicose unidos por ligações β -(1-4)

C3H - *p-coumarate 3-hydroxylase*

C4H - *Cinnamate 4-hydroxylase*

CAD - *Cinnamyl alcohol dehydrogenase*

CCoAOMT - *Caffeoyl-CoA O-methyltransferase*

CCR - *Cinnamoyl-CoA reductase*

CDD - Conserved Domain Database (Base de dados de domínios conservados)

cDNA - Sequência de DNA complementar ao mRNA gerada artificialmente por transcrição reversa

CESA - *Cellulose Synthase*

CIA - Solução contendo clorofórmio e álcool isoamílico

CO₂ - Dióxido de carbono

COMT - *Caffeic acid 3-O-methyltransferase*

contig - Conjunto de sequências de nucleotídeos obtidas no sequenciamento que se sobrepõem

CTAB - Detergente brometo de cetiltrimetilamônio

DEPC - Dietilpirocarbonato

DMSO - Dimetil sulfóxido

DNA - Ácido desoxirribonucléico

dNTP - Desoxirribonucleotídeo trifosfato

DTT - Ditionitreitol

E. coli - *Escherichia coli*

EDTA - Ácido etileno diamino tetracético

EG - *Endo-1,4-β-glucanase (Korrigan)*

EST - *Expressed Sequence Tag* - Sequências de cDNA, completas ou não, sintetizadas a partir de mRNA.

F5H - *Ferulic acid/coniferaldehyde 5-hydroxylase*

g - Grama

GAPDH - *Glyceraldehyde-3-phosphate dehydrogenase*

GLP - *Germin-like protein*

GS - *Glutamine synthetase*

GS/GOGAT – *Glutamine synthetase/Glutamate synthase*

H₂O - Molécula de água

H₂O₂ - Peróxido de hidrogênio

HCT - *Hydroxycinnamoyltransferase*

kb - Quilobase

kV - Quilo Volts

L - Litro

LC-MS/MS - Cromatografia líquida acoplada ao sistema de espectrometria de massas em *tandem*

LHCII - *Light-harvesting complex II*

LiCl - Cloreto de lítio

M - Molar

mg - Miligrama

MgCl₂ - Cloreto de magnésio

MilliQ - Água destilada deionizada e filtrada a 0,2μm

mL - Mililitro

mM - Milimolar

MPC - *Magnetic Particles Concentrator* ou concentrador de partículas magnéticas

mRNA - RNA mensageiro

ms - Milisegundos

m/z - razão massa/carga

NaCl - Cloreto de sódio

NAD⁺ - Nicotinamida adenina dinucleotídeo

NADPH - Nicotinamida adenina dinucleotídeo fosfato

NCBI - *National Center for Biotechnology Information*

O₂ - Molécula de oxigênio

°C - Medida de temperatura, grau centígrado

Oligo (dT) - Oligo desoxitimidina trifosfato (oligonucleotídeo poli-T)

ORF - *Open Reading Frame* ou fase aberta de leitura

ORF Finder - Ferramenta do NCBI capaz de identificar todas as fases abertas de leitura possíveis para uma dada sequência de nucleotídeos

p/v - Peso por volume

PAL - *Phenylalanine ammonia-lyase*

pb - Pares de bases

PCD - Morte celular programada (*Programmed Cell Death*)

PCR - Reação em cadeia da polimerase

PIB - Produto Interno Bruto

PIP - *Aquaporin*

PME - *Pectin methylesterases*

primer - Oligonucleotídeo iniciador

PVP - Polivinil pirrolidona

q.s.p. - Quantidade suficiente para

RCD - *Radical-induced Cell Death*

RNA - Ácido ribonucléico

RNA_r - RNA ribossômico

RNases - Enzimas que degradam o RNA

rpm - Rotações por minuto

RT-PCR - Síntese de cDNA por uma transcriptase reversa, seguida de PCR

SAGE - *Serial Analysis of Gene Expression*

SAM - *S-adenosylmethionine*

SAMS - *S-adenosylmethionine synthetase*

SDS - Dodecil sulfato de sódio

SUSY - *Sucrose synthase*

tag - Fragmento de DNA de 9-11 pares de base

Taq - DNA polimerase da bactéria *Thermus aquaticus*

T/G - Relação entre o número de transcritos e o número de genes

Tris - Tris-hidroximetil-aminometano

U - Unidades de enzima

UDP - Uridinodifosfato

UGDH - UDP-D-glicose desidrogenase

UXS - Proteína UDP-D-glucuronato descarboxilase

v/v - Volume por volume

Valor E - *E-value (Expected Value)* indica o número de alinhamentos que pode ser encontrado com uma pontuação maior ou igual à pontuação de alinhamento observada.

x g - Força centrífuga relativa à aceleração padrão de gravidade

XET - *Xyloglucan endotransglycosylase*

1 INTRODUÇÃO

As árvores possuem um papel fundamental no ambiente terrestre, fornecendo matéria prima e ajudando na conservação de inúmeros recursos naturais que são essenciais para a vida. O melhor entendimento de como as árvores desenvolvem o seu material lenhoso pode ter impactos diretos na economia mundial, uma vez que os produtos florestais, que movimentam bilhões de dólares, são manufaturados, vendidos e utilizados a cada ano. Além do seu alto valor econômico não se pode negar os efeitos ambientais da utilização das florestas plantadas, como por exemplo, o sequestro de carbono, reduzindo o efeito estufa e até mesmo a contribuição estética das árvores à paisagem natural (LI et al., 2006)

A madeira é o mais importante recurso natural de energia renovável, sendo uma alternativa ao uso de combustíveis fósseis. No comércio mundial, a madeira é o quinto produto em importância, devido à sua utilização na confecção de uma vasta gama de produtos (PLOMION et al., 2001). Dentre as espécies fornecedoras de madeira, as espécies de *Eucalyptus*, árvores tropicais de extraordinário crescimento, merecem grande destaque, uma vez que apresentam excelente potencial para serem utilizadas em diversos fins, principalmente na produção de papel e celulose. No Brasil, a produção de papel e celulose constitui um componente muito importante do setor industrial, com perspectivas de expansão para os próximos anos, tendo o país apresentado forte crescimento desse setor e tornando-se o principal produtor de celulose de fibra curta a partir do eucalipto como matéria prima.

Para as indústrias que utilizam a madeira como matéria-prima, o tecido de maior interesse é o xilema. Nas espécies perenes, a sucessiva adição de xilema secundário, diferenciado a partir do câmbio vascular, origina um tecido único chamado madeira (HAYGREEN et al., 1996). O processo de desenvolvimento do xilema secundário, conhecido como xilogênese, ocorre através de vários passos que envolvem o alongamento e alargamento radial de células derivadas do câmbio vascular; formação de uma parede celular espessa que se torna altamente lignificada e finalmente um processo de morte celular programada (CHAFFEY, 1999).

A maior parte das propriedades da madeira está sobre intenso controle genético (ZOBEL et al., 1995), e os conhecimentos sobre os mecanismos de controle molecular da formação da madeira são ainda muito pequenos.

Devido ao desenvolvimento recente dos projetos genomas em vários organismos, a informação sobre as sequências dos genes tem se acumulado. Com o sequenciamento dos genomas de *Arabidopsis thaliana*, *Oryza sativa* e a recente conclusão da primeira sequência do genoma de uma espécie arbórea (*Populus trichocarpa*) as atenções voltaram-se agora para a pesquisa genômica em *Eucalyptus*. A análise da expressão gênica compreende uma das partes mais importantes no estudo da função dos genes, e, métodos quantitativos de alta capacidade para a análise da expressão gênica são normalmente requeridos para esse fim.

A técnica de SAGE (*Serial Analysis of Gene Expression*), desenvolvida por Velculescu et al. (1995), permite uma análise quantitativa e precisa da expressão de centenas e até milhares de genes ao mesmo tempo, sendo, portanto, uma eficiente ferramenta para a realização de estudos de expressão gênica na era pós-genômica.

Mais informações sobre genes e proteínas são necessárias para aprofundar o entendimento das vias metabólicas envolvidas na fisiologia da madeira para criar ferramentas que poderão ser muito importantes na aproximação da biotecnologia aos anseios sobre as propriedades da madeira. A identificação de genes especificamente expressos no câmbio vascular altamente especializado ou tecidos do xilema diferenciados, pode fornecer informações substanciais a respeito dos processos moleculares envolvidos na xilogênese e também quais genes serão chaves na engenharia genética das propriedades da madeira. O conhecimento dos mecanismos moleculares e fisiológicos da atividade cambial e da formação da madeira é, portanto, considerado de fundamental importância, reunindo esforços de diversas áreas (PLOMION et al., 2001).

Pensando nisso, o Laboratório Max Feffer de Genética de Plantas iniciou uma série de projetos aplicados. A idéia geral é criar um banco de dados com informações do transcrito e do proteoma a partir de diferentes tecidos (madeira, folhas e raízes) de eucalipto em diferentes idades de crescimento e também submetido a diversos tipos de tratamentos (resistência a doenças, tolerância a seca, formação de madeira de

reação). A identificação e análise da expressão dos genes envolvidos nas vias metabólicas responsáveis pela formação e qualidade da madeira, bem como, na resistência ou tolerância a doenças e/ou adversidades ambientais em eucalipto, fornecerá informações valiosas a respeito da função desses genes, com possibilidades de manipulações futuras e seleção de novos clones, a fim de se obter árvores mais produtivas, com madeira de melhor qualidade e causando um menor impacto ao ambiente.

Dessa forma, o presente trabalho está inserido em um projeto maior de pesquisa que vem sendo desenvolvido no laboratório e teve como objetivo principal, utilizar a tecnologia SAGE para identificar e quantificar a expressão dos genes associados principalmente a xilogênese (divisão celular, expansão, formação da parede celular secundária, lignificação, e morte celular programada) no caule de plantas de *Eucalyptus grandis* com 6 meses de idade. Este projeto complementa outros projetos que estão analisando o padrão de expressão gênica durante a xilogênese na região cambial de árvores juvenis e adultas, e também outros projetos cujo objetivo é, através da espectrometria de massas, identificar as proteínas presentes nos mesmos tecidos e idades. Através desses projetos será possível associar o padrão de expressão gênica ao conjunto de proteínas presentes no mesmo tecido das plantas de *Eucalyptus grandis* e também será possível a construção de um banco de dados local com o depósito de todos os genes e proteínas identificadas. Conjuntamente, todos esses resultados poderão fornecer um panorama do padrão de expressão gênica durante a formação da madeira em diferentes idades de crescimento das árvores de eucalipto.

2 REVISÃO BIBLIOGRÁFICA

2.1 O Gênero *Eucalyptus* – origem e distribuição

O nome eucalipto deriva do grego: eu (= bem) e kalipto (= cobrir), referindo-se à estrutura globular arredondada de seu fruto, caracterizando o opérculo que protege bem as suas sementes. O eucalipto é principalmente nativo da Austrália, onde cobre praticamente 90% da área do país, formando densos maciços florestais nativos, sendo algumas poucas espécies nativas das ilhas de Java, Filipinas, Timor, Papua Nova Guiné e outras (PRYOR, 1976; ELDRIDGE et al., 1994; PRYOR et al., 1995; O EUCALIPTO, 2001a).

Pertencente à família das Mirtáceas, o gênero *Eucalyptus* engloba cerca de 700 espécies, além de subespécies, variedades e híbridos naturais, que podem ser encontrados desde a forma arbustiva até árvores de grande porte, estando amplamente distribuído em países de climas tropicais e subtropicais entre latitudes de 40° N e 45° S e altitudes que variam de 30 a 1000 metros, com exemplares adaptados a diversas condições de clima e solo. (CRESSWELL et al., 1985; BROOKER, 2000).

As espécies do gênero *Eucalyptus* são monóicas, protândricas e preferencialmente alógamas (apresentam em média 30% de autogamia) (PRYOR, 1976) sendo a autogamia maior em populações naturais do que em populações melhoradas, segundo Mori¹ (informação verbal). A polinização é principalmente entomofílica, e as abelhas (*Apis mellifera* L.) são os principais polinizadores (PACHECO, 1982), podendo ocorrer também a polinização pelo vento e por pássaros (SAMPSON et al., 1995).

A análise do cariótipo revelou que grande parte das espécies do gênero são diplóides, com número básico de $2n=22$ cromossomos de tamanho extremamente pequeno, como a maioria dos gêneros da família Mirtácea. Estimativas do conteúdo de DNA nuclear revelaram que esse pode variar entre 370 e 700 milhões de pares de base, sendo que a maior parte das espécies possui em média 650 milhões de pares de bases (ELDRIDGE et al., 1993; GRATTAPAGLIA; BRADSHAW, 1994).

¹ MORI, E.S. (Departamento de Ciências Florestais/UNESP), 06 de fevereiro de 2007, informação verbal durante defesa de tese na ESALQ/USP.

Os historiadores mencionam que, por volta de 1774, o eucalipto foi introduzido na Europa por Antônio Guichenot e, em 1788, foi descrito pela primeira vez pelo botânico francês L'Héritier de Brutelle, no *Sertum Anglicum*, em Paris, valendo-se do material recolhido em expedições no território australiano. Até a metade do século XIX, o eucalipto figurou apenas em coleções de alguns jardins botânicos, sem nenhuma importância comercial. Na Índia, os primeiros plantios ocorreram em 1843 e, por volta de 1856, já havia plantações de eucaliptos bem desenvolvidas. Na África do Sul, os primeiros ensaios com eucaliptos ocorrem em 1828, na colônia do Cabo. Os primeiros ensaios na Europa, visando à produção comercial, datam de 1854, principalmente com o *Eucalyptus globulus*, plantados em Portugal pelo Barão de Massarellos e por J. M. Eugênio de Almeida. Em 1863 foi introduzido na Espanha e, em 1869 na Itália (O EUCALIPTO, 2001a).

Na América do Sul, talvez o Chile tenha sido o primeiro país a introduzir o eucalipto em 1823, recebendo as sementes de um navio inglês. A Argentina teria introduzido o eucalipto em seu país em 1865, através do Presidente Garcia Moreno. No Uruguai, as primeiras sementes de eucalipto foram recebidas em 1853. É difícil se determinar, com segurança, a data de introdução do eucalipto no Brasil. Até algum tempo, tinha-se como certo que os primeiros plantios aconteceram no Rio Grande do Sul, em 1868, por Frederico de Albuquerque, mas hoje alguns historiadores acreditam que os exemplares mais antigos encontrados no país são de *Eucalyptus robusta* e *Eucalyptus tereticornis*, plantados pelo Imperador D. Pedro I no Jardim Botânico do Rio de Janeiro, em 1825 (O EUCALIPTO, 2001a).

2.2 O setor florestal brasileiro e a importância do eucalipto

A madeira, material encontrado unicamente nos vegetais superiores, constitui-se no quinto maior produto comercial do mundo devido a sua enorme gama de utilizações, como por exemplo, produção de combustível, papel e celulose, construção civil, e muitos outros. O complexo químico que constitui a madeira também faz dela um material ideal, para no futuro, substituir a indústria petroquímica na produção de plástico e outros tipos de produtos químicos, podendo também ser utilizada na indústria de alimentos e produtos têxteis (PLOMION et al., 2001).

As espécies de eucalipto apresentam excelente potencial para produção de madeira em função de sua diversidade, adaptabilidade, alta produtividade e características físico-mecânicas que permitem inúmeras utilizações como matéria-prima para marcenaria (serraria e madeira processada), estruturas para construção civil, moirões, postes, lenha e carvão vegetal, indústria química e farmacêutica, apicultura e ornamentação, e principalmente para a produção de papel e celulose (GONZÁLEZ, 2002).

O rápido crescimento (média de 45 m³ ha⁻¹ por ano em plantios clonais) associado à produção de fibra e polpa de alta qualidade, baixo custo sobre curto período (5-7 anos) em um regime que permite até 3 rotações sucessivas e econômicas com ciclo de até 21 anos, são as principais razões do extensivo uso do eucalipto em reflorestamentos comerciais em várias partes do mundo (HO et al., 1998).

Segundo levantamentos da FAO (FAO, 2000), mais de noventa países possuem programas de reflorestamento com eucalipto. A área total reflorestada com o eucalipto no mundo equivale a 18 milhões de hectares com grandes plantios comerciais na Índia, Brasil, África do Sul, Portugal e Espanha. Desses 18 milhões de hectares, 3,4 milhões estão no Brasil (SOCIEDADE BRASILEIRA DE SILVICULTURA - SBS, 2006), sendo os estados de Minas Gerais, São Paulo, Bahia e Espírito Santo, os que apresentam as maiores áreas plantadas (MORA; GARCIA, 2000). As espécies mais utilizadas em nosso país são: *Eucalyptus grandis* (55%), *Eucalyptus saligna* (17%), *Eucalyptus urophylla* (9%), *Eucalyptus viminalis* (2%), híbridos de *E. grandis* x *E. urophylla* (11%) e outras espécies (6%) (O EUCALIPTO, 2001b).

No Brasil, o plantio de eucalipto 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 tratamentos culturais de manejo e melhoramento, adequação aos mais diferentes usos e aceitação de mercado (MORA; GARCIA, 2000; SILVA, 2005).

O setor florestal brasileiro que é principalmente dividido nos segmentos de papel e celulose, siderurgia e carvão vegetal, e indústria moveleira, apresentou em 2005 um faturamento de US\$ 27,8 bilhões (3,5 % do PIB nacional) e contribuiu com US\$ 4,2 bilhões em impostos. As exportações brasileiras ligadas ao setor atingiram US\$ 7,4

bilhões, correspondendo a aproximadamente 7% do total exportado pelo país, gerando ainda cerca de 3 milhões de empregos diretos e indiretos (SBS, 2006; LEITE, 2005).

A produção de papel e celulose constitui um componente muito importante do setor industrial, com perspectivas de expansão para os próximos anos, tendo o país apresentado forte crescimento desse setor. Hoje, o Brasil ocupa a sétima posição mundial na produção de celulose e a 11ª como fabricante de papel. Além disso, nos últimos 10 anos, investimentos da ordem de US\$ 12 bilhões feitos pelo Setor de Papel e Celulose para a ampliação da capacidade produtiva, promoveram um extraordinário aumento na produção de celulose, tornando o Brasil o maior produtor mundial de celulose de fibra curta, utilizando o eucalipto como matéria prima (SBS, 2006). Dessa forma, o setor brasileiro de papel e celulose contribui de forma relevante para o desenvolvimento do país em termos de geração de renda, tributos, empregos e divisas, bem como, de promoção do desenvolvimento regional (BRACELPA, 2006).

Em 2007, a produção brasileira de celulose deverá alcançar 11,75 milhões de toneladas e a de papel, 9,0 milhões de toneladas, registrando um crescimento sobre 2005 de 5,9% e 2,9%, respectivamente, o que levará o Brasil a passar da sétima para a sexta posição como produtor mundial de celulose. (BRACELPA, 2006).

Segundo estimativas, em cinco décadas, a população mundial passará de 6 para 9,5 bilhões de habitantes e o consumo de madeira subirá quase na mesma proporção do crescimento populacional. De acordo com pesquisas internacionais recentes, o consumo médio mundial de madeira deve aumentar a uma taxa de 1,2 a 3,4% ao ano (PERSPECTIVAS, 2003). Atento a isso, o setor de papel e celulose está dando continuidade a seu programa de investimento, no valor de US\$ 14,4 bilhões, para o período 2003 – 2012, cujo objetivo é ampliar a capacidade produtiva e a competitividade das indústrias brasileiras de papel e celulose, assegurando assim, o pleno atendimento do mercado interno, aumentando as exportações (principalmente de celulose) e criando novas oportunidades de trabalho (BRACELPA, 2006).

Do ponto de vista ambiental, o uso de madeira reflorestada contribui significativamente para a proteção e conservação dos recursos naturais existentes, preservando-se assim, áreas de mata nativa já escassas e prevenindo-se também a extinção de espécies consideradas nobres, geralmente de madeira dura e crescimento

muito lento (FERREIRA, 1979). Atualmente o mundo atravessa um período de mudanças climáticas e ambientais. Uma das principais causas apontadas pelos cientistas é o chamado efeito estufa, o aumento gradativo da concentração de carbono na atmosfera, decorrente principalmente da queima de combustíveis fósseis. Reflorestar ajuda a sequestrar esse excesso de carbono da atmosfera, fato reconhecido e valorizado pelo Protocolo de Kyoto, criando os chamados “mecanismos para o desenvolvimento limpo”, sendo os créditos de carbono um desses mecanismos. Segundo esse protocolo, os países intensamente industrializados devem diminuir seus níveis de poluição na atmosfera. Caso as metas não sejam atingidas através da redução da emissão dos gases poluentes, estes poderão conseguí-la por meio da compra de créditos de carbono de países que não têm a mesma obrigatoriedade, como é o caso do Brasil. Assim, o país deve se colocar como um grande vendedor de créditos de carbono e como alvo de investimentos em projetos engajados com a redução da emissão de poluentes. Dessa forma, espera-se que o faturamento decorrente da venda desses créditos alcance US\$ 13 bilhões em 2007 (BRASIL, 2005).

Diante de toda essa importância econômica e também ambiental, o desafio para o setor de papel e celulose nos próximos anos, não está apenas em investimentos na área industrial, mas também de aumento da produtividade e qualidade. Isso deve ser feito aliando-se o melhoramento convencional as técnicas de biotecnologia.

2.3 Melhoramento genético e a biotecnologia aplicada ao eucalipto

Até o início do século XX, o eucalipto era plantado como árvore decorativa, pelo seu extraordinário desenvolvimento como quebra-vento ou por supostas propriedades sanitárias. Pouquíssimas foram as plantações com fins industriais e caráter florestal. Deve-se à Companhia Paulista de Estradas de Ferro e ao Dr. Edmundo Navarro de Andrade o início da silvicultura intensiva e a expansão da eucaliptocultura no Brasil. Os primeiros plantios de eucalipto destinavam-se 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 (Rio Claro - S.P.) no mais completo banco de germoplasma de eucalipto no Brasil (O EUCALIPTO, 2001a; ANDRADE, 1961).

Em função da grande demanda de madeira para futuros projetos industriais e ao esgotamento das matas nativas, o governo brasileiro instituiu, a partir de 1967, um programa de incentivos fiscais ao setor de reflorestamento para aumentar a área plantada especialmente com espécies exóticas de rápido crescimento. Em poucos anos, a área com plantações de eucaliptos saltou de 400 mil para 3 milhões de hectares. Nessa época foram criados cursos de engenharia florestal e várias instituições de pesquisa florestal, que deram suporte ao desenvolvimento da cultura. No final da década de 70 e durante os anos 80, intensificaram-se estudos sobre a qualidade fisiológica e genética das sementes e, foram instalados os primeiros testes de progênes e pomares de sementes para atender a demanda crescente de sementes no mercado (O EUCALIPTO, 2001a).

Dessa maneira, o melhoramento florestal, incluindo o melhoramento do eucalipto, é uma ciência relativamente nova no Brasil, incorporando conceitos aplicados ao melhoramento animal e ao melhoramento de culturas agrícolas anuais, bem como algumas particularidades desenvolvidas ao longo dos anos em decorrência das exigências das espécies florestais (RESENDE, 1999). No início, os programas de melhoramento genético de eucalipto tinham como objetivo determinar a adaptabilidade das espécies e de diferentes genótipos em ambientes específicos por meio de ensaios de progênes em diferentes locais, compreender a interação genótipo x ambiente e, dessa forma, analisar a viabilidade dos diferentes genótipos em diferentes condições ambientais (KAGEYAMA, 1980). O passo seguinte seria aumentar a produtividade volumétrica da madeira a cada ciclo de seleção, sem comprometer com isso a base genética da população (MENK, 1989).

Com o fim do subsídio para o setor florestal em 1986, as pesquisas básicas deixam de ser prioridade e o enfoque passa a ser a busca pela árvore superior, como unidade de propagação clonal, baseado principalmente nas técnicas de hibridação e clonagem (ODA et al., 2007).

A hibridação interespecífica tornou-se a forma mais rápida e eficiente de obtenção de ganhos genéticos no melhoramento de espécies de *Eucalyptus*, com reflexos diretos e significativos na indústria de base florestal, uma vez que, a grande diversidade interespecífica deste gênero facilita a busca de combinações desejáveis

nos programas de melhoramento. A busca de complementaridade nas características tecnológicas da madeira, a tolerância a estresses bióticos e abióticos, bem como a manifestação da heterose, foi verificada em vários cruzamentos híbridos, como por exemplo, o híbrido interespecífico *E. urograndis* (*E. grandis* x *E. urophylla*) (ASSIS; MAFIA, 2007).

Num segundo momento, o emprego da clonagem em vez dos métodos sexuais de produção de mudas, permitiu a transferência da variância genética total, resultando na obtenção de ganhos máximos, sejam de produtividade volumétrica, sejam relacionados a propriedades tecnológicas da madeira ou resistência a fatores bióticos e abióticos. A clonagem possibilitou ainda, a produção de matéria-prima mais uniforme que, do ponto de vista industrial, representa significativos benefícios tanto na maximização dos custos do processo industrial, quanto na qualidade dos produtos. Dessa forma, técnicas de clonagem por micropropagação, como a microestaquia e a miniestaquia, ganharam destaque e, milhares de árvores são selecionadas nas populações existentes e propagadas vegetativamente, para serem incluídas nos jardins clonais, testes clonais e plantios comerciais clonais (ODA et al., 2007; ASSIS; MAFIA, 2007).

O melhoramento genético sexual tornou-se, portanto, a base para fornecer materiais genéticos melhorados em gerações avançadas de melhoramento para propagação clonal. A seleção desses materiais genéticos superiores, em programas de melhoramento florestal, passou a ser feita visando às características silviculturais de interesse como volume de madeira, forma do tronco da árvore, qualidade da madeira para usos múltiplos e resistência a doenças. Já, características do aspecto reprodutivo, como capacidade e periodicidade de florescimento e produção de sementes, dificilmente são considerados na seleção (ODA et al., 2007).

O termo “qualidade da madeira” surge somente nos anos 90, quando os setores florestais passam a questionar as características desejáveis da madeira em si, em função do seu uso final e dos métodos de processamento a serem utilizados, para dessa forma, definir as características que deveriam ser buscadas e avaliadas nos programas de melhoramento (RAYMOND, 2002; POT et al., 2002). Assim, para a indústria siderúrgica, as características mais importantes da madeira são aquelas que

afetam positivamente a qualidade do carvão, tais como elevadas densidades da madeira e alto teor de lignina as quais somam maior poder calorífero ao carvão tornando-o mais eficiente no processo de fusão do minério de ferro (BRITO; BARRICHELO, 1977; TRUGILHO et al., 2001). Para a indústria de móveis, que utiliza produtos sólidos, a madeira deve apresentar propriedades físicas e mecânicas satisfatórias, dessa forma, 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, extrativos e umidade, visando diminuir o custo de produção através de um maior rendimento durante o processo de polpação, bem como, aumentar a qualidade do produto final (POT et al., 2002; ASSIS, 1996).

Não se pode negar a grande contribuição do melhoramento genético e da silvicultura para a formação de florestas altamente produtivas e com caracteres específicos desejáveis ao uso final, principalmente para as florestas de eucalipto. No entanto, os programas de melhoramento genéticos tradicionais de espécies florestais são dificultados por características inerentes a biologia das árvores, como por exemplo: longo ciclo de vida das espécies; porte das árvores; maturação tardia; polinização cruzada; elevada heterozigosidade; alta complexidade da análise dos descendentes após cruzamentos e retrocruzamentos; dificuldade em conseguir ganhos significativos para caracteres complexos, como propriedades da madeira, controle de pragas e doenças, e tolerância a estresses abióticos; além da necessidade de grandes áreas para os plantios experimentais (VALVERDE et al., 2007; CAMPBELL et al., 2003).

Neste cenário, a biotecnologia florestal surge como uma alternativa potencial para alcançar progressos de forma rápida, manipular e estudar caracteres até então desconhecidos ou que não podiam ser manipulados através do emprego das técnicas de melhoramento convencional. As técnicas da biotecnologia aliadas ao melhoramento convencional podem contribuir significativamente para um aumento no desenvolvimento vegetativo e biomassa das plantas. Atualmente, o grande desafio dos programas de melhoramento de eucalipto é melhorar a qualidade da madeira, embora o interesse por ganhos em volume continue elevado. A melhoria da qualidade da madeira afeta a

quantidade do produto comercial a ser obtida por unidade de área e, também permite a diferenciação de produtos de forma a melhorar a competitividade das empresas no mercado (CHAFFEY, 1999).

Uma importante ferramenta da biotecnologia para auxiliar no melhoramento clássico de espécies arbóreas é a seleção assistida por marcadores (SAM), que utiliza marcadores moleculares para alelos relacionados ao gene alvo facilitando a seleção de genótipos superiores, por permitir a seleção precoce de caracteres que normalmente são identificados na fase adulta reduzindo assim, o tempo para obtenção desses indivíduos superiores (FERREIRA; GRATTAPAGLIA, 1995).

Os marcadores genéticos vêm sendo utilizados em espécies florestais desde os anos 80, quando isoenzimas permitiram realizar os primeiros sistemas de cruzamento em pomares de sementes e produzir as primeiras versões de mapas genéticos em coníferas. Os mapas genéticos têm se revelado como um recurso de grande valor, pois permitem a localização de genes e regiões genômicas envolvidas no controle de características quantitativas (QTL), sem a necessidade de conhecimento prévio da característica. Assim, através da análise dos QTLs, regiões do genoma que afetam a variação de uma característica quantitativa podem ser identificadas, fornecendo mapas de ligação que permitem uma representação do genoma bem como o estudo do controle genético desses caracteres quantitativos (GRATTAPAGLIA, 2007; POKE et al., 2005).

A partir da década de 90 então, iniciou-se a construção de uma série de mapas genéticos para espécies dos gêneros florestais *Pinus*, *Populus* e *Eucalyptus*, através da combinação de marcadores dominantes RAPD e AFLP (RITTER et al., 1990; GRATTAPAGLIA; SEDEROFF, 1994). Em seguida, a técnica passou a ser complementada com informações obtidas a partir de marcadores co-dominantes RFLP (BYRNE et al., 1995), marcadores microsatélites (BRONDANI et al., 1998) e mais recentemente, com marcadores de polimorfismos de base individual (SNPs) (BROWN et al., 2004).

O que tem se buscado hoje, é a integração dos dados fornecidos pelos mapas genéticos com os dados de expressão gênica em larga escala, de modo a compreender como a variabilidade em um nível leva a mudanças em todos os outros, e

principalmente à variabilidade fenotípica. Jansen e Nap (2001), em uma revisão de diversos estudos, propuseram uma possibilidade para tal integração. Segundo os autores, os dados de expressão gênica coletados a partir de uma população segregante poderiam ser analisados por meio do mapeamento de QTLs. Surge então, os QTLs de expressão (eQTLs), que como o mapeamento de QTLs tradicional define regiões genômicas que controlam características fenotípicas. O mapeamento de eQTLs identifica regiões que controlam o nível de expressão dos genes, sendo assim, a análise de eQTL é capaz de distinguir se a regulação da expressão gênica é do tipo *cis* ou *trans*², permite a identificação de agrupamentos de eQTLs (eQTL *hotspots*) e fornece uma oportunidade real para se entender a complexa rede de regulação gênica e epistasia (POKE et al., 2005; KIRST, 2007). Kirst e colaboradores (2004) foram os primeiros a demonstrar tal aplicação em eucalipto.

Outra importante alternativa da biotecnologia para auxiliar no melhoramento clássico das espécies arbóreas é a transformação genética. Muito mais do que um método para avaliar a função de genes em plantas, a transgenia tem permitido a geração de cultivares com características superiores dos pontos de vista agrônomo/silvícola, nutricional ou industrial. Essa técnica permite a inserção de genes previamente caracterizados de uma espécie em outra espécie e a subsequente regeneração de plantas férteis através da cultura *in vitro* de tecidos. As grandes vantagens da transgenia são a redução do tempo necessário para a introdução de novas características de interesse em linhagens vegetais já estabelecidas por longos ciclos de cruzamento e seleção e, também, a quebra da barreira sexual entre diferentes espécies (PASQUALI; ZANETTINI, 2007).

Diversas espécies arbóreas tem sido utilizadas como modelos para a avaliação de respostas a modificações genéticas. Entre elas, o gênero *Populus* que possui grande importância econômica para a indústria de base florestal de diversos países europeus de clima temperado com sólida estrutura de pesquisa (França, Alemanha, Bélgica e Itália), destaca-se por ter sido a primeira árvore a ser cultivada e regenerada *in vitro* já no final dos anos 60, assim como, pela facilidade de ser transformada geneticamente (FILLATTI et al., 1987). Devido ao pequeno tamanho do seu genoma (cerca de 500

² Regulação *cis*: posição do eQTL é a mesma do gene. Regulação *trans*: posição do eQTL não coincide com a do gene.

Mb), também foi a primeira espécie florestal a ter o genoma totalmente sequenciado (TUSKAN et al., 2006).

Apesar da grande superioridade econômica do eucalipto, a recalcitrância à transformação genética ainda é um fator limitante quando comparado ao *Populus*. Dessa forma, apesar dos esforços de vários grupos de pesquisa do mundo todo, poucos trabalhos relatam a obtenção de plantas transgênicas de eucalipto (PASQUALI; ZANETTINI, 2007). Serrano et al (1996) utilizando o sistema de transformação por biolística obtiveram embriões zigóticos de *E. camaldulensis* transformados, porém não conseguiram recuperar plantas transformadas. Mullins et al. (1997) e Ho et al. (1998), utilizando o sistema de transformação genética via *Agrobacterium tumefaciens*, obtiveram plantas de *E. camaldulensis* transformadas a partir de explantes de folha e hipocótilos, respectivamente. Moralejo et al. (1998), a partir de ferimentos produzidos por bombardeamento seguido de transformação por *A. tumefaciens*, obtiveram plantas transgênicas de *E. globulus*. Somente em 2002, González e colaboradores (2002) utilizando um sistema de transformação por *A. tumefaciens*, assistida por sonicação, conseguiram obter plantas transgênicas para o híbrido *E. grandis* x *E. urophylla* que possui grande interesse comercial no Brasil.

Recentemente, muitas pesquisas têm sido realizadas para identificação de genes envolvidos na formação da madeira, ou seja, genes relacionados à síntese de celulose e lignina durante a formação da parede secundária. Seguindo essa linha, Jouanin e colaboradores (2000), obtiveram plantas de *Populus* transgênicas através da superexpressão de um transgene que codifica a enzima *Caffeic acid 3-O-methyltransferase* (COMT). Em uma das linhagens transgênicas contendo duas cópias do promotor utilizado (35S), a atividade da COMT foi próxima a zero graças ao fenômeno de silenciamento gênico. Essa mudança levou a uma redução do nível de lignina em 17% e alterou fortemente sua estrutura, levando a um aumento no conteúdo de celulose e afetando positivamente em cerca de 10% a produção de polpa na produção de pasta celulósica. Também, através da transgenia, VALÉRIO et al. (2003) obtiveram plantas de eucalipto antisense para o gene que codifica a enzima *Cinnamyl alcohol dehydrogenase* (CAD). Embora as plantas transgênicas recuperadas tenham exibido forte redução no nível de transcritos para a enzima, as análises químicas, de

composição e de polpação da madeira dessas plantas com 10 meses de idade, não apresentaram diferenças no conteúdo de lignina em relação às plantas controles como era esperado, sugerindo que essa rota metabólica seja altamente regulada.

Como o principal destino da madeira de eucalipto é a indústria de papel e celulose, vem-se buscando melhoria na qualidade da madeira para atender tal finalidade e, uma boa forma de se alcançar sucesso é através da transgenia, com a geração de plantas com diferentes composições e arquitetura da parede celular. Dessa forma, a parede celular está se tornando alvo-chave no melhoramento de plantas superiores (CARPITA et al., 2001). A principal consequência da redução, por exemplo, da quantidade e composição da lignina em árvores é a diminuição do uso de produtos químicos empregados durante o processo de sua extração na produção de papel e celulose, reduzindo custos e minimizando os impactos ambientais (BAUCHER et al., 2003).

2.4 Genômica

A genômica pode ser definida como a parte da biologia molecular responsável pelo estudo dos genes (genoma) dos organismos vivos. Em particular, a genômica se ocupa da investigação da estrutura, conteúdo, funções e evolução dos genomas, tendo como base, métodos de clonagem de genes e o sequenciamento de DNA (BINNECK, 2004).

A caracterização dos genomas foi uma das principais forças motrizes da ciência nos anos 90. Desde o sequenciamento completo do primeiro organismo de vida livre em 1995 (FLEISCHMANN et al., 1995), a inovação e consolidação das técnicas de análise genômica tornaram possível o sequenciamento do genoma de seres complexos, como plantas, animais e seres humanos.

Como resultados dos crescentes investimentos na área, a lista de sequências de genomas completos vem crescendo a uma velocidade cada vez maior e contribuindo com um volume de dados disponíveis para acesso público sem precedentes na história (BINNECK, 2004). Por exemplo, até março de 2007 já haviam sido publicados 525 genomas completos, dos quais 440 referiam-se a genomas de bactérias, 37 de *Archea*

e 48 de eucariotos. Além disso, haviam 2468 projetos genomas em andamento, sendo 1091 de bactérias, 59 de *Archea* e 720 de eucariotos (<http://www.genomesonline.org/>).

No caso de espécies arbóreas, o primeiro genoma completo sequenciado foi concluído em 2006 para *Populus trichocarpa* (TUSKAN et al., 2006). No caso do eucalipto, já está disponível o sequenciamento do genoma do cloroplasto de *E. globulus* (STEANE, 2005), e a partir de 2003, através de uma associação internacional de pesquisadores, universidades e indústrias, foi criado o *Eucalyptus* Genome Network - EUCAGEN (www.ieugc.up.ac.za), para realizar o sequenciamento do genoma do eucalipto, utilizando espécies de interesse comercial (*E. globulus*, *E. camaldulensis* e *E. grandis*). Em meados de 2006, a proposta de sequenciamento foi enviada ao *Joint Genome Initiative* (JGI) do Departamento de Energia dos EUA e acredita-se que até o final de 2007 já tenha sido gerada uma sequência preliminar.

Com a enorme quantidade de dados gerados, acreditava-se que a genômica seria a solução para todos os problemas, mas na verdade, a determinação da sequência completa de vários genomas não é o final da história. Na realidade, o conhecimento sobre os genes é apenas o primeiro passo, sendo necessária ainda uma compreensão mais aprofundada dos processos que ocorrem após a duplicação, a transcrição e a tradução do material genético. A chamada “era pós-genômica” começou com a constatação de que os dados trazidos pelo sequenciamento das moléculas de DNA, embora relevantes, são limitados, sendo imprescindível investigar tanto os processos de transcrição das informações contidas no genoma através da transcritômica, quanto os seus produtos, as proteínas, através da proteômica e até mesmo os metabólitos através da metabolômica, para dessa forma estabelecer um perfil funcional da célula e/ou tecido, fornecendo uma visão global do funcionamento dos organismos e as interações que levam um genótipo a expressar um fenótipo em determinadas condições (PIMENTA, 2003).

2.5 Transcritômica

A Transcritômica (genômica funcional) é o estudo do conjunto de transcritos expressos em uma célula ou tecido, em um dado momento, representando o primeiro passo na caracterização funcional dos genes e identificação de grupos de genes

funcionalmente associados. A análise detalhada do transcrito pode tornar possível a identificação e a dissecação das complexas redes metabólicas a que estão sujeitos os processos fisiológicos, desenvolvimento e resposta a estímulos (ALBA et al., 2004).

Após a identificação dos genes, a prioridade passa a ser o estabelecimento dos padrões de expressão e funções. Apesar de serem bem definidos, sensíveis e robustos, os métodos tradicionais de análise da expressão gênica, como, *Northern Blot*, Hibridização *in situ*, PCR (*Polymerase Chain Reaction*) ou RT-PCR (*Reverse Polymerase Chain Reaction*), são adequados para a análise de um número pequeno de genes. O estabelecimento dos perfis transcricionais proporciona um enorme avanço, pois permite uma análise eficiente em alta velocidade (*high-throughput*) de um grande número de genes ao mesmo tempo (DONSON et al., 2002; MEYERS et al., 2004).

As metodologias utilizadas para estabelecer perfis transcricionais baseiam-se na geração de populações de cDNA a partir da população de RNA mensageiro (mRNA) presente. Essas metodologias podem ser divididas em métodos globais (abertos) e métodos limitados (fechados) de análise. Os métodos globais permitem o acesso à potencialmente todos transcritos expressos em um determinado momento porque não requerem informações prévias de sequências 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 (ALBA et al., 2004).

Atualmente, existem diferentes metodologias que permitem uma análise global do transcrito, entre as mais utilizadas estão o sequenciamento de ESTs (ADAMS et al., 1991), *Differential display* (DD) (LIANG; PARDEE, 1992), *Serial analysis of gene expression* (SAGE) (VELCULESCU et al., 1995), *Amplified restriction fragment polymorphism* baseado em cDNA (cDNA-AFLP) (VOS, et al., 1995; BACHEM et al., 1996), *Suppression subtractive hybridization* (SSH) (DIATCHENKO et al., 1996) e *Massively parallel signature sequencing* (MPSS) (BRENNER et al., 2000).

Devido à existência de grandes regiões não codificadoras no DNA, tem-se utilizado, como alternativa, sequenciar apenas as regiões codificadoras, dando origem a bancos de etiquetas de sequências expressas (ESTs - *Expressed Sequence Tags*). A base do método consiste no fato de que o nível de mRNA tecido-específico presente

em uma espécie é refletido pela frequência da ocorrência de seu EST correspondente em uma biblioteca de cDNA (ADAMS et al., 1991)

A partir do conceito original de ESTs proposto por Adams et al. (1991), para ajudar na identificação de genes diferencialmente expressos em um tecido ou condição específica, inicia-se uma nova era no campo da genética de árvores, seguindo o exemplo de outras plantas como *Arabidopsis*, milho e arroz. Surgem em 1998, os dois primeiros estudos descrevendo 5.692 ESTs de *Populus* (STERKY et al., 1998) e 1.000 ESTs de *Pinus* (ALLONA et al., 1998). Desde então, o número de ESTs de árvores tem aumentado progressivamente, sendo que atualmente já encontram-se disponíveis nos bancos de dados públicos (<http://www.ncbi.nlm.nih.gov/dbEST>) mais de 359 mil ESTs de *Pinus sp.* e mais de 369 mil ESTs de *Populus sp.*

No caso do eucalipto, foram conduzidos no Brasil, através de iniciativas de diversas empresas do setor florestal juntamente com universidades e órgãos nacionais de fomento à pesquisa, dois projetos de sequenciamento de ESTs de diferentes espécies e tecidos. O Projeto *Forests* que resultou na construção de um banco de dados com 123.889 ESTs e o Projeto *Genolyptus* que resultou no sequenciamento de cerca de 135.000 ESTs (POKE et al., 2005). Porém, infelizmente, ambos os projetos não disponibilizaram publicamente as sequências, inviabilizando pesquisas mais avançadas e com futuro promissor. Graças a grupos estrangeiros, hoje existem cerca de 15 mil ESTs de *E. globulus*, *E. camaldulensis* e *E. grandis* depositados em bancos públicos, no entanto, a maior parte dos ESTs produzidos por esses grupos, ainda não foram disponibilizados publicamente.

Embora o sequenciamento de ESTs represente um método rápido para identificar genes expressos, fornecendo informações importantes para a construção de plataformas de análise baseadas em arranjos (*arrays*), ele não substitui completamente o sequenciamento genômico, uma vez que, genes que apresentam baixa expressão são dificilmente representados em bibliotecas de ESTs (SUN et al., 2004; BHALERAO et al., 2003).

A técnica de *Differential display* (DD) (LIANG; PARDEE, 1992) é baseada em métodos simples, necessita de pouco material inicial e permite que duas ou mais amostras sejam comparadas simultaneamente. Basicamente a DD consiste na

transcrição reversa do mRNA com *primers* oligo(dT), amplificação do DNA com *primers* curtos 5' arbitrários, marcação radioativa dos cDNAs amplificados e eletroforese em gel de poliacrilamida, isolamento dos fragmentos separados no gel, amplificação e sequenciamento. Apesar de permitir a comparação de um grande número de transcritos entre populações de interesse, e ter sido muito utilizada em estudos que envolvem eventos fisiológicos, transdução de sinais, resposta a estresse e metabolismo secundário, a dificuldade de se acessar todos os transcritos e o uso de *primers* pequenos e arbitrários torna muito difícil a confirmação das sequências encontradas. (DONSON et al., 2002; STEIN; LIANG, 2002; YAMAZAKI; SAITO, 2002; MEYERS et al., 2004)

Devido a essas dificuldades, outras técnicas de análise da expressão gênica podem ser utilizadas para complementar ou mesmo substituir a técnica de DD. Uma delas é a *Amplified restriction fragment polymorphism* baseado em cDNA (cDNA-AFLP) (VOS, et al, 1995; BACHEM et al., 1996). Nesse caso, uma população de cDNA é sintetizada, cortada com enzimas de restrição e ligada a adaptadores com sítios para amplificação por PCR de alta estringência e seletividade (devido ao uso de adaptadores e pela adição de uma ou mais bases aleatórias ao *primer*). Em seguida, a amostra é submetida à eletroforese em gel de poliacrilamida e as bandas de interesse são excisadas e sequenciadas (BACHEM et al., 1996, ALBA et al., 2004). A técnica de cDNA-AFLP tem sido bastante utilizada para identificar genes diferencialmente expressos em uma larga variedade de tecidos e estádios de desenvolvimento em plantas (van RAEMDONCK et al., 2005; RANIK, et al., 2006). Embora a técnica permita acesso a uma grande quantidade de transcritos, com maior precisão e confiabilidade do que a DD, essa possui pequena sensibilidade na identificação de sequências com baixa frequência (BACHEM et al., 1996, ALBA et al., 2004).

Desenvolvido em 1996, o método de *Supression subtractive hybridization* (SSH) permite amplificar de maneira seletiva fragmentos de cDNA alvo 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 a produção de híbridos de cDNA diferencialmente expressos entre duas amostras. As bibliotecas são formadas somente após uma etapa de hibridização entre a amostra e o controle.

Somente os cDNAs que não hibridizaram no primeiro passo e que são portanto diferencialmente expressos entre as duas bibliotecas, estarão disponíveis para hibridização e seleção no segundo passo, onde uma nova quantidade de cDNA desnaturado da amostra é fornecida para hibridização. Como as populações de cDNA tanto da amostra quanto do controle são ligadas a adaptadores com sítios de anelamento para *primers* específicos no início do processo, os híbridos-alvos formados podem ser especificamente amplificados através de PCR, clonados e sequenciados (DIATCHENKO *et al.*, 1996; STEIN; LIANG, 2002). Apesar de ser um método que permite com grande eficiência a detecção de transcritos expressos tanto em baixo quanto em alto nível de expressão, o SSH apresenta algumas desvantagens como protocolo tecnicamente trabalhoso, só podem ser conduzidos experimentos para comparação “*one-way*” e pequenas diferenças podem ser mascaradas. Atualmente, o SSH tem sido utilizado para estudar diferenças na abundância de cDNA tanto em procariotos quanto em eucariotos (STEIN; LIANG, 2002), sendo que Paux *et al.* (2004) e Foucart *et al.* (2006) utilizaram a técnica em *Eucalyptus*.

Na técnica de *Massively parallel signature sequencing* (MPSS) milhões de *microbeads* são ligadas a adaptadores, associadas as sequências de cDNA da amostra e posicionadas em uma plataforma de análise. O sequenciamento das amostras (cDNAs) é produzido a partir de passos sucessivos de clivagem com uma enzima de restrição endonuclease tipo II, que reconhece seu sítio na sequência dos adaptadores e cliva adiante na sequência da amostra liberando 16 nucleotídeos em cada passo de clivagem. Como os adaptadores são fluorescentes, uma câmera com um software para processar imagens capta os sinais de fluorescência durante os passos de hibridização, ligação e clivagem. A possibilidade de gerar uma enorme quantidade de dados em pouco tempo, e o tamanho das sequências produzidas (geralmente entre 16 e 20 bases), fazem com que a técnica seja comparada ao SAGE. No entanto, por ser tecnicamente complexa e apresentar um elevado custo total, poucos grupos têm utilizado esta técnica, que ainda não foi usada na análise da expressão gênica em espécies florestais. (BRENNER *et al.*, 2000; DONSON *et al.*, 2002; MEYERS *et al.*, 2004).

Dentre os métodos fechados para análise da expressão gênica, usualmente distinguem-se dois tipos principais de arranjos (*arrays*), os macro-arranjos e os micro-arranjos. As diferenças entre eles baseiam-se no tipo de suporte, na marcação das amostras, no tamanho do arranjo e no número/densidade de sondas fixadas sobre o suporte (LAU et al, 2007).

Os macro-arranjos, considerados antecessores dos micro-arranjos, são aqueles cujos suportes nos quais as sondas são hibridizadas são membranas, normalmente de *nylon* ou nitrocelulose, a marcação é do tipo radioativa e geralmente contém entre dezenas e milhares de sondas. Embora tenha menor sensibilidade e permita testar menos genes do que os micro-arranjos, esse apresenta um menor custo e por isso tem sido utilizado por diversos grupos de pesquisa em trabalhos envolvendo *Populus* e *Eucalyptus* (LAU et al., 2007; CHRISTOPHER et al., 2004; PAUX et al., 2004; FOUCART et al., 2006). Já os micro-arranjos ou DNA *chips*, caracterizam-se por serem suportes, geralmente de vidro, com *spots* microscópicos sobre os quais milhares de sequências (genes) podem ser analisados simultaneamente em um único experimento. Os cDNAs da amostra e do controle são marcados com dois fluoróforos diferentes e submetidos a hibridização, sendo que um gene é considerado induzido quando há predominância da emissão de fluorescência do fluoróforo da amostra. Atualmente, dois diferentes tipos de micro-arranjos podem ser utilizados, os micro-arranjos de cDNA ou de oligonucleotídeos (STEIN; LIANG, 2002; DONSON et al., 2002; ALBA et al., 2004)

Frequentemente costuma-se utilizar os micro-arranjos de cDNA, uma vez que o *chip* pode ser preparado diretamente a partir de sequências de cDNA disponíveis em bancos de dados públicos e também fragmentos de PCR. Os micro-arranjos de oligonucleotídeos por sua vez, apresentam como grande vantagem o fato de permitirem a avaliação de um número muito maior de genes em um *chip*, assim como, uma hibridização com maior especificidade de um único gene com cada oligonucleotídeo. Esta última característica pode representar também sua maior desvantagem, uma vez que, a especificidade da hibridização demanda grande quantidade de informações de sequências para o desenho de oligonucleotídeos específicos, tornando um experimento de micro-arranjos de nucleotídeos muito mais caro do que um experimento de micro-arranjos de cDNA (ALBA et al., 2004).

Atualmente, de todas as técnicas disponíveis para estudos de expressão gênica, certamente a técnica de micro-arranjos é a mais utilizada em diversos organismos, incluindo plantas. Sem dúvida, os micro-arranjos proporcionaram uma revolução na análise da expressão gênica por permitir de maneira rápida, a análise do padrão de expressão de milhares de genes sob vários tratamentos ou pontos no tempo. Todavia, problemas como o alto custo dos equipamentos, a disponibilidade dos arranjos, hibridização cruzada, baixa sensibilidade para transcritos de baixa expressão e principalmente, a necessidade da existência de informações de sequências genômicas completas ou grandes quantidades de ESTs disponíveis, tem feito com que a tecnologia fique restrita a trabalhos envolvendo um número limitado de espécies que possuem essas informações disponíveis, como por exemplo as espécies modelos, *Arabidopsis*, arroz, e *Populus* (MEYERS et al., 2004)

No caso de espécies arbóreas, a grande maioria dos estudos utilizando micro-arranjos está concentrada em *Populus*, devido à disponibilidade de informações (HERTZBERG et al., 2001; ASPEBORG et al., 2005; MOREAU et al., 2005; ANDERSSON-GUNNERÅS et al., 2006). Possivelmente, seja justamente devido a essa pequena disponibilidade pública de informações, que apenas poucos grupos que detêm uma certa quantidade de ESTs tenham desenvolvido trabalhos com a técnica de micro-arranjos em espécies de eucalipto (KIRST et al., 2004; KIRST et al., 2005).

2.5.1 A técnica de SAGE (Análise Serial da Expressão Gênica)

A técnica de SAGE (*Serial Analysis of Genes Expression*) (Figura 1), desenvolvida por Velculescu et al. (1995), permite identificar quais genes são expressos em um dado momento, possibilitando uma análise quantitativa acurada da expressão gênica de um grande número de transcritos simultaneamente, inclusive daqueles ainda não conhecidos, com a exclusividade de permitir que o perfil dos transcritos seja apresentado na forma de dados digitais. Consequentemente, o acúmulo dos resultados da análise tipo SAGE possibilita a construção de bancos de dados de transcritos de plantas que poderão ser acessados pela comunidade científica (MATSUMURA et al., 1999).

Na técnica SAGE, um fragmento de DNA de 9-11 pares de base (*tag*) adjacentes ao sítio de restrição *Nla* III localizado próximo à extremidade 3' de cada cDNA é extraído por uma série de ligações com adaptadores e digestão com enzimas de restrição. As *tags* são amplificadas por PCR, concatenadas, clonadas em um vetor plasmidial e sequenciadas. Devido à concatenação, o sequenciamento de uma biblioteca SAGE é mais rápido e proporciona maior economia quando comparado ao sequenciamento de bibliotecas de ESTs, que baseia-se na clonagem de sequências expressas que representam apenas um único transcrito por clone. Dessa forma, a partir de um mesmo número de clones sequenciados, a técnica SAGE permite identificar um número muito maior de transcritos (SUN et al., 2004).

Figura 1 - Representação esquemática da construção da biblioteca de SAGE. Adaptado do site (<http://www.ncbi.nlm.nih.gov>)

A frequência de cada *tag*, ou seja, número de vezes que uma determinada *tag* foi sequenciada, reflete diretamente a abundância de cada RNA mensageiro presente no tecido. Além disso, o tamanho das *tags* (10 pb) contém informação suficiente para

identificar unicamente os genes (ou pelo menos uma família de genes similares), uma vez que esta é extraída a partir de uma posição definida do transcrito. O número máximo possível de *tag* com sequências diferentes ($4^{10}=1.048.576$) é muito maior do que o número estimado de genes em plantas (15.000 – 33.000 em *Arabidopsis*, COOKE et al., 1996), assegurando teoricamente a identificação única pela sequência da *tag*.

Um dos pontos críticos da técnica é o chamado *tag mapping*, ou seja, o estabelecimento de uma correta associação entre cada uma das *tags* sequenciadas e o transcrito que lhe deu origem (LASH et al., 2000). Essa associação é feita através de um banco de dados de sequências, onde somente devem ser considerados os resultados nos quais a *tag* encontra-se posicionada imediatamente após o último sítio de reconhecimento da enzima utilizada, antes da cauda poli(A) na extremidade 3'. As *tags* devem, preferencialmente, ser mapeadas em bancos de ESTs 3' ou sequências completas de cDNA, o que é muito difícil no caso de plantas e principalmente para espécies arbóreas, uma vez que os investimentos concentram-se principalmente nos estudos envolvendo humanos.

Dessa forma, algumas alternativas têm sido utilizadas na tentativa de se realizar o *tag mapping*. Uma delas é o SAGEmap disponível no site do NCBI (<http://www.ncbi.nlm.nih.gov/sage>), que extrai *tags* “virtuais” a partir de sequências expressas (ESTs e cDNAs) depositadas que contenham a terminação poli(A) ou o sinal de poliadenilação e as compara com o conjunto de *tags* obtidas experimentalmente (PLEASANCE et al., 2003). Outra alternativa, é o uso de bancos genômicos, onde todas as possíveis *tags* “virtuais” do genoma de uma espécie são extraídas e comparadas com um conjunto das *tags* obtidas experimentalmente, permitindo não só uma maior porcentagem de identificação dos genes, como também evidenciando a regulação por *splicing* (FIZAMES et al., 2004; ROBINSON et al., 2004).

A grande vantagem da técnica de SAGE, principalmente quando comparada a outras técnicas como micro-arranjos e ESTs, reside no fato desta permitir o acesso a grande maioria dos genes expressos em uma amostra em um determinado momento, incluindo aqueles genes ainda não conhecidos, revelando os valores absolutos dos níveis de expressão. Da mesma forma, a técnica propicia maior sensibilidade à análise,

permitindo acessar transcritos de baixa abundância geralmente envolvidos na regulação da expressão gênica durante a transcrição e a tradução, e que normalmente são perdidos em outros tipos de análise (SUN et al., 2004; GIBBINGS et al., 2003). Além disso, a técnica de SAGE permite que as *tags* sejam utilizadas na forma de *primers* ou sondas para identificar genes ainda desconhecidos como demonstrado em banana (COEMANS et al., 2005) e humanos (CHEN, J et al., 2002), bem como possibilita a identificação da ocorrência de regulação anti-sense, como demonstrado em estudos envolvendo diferentes organismos, como humanos (GE et al., 2006), arroz (GIBBINGS et al., 2003) e cana-de-açúcar (CALSA; FIGUEIRA, 2007).

Apesar de todas as vantagens e das aplicações que a técnica oferece, alguns autores ainda questionam o potencial de uso das *tags* geradas pela SAGE na identificação de genes únicos, principalmente para genes que pertencem a famílias gênicas, uma vez que nesse caso, uma única *tag* pode identificar mais de um gene (GE et al., 2006). Para aumentar a fidelidade do *tag mapping*, diversas modificações vêm sendo propostas e incorporadas a técnica, buscando aumentar a eficiência da clonagem e tamanho dos insertos. Por exemplo, o *SuperSAGE* é baseado no uso de uma enzima de restrição tipo III para gerar *tags* maiores que 25 pb (MATSUMURA et al., 2003) já *LongSAGE*, *tags* de 20 pb são produzidas tanto a partir da região 5', quanto da região 3' dos transcritos, aumentando a especificidade de identificação dos genes (WEI et al., 2004).

Devido a enorme quantidade de sequências disponíveis para humanos, a análise SAGE tem sido utilizada amplamente em estudos envolvendo a identificação da expressão diferencial de transcritos em células humanas normais e cancerosas para a identificação de genes tumores-específicos que poderão ser utilizados como marcadores para facilitar tanto o diagnóstico quanto o tratamento do câncer (LEE et al., 2003; NACHT et al., 1999; ZHANG et al., 1997).

Sem dúvida, a aplicação mais importante da SAGE é a identificação da expressão diferencial de genes, sendo um método efetivo para o reconhecimento de genes que são expressos em tecidos específicos em resposta a estresses ou doenças e, portanto, bastante indicado para estudos envolvendo vegetais superiores (MATSUMURA et al., 1999).

O primeiro estudo quantitativo da expressão global de genes em plantas superiores foi realizado por Matsumura e colaboradores (1999). Os autores obtiveram grande êxito na análise da expressão diferencial de genes em plântulas de arroz cultivadas em condições aeróbias e em condições anaeróbias, utilizando a técnica SAGE. Os resultados obtidos permitiram demonstrar que alguns genes foram induzidos em condições anaeróbias, como por exemplo, o gene da prolamina (proteína de reserva comumente expresso em sementes, mas não em folhas ou brotos, sugerindo uma nova função em anaerobiose para ele), enquanto outros foram reprimidos. Existem ainda, outros trabalhos utilizando *Oryza sativa* cujos objetivos vão desde identificar genes envolvidos na resistência a doenças (MATSUMURA et al., 2003), até caracterizar o padrão de expressão gênica em diferentes tecidos tais como folhas, sementes (GIBBINGS et al., 2003) e plântulas (MATSUMURA et al., 1999).

Nos últimos anos foram desenvolvidos diversos trabalhos envolvendo a espécie modelo *Arabidopsis thaliana*. Jung et al. (2003) utilizaram a técnica SAGE para identificar genes diferencialmente expressos em plantas cultivadas em condições normais de crescimento e plantas submetidas ao estresse em baixas temperaturas, já Lee & Lee (2003) utilizaram SAGE para identificar mudanças no padrão de expressão gênica do pólen também submetido ao estresse em baixas temperaturas. Em 2004, Fizames e colaboradores (2004) utilizaram a técnica para identificar genes especificamente expressos em raízes.

Mais recentemente a técnica SAGE também vem sendo utilizada em estudos com outras plantas superiores, como é o caso de banana (COEMANS et al., 2005) e cana-de-acúcar (CALSA; FIGUEIRA, 2007). No caso de espécies arbóreas, Lorenz e colaboradores (2002) foram os primeiros a utilizar a técnica. Os autores estabeleceram o perfil da expressão gênica no xilema secundário e identificaram genes que são diferencialmente expressos ao longo do gradiente de desenvolvimento do xilema em *Pinus*, construindo duas bibliotecas de células de xilema isoladas tanto do topo quanto da base do tronco. Foram extraídas 85.000 *tags* representando 27.398 genes expressos da biblioteca do topo e 65.000 *tags* representando 25.983 genes expressos na biblioteca da base. Combinando essas amostras para refletir o total de genes expressos no xilema lignificado, 150.000 *tags* foram catalogadas, representando um

total de 42.641 genes, entretanto os autores analisaram apenas as 500 *tags* mais expressas. Para eucalipto, Salvatierra (2006) utilizou a técnica de SAGE para identificar genes diferencialmente expressos em árvores de *E. grandis* susceptíveis e resistentes à ferrugem (*Puccinia psidii*), sendo capaz de identificar 232 genes com expressão diferencial e que podem estar envolvidos de alguma forma na resistência a doença. Já Carvalho (2007), utilizou a técnica para caracterizar os perfis transcricionais de folhas e da região cambial de árvores juvenis de *E. grandis*, o que possibilitou a identificação de genes exclusivamente expressos em folhas e em madeira, revelando ainda uma lista de potenciais genes para busca de promotores tecido-específicos.

Todos esses trabalhos mostram, portanto, que a técnica tem grande utilidade para a identificação de novos genes, bem como para o estudo dos níveis de expressão de genes envolvidos nas diferentes fases da formação da madeira, determinando suas características, como por exemplo, aquelas necessárias à indústria de papel e celulose.

2.6 Proteômica

Para compreender melhor a função de todos os genes de um organismo, é necessário conhecer não só quais genes são expressos, quando e onde, mas também quais são os produtos da expressão, e em que condições esse produtos, as proteínas, são sintetizados em certos tecidos. A proteômica, estudo sistemático em larga escala das proteínas (proteoma) de um determinado tecido ou organismo, tem o objetivo de fornecer informações detalhadas sobre a estrutura e a função dessas proteínas nos diferentes sistemas biológicos, permitindo ainda verificar se existe correlação entre o nível de expressão dos mRNA e das proteínas, a ocorrência de isoformas gênicas e modificações pós-traducionais (PANDEY; MANN, 2000; PATTERSON; AEBERSOLD, 2003).

Os organismos podem sintetizar milhares de proteínas ao mesmo tempo, e a diversidade potencial de tipos de proteínas no proteoma certamente excede o número estimado de genes no genoma. Isso ocorre porque os produtos de um gene podem diferir devido ao processamento alternativo do mRNA antes da tradução em proteína (*Splicing* Alternativo) e a uma variedade de modificações pós-traducionais, como por exemplo, fosforilação, glicosilação, metilação, clivagem, entre outras. Dessa forma, o

proteoma celular é muito complexo e formado por um grande número de proteínas que variam conforme o tecido ou condição ambiental a que é submetido o organismo (LOPEZ, 1999; BINNECK, 2004).

De modo geral, a análise proteômica baseia-se em duas plataformas metodológicas complementares. A primeira plataforma realiza uma separação prévia dos componentes a serem analisados com o objetivo de aumentar, através de uma separação de moléculas, o poder resolutivo da futura análise dos componentes da amostra. Além disso, a primeira plataforma fornece uma visão dos compostos separados, permitindo uma comparação não só qualitativa, mas também quantitativa entre duas ou mais amostras. Para isso são utilizadas técnicas como a eletroforese bidimensional em gel de poliacrilamida (2D-PAGE) ou a cromatografia líquida multidimensional. Já a segunda plataforma realiza a análise e caracterização estrutural das moléculas separadas na primeira, principalmente através da espectrometria de massas (PIMENTA, 2003). Atualmente a espectrometria de massas pode ser realizada através de dois métodos ionização branda de moléculas: MALDI (*Matrix Assisted Laser Desorption/Ionization*) e ESI (*Electrospray ionization*) (MANN; HENDRICKSON, 2001).

A análise do tipo MALDI (KARAS et. al., 1987) tem como princípio a ionização de biomoléculas através do bombardeamento com laser. Neste tipo de análise, a amostra é misturada a um solvente contendo um material com capacidade de absorver luz no comprimento de onda do laser, conhecido como matriz e, aplicado em uma placa. Essa matriz absorve e transfere a energia do laser, permitindo a vaporização e ionização da amostra, além de protegê-la da decomposição por excesso de energia e prevenir a formação de agregados, prejudicando e comprometendo a análise (MANN; HENDRICKSON, 2001).

Na ionização por eletrospray (ESI) (FENN et al, 1989) a amostra é dissolvida em um solvente volátil ao passar por um estreito tubo capilar metálico, ao qual é aplicada uma voltagem, gerando um aerossol de analíto e solvente. Com a evaporação do solvente originam-se íons que são subsequentemente submetidos a alto vácuo, acelerados através de um campo elétrico para o analisador de massas do espectrômetro e separados de acordo com sua razão m/z (massa/carga). Um grande número de compostos pode ser analisado por ES e a principal vantagem do ESI é a

possibilidade de combiná-lo com técnicas de separação e purificação como a cromatografia líquida e a eletroforese capilar (MANN; HENDRICKSON, 2001).

Dessa forma, a separação das proteínas por eletroforese bidimensional combinada com a identificação por espectrometria de massas tem sido a estratégia mais utilizada em estudos de proteômica (GÖRG et al., 2004). No caso de espécies arbóreas, diversos estudos vêm sendo conduzidos em *Pinus* (COSTA et al., 1999; GION et al. 2005), *Populus* (MIJNSBRUGGE et al., 2000) e *Eucalyptus* (PLOMION et al., 2003; CELEDON et al., 2007).

A complexidade da formação da madeira também tem sido estudada do ponto de vista proteômico, uma vez que proteínas produzidas preferencialmente durante o desenvolvimento do xilema podem fornecer informações substanciais a respeito da xilogênese. Para revelar a identidade dessas proteínas, Mijnsbrugge e colaboradores (2000), utilizando a técnica de 2D-PAGE, estabeleceram o perfil protéico do xilema em diferenciação (jovem), do xilema maduro e da casca, em *Populus* cortados em diferentes épocas do ano. As proteínas mais abundantes do xilema foram identificadas pela análise de micro-sequências, e dessa forma, puderam ser estabelecidas as possíveis funções para 15 dessas proteínas, baseada na similaridade com enzimas vegetais previamente caracterizadas e presentes nos bancos de dados *Poplar EST database*.

Proteínas de madeira de tensão e de madeira normal de eucalipto foram comparadas e apresentaram padrão diferencial nos tecidos indicando possível ligação com o desenvolvimento do xilema (PLOMION et al., 2003). Recentemente, Celedon e colaboradores (2007), utilizando um sistema de 2D-LC-MS/MS para identificar proteínas que participam do processo de formação da madeira a partir de amostras da região cambial de *Eucalyptus* em 3 diferentes idades de crescimento (6 meses, 3 e 6 anos), conseguiram identificar 240 proteínas, sendo que a comparação entre as idades revelou que existem diferenças no padrão de expressão para algumas proteínas.

Contudo, a análise conjunta dos dados gerados pela genômica, transcritômica e proteômica, por meio da bioinformática, é extremamente importante, já que pode indicar se a regulação de um gene ocorre em nível transcricional ou traducional, representando

ainda, a possibilidade para se entender melhor os caminhos entre os genes e sua verdadeira função nas células (van WIJK, 2001).

2.7 A formação da madeira

A madeira pode ser definida como sendo um conjunto de células mortas do xilema secundário que são acumuladas ao longo do desenvolvimento das espécies lenhosas (LARSON, 1994). O xilema secundário, cujo processo de formação é chamado de xilogênese, é responsável por transportar e estocar, água, nutrientes e hormônios (ácido abscísico, citocininas), funcionando também como suporte mecânico ao corpo das árvores (FUKUDA, 1996). Dessa forma, a madeira é um complexo tecido composto pelos vários polímeros orgânicos que compõem a parede celular vegetal (celulose, hemiceluloses e lignina), estrutura que permanece após a morte das células do xilema secundário (REITER, 2002). Aproximadamente 40% a 50% da madeira é formada por celulose, 25% por hemiceluloses, 25% a 35% por lignina e, uma pequena porcentagem, por proteínas de parede celular. Além disso, outras substâncias químicas podem estar presentes nas madeiras, como por exemplo, os extrativos, que compreendem resinas, taninos, óleos, gomas, compostos aromáticos e sais de ácidos orgânicos (PLOMION et al., 2001).

2.7.1 Parede celular vegetal

Além de exercerem uma série de funções estruturais, de proteção e regulação do crescimento durante a vida da planta, as paredes das células vegetais apresentam grande importância prática na nutrição humana e animal, funcionando também como fonte de fibras para a produção têxtil e produção de papel. Por estas razões, o estudo da biossíntese da parede celular é considerado interessante por ambos pontos de vista: básico e aplicado (REITER, 2002).

Dois tipos de parede celular podem ser distinguidos, as paredes primárias e as paredes secundárias. A parede celular primária se origina durante a divisão celular e aumenta sua área superficial rapidamente durante a expansão da célula (CARPITA; McCANN, 2000). A interface entre as paredes primárias de células vizinhas é denominada de lamela média, cuja composição difere do restante da parede, por ser

rica em pectina e conter proteínas diferentes (TAIZ; ZEIGER, 2004). A parede secundária, por sua vez, forma-se depois de cessado o crescimento da célula, e pode se tornar altamente especializada em estrutura e composição, refletindo o estado diferenciado da célula (TAIZ; ZEIGER, 2004).

2.7.1.1 Parede celular primária

A parede celular primária é composta basicamente pela interação de três polissacarídeos (celulose, hemicelulose e pectinas) estruturalmente interligados e por proteínas de parede (Figura 2) (CARPITA; GIBEAUT, 1993). A estrutura da parede celular é formada por uma rede de celulose e hemiceluloses embebida em uma matriz de polissacarídeos pécnicos, o que fornece ao mesmo tempo, resistência e flexibilidade (TAIZ; ZEIGER, 2004).

Figura 2 - Representação esquemática da parede celular primária, mostrando a interação dos seus constituintes (celulose, hemicelulose e pectina). Estrutura originalmente proposta por McCann e Roberts (1992)

As microfibrilas de celulose são estruturas relativamente rígidas que contribuem para resistência e predisposição estrutural da célula, sendo sintetizadas na membrana plasmática, utilizando UDP-D-glucose como precursor. Essas microfibrilas são formadas por várias (aproximadamente 36) cadeias lineares de β -(1 $\rightarrow$ 4) D-glicose, unidas umas às outras, ao longo de seu comprimento, por pontes de hidrogênio. (CARPITA; McCANN, 2000).

As hemiceluloses são sintetizadas no complexo de Golgi e caracterizam-se por serem polissacarídeos flexíveis que ligam-se à superfície da celulose. Diversos tipos de

hemiceluloses são encontrados em paredes celulares, entretanto, na parede primária das dicotiledôneas, a hemicelulose mais abundante é o xiloglucano, que na maioria das vezes, está presente na mesma quantidade que a celulose. Este polímero possui uma estrutura básica de resíduos de β -(1 $\rightarrow$ 4) D-glicose, com cadeias laterais de xilose, podendo ainda esses resíduos de xilose serem substituídos por arabinose e galactose, sendo que dependendo da espécie, a galactose pode ser substituída ainda por fucose. (TAIZ; ZEIGER, 2004; REITER, 2002).

Já, as pectinas, sintetizadas também no complexo de Golgi, incluem vários tipos diferentes de polissacarídeos ricos em ácido galacturônico (GalA). As pectinas formam uma fase gel hidratada na qual está implantada a rede celulose-hemicelulose atuando como preenchimento hidrofílico e impedindo assim a agregação e o colapso da rede de celulose. (TAIZ; ZEIGER, 2004; REITER, 2002).

Além dos polissacarídeos, a parede celular contém várias classes de proteínas estruturais, classificadas de acordo com sua composição predominante de aminoácidos, por exemplo, glicoproteína rica em hidroxiprolina (HRGP), proteína rica em glicina (GRP) e proteína rica em prolina (PRP). Essas proteínas são reguladas pelo desenvolvimento e suas quantidades variam de acordo com o tecido e espécie (TAIZ; ZEIGER, 2004).

A composição e organização estrutural da parede celular primária permitem que a mesma seja forte, fina, flexível e capaz de sofrer extensão plástica e elástica (FRY, 2004).

2.7.1.2 Parede celular secundária

A parede celular secundária é depositada após a expansão celular e frequentemente apresenta uma composição e organização distinta. Esse tipo de parede é sintetizado geralmente por células que requerem uma grande força mecânica e um reforço estrutural (COSGROVE, 2005), como por exemplo, traqueídes e elementos de vaso (TAIZ; ZEIGER, 2004). Essa formação é dirigida pela expressão coordenada de um grande número de genes especialmente envolvidos na biossíntese e reunião de quatro componentes principais: polissacarídeos (celulose e hemiceluloses); ligninas; proteínas da parede celular e outros constituintes menores solúveis (estilbeno,

flavonóides, taninos e terpenóides) e insolúveis (pectina e proteínas) em solvente neutro (HIGUCHI, 1997).

A parede secundária pode ser dividida em três camadas distintas, chamadas S_1 , S_2 e S_3 , (PLOMION et al., 2001) que apresentam diferenças na orientação das microfibrilas de celulose (BIDLACK et al., 1992) (Figura 3). A primeira camada formada, a S_1 , é a mais fina (0,1-0,35 μm) e representa apenas 5-10% da espessura da parede celular (PLOMION et al., 2001). Nessa camada, as microfibrilas de celulose formam um ângulo de 60-80° em relação ao eixo da célula. A camada S_2 é a mais espessa (1-10 μm) e importante com relação ao suporte mecânico. Essa camada, por sua vez, representa 75-85% da espessura total da parede celular e o ângulo formado pelas microfibrilas de celulose em relação ao eixo da célula pode variar de 5-30°. A orientação das microfibrilas de celulose da camada S_2 pode influenciar grandemente nas propriedades físicas e mecânicas da célula. A camada S_3 é relativamente fina (0,5-1,10 μm) e as microfibrilas de celulose apresentam um arranjo paralelo, formando um ângulo de 60-90°, em relação ao eixo da célula. Os microtúbulos formados por α e β – tubulina determinam a posição e orientação das microfibrilas de celulose durante a formação da parede secundária em plantas (CHAFFEY, 2000). Nessas três camadas também estão presentes hemiceluloses e lignina (PLOMION et al., 2001; MELLEROWICZ, et al., 2001). Além de uma proporção mais alta de celulose, a parede secundária tipicamente contém menos xiloglucano e mais xilanos e glucumananos (TAIZ; ZEIGER, 2004).

Figura 3 - Representação esquemática da parede celular secundária, mostrando a presença da parede primária; das camadas S_1 , S_2 e S_3 da parede secundária; lúmen e lamela média. Retirado de Plomion et al. (2001)

A lignina, um dos maiores componentes de algumas paredes secundárias (CARPITA; McCANN, 2000), é formada pela polimerização de três diferentes álcoois de fenilpropanóides (ou monolignóis): coniferil, cumaril e sinapil (TAIZ; ZEIGER, 2004). Esses monolignóis dão origem às unidades G, H e S, respectivamente, do polímero de lignina, que diferem entre si apenas pelo grau de metoxilação (PLOMION et al., 2001; MELLEROWICZ, et al., 2001). Essas unidades são sintetizadas a partir de fenilalanina, secretadas na parede, e oxidadas no local apropriado pelas enzimas peroxidase e lacase (TAIZ; ZEIGER, 2004). O conteúdo de lignina, assim como de seus monômeros, diferem entre taxa, tecidos, tipos celulares e camadas da parede, dependendo do estágio de desenvolvimento da planta e condições ambientais (PLOMION et al., 2001; MELLEROWICZ, et al., 2001). A lignina embebida na matriz de polissacarídeos, além de conferir rigidez e coesividade, proporciona hidrofobicidade superficial aos tecidos condutores, necessária para o transporte de água (PLOMION et al., 2001). Outro papel importante da lignina é a sua atuação na proteção das plantas contra patógenos (FUKUDA, 1996). Dada a sua importância, as plantas de habitat terrestres apenas apareceram após a evolução da biossíntese de lignina, devido às funções essenciais a biologia dessas plantas (WHETTEN; SEDEROFF, 1995).

A deposição de lignina pode ser induzida em muitos tipos de células por doenças ou fermento, indicando que os genes envolvidos na biossíntese de lignina respondem tanto aos sinais de desenvolvimento, quanto aos sinais ambientais (WHETTEN; SEDEROFF, 1995). Embora vital para a arquitetura da planta, a lignina apresenta um impacto negativo no processamento do papel e também na alimentação de animais de criação. A maior parte do custo, complexidade e poluição associadas ao processo da fabricação de papel se deve a remoção da lignina da madeira para obtenção de um papel de alta qualidade. Similarmente, a quantidade de energia digestível disponível para os animais ruminantes, a partir da fermentação dos polissacarídeos da parede celular é restringido pela lignina, tornando inacessível uma grande parte da celulose e hemicelulose para os microrganismos do rúmen. Devido aos benefícios econômicos que podem ser alcançados, consideráveis esforços em pesquisa têm buscado a redução da quantidade de lignina ou modificação da estrutura da lignina para facilitar a

extração da polpa celulósica para fabricação de papel, assim como, melhorar a digestibilidade das plantas forrageiras (JUNG; NI, 1998).

2.7.2. Xilogênese

A xilogênese é um processo contínuo de diferenciação que se inicia a partir das células do câmbio vascular e resulta na formação do xilema secundário (madeira).

O câmbio vascular é geralmente descrito como uma camada única de células não diferenciadas 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 das células mantém as características meristemáticas, enquanto a outra, mais interna, se torna uma célula mãe do xilema. Tanto as iniciais fusiformes quanto as iniciais radiais estão envolvidas na gênese do xilema através da formação dos elementos longitudinais (elementos de vaso ou traqueídes, parênquima longitudinal e fibras) e dos elementos transversais do parênquima radial, respectivamente (LARSON, 1994). Dessa forma, através da atividade do câmbio ocorre uma constante renovação das células do xilema e floema, resultando em um aumento do diâmetro da árvore. As células do xilema, próximas ao câmbio, sempre se dividem mais que as células do floema, explicando a variação de espessura que existe entre o tecido do floema e xilema (STRASBURGER, 1994; PLOMION et al., 2001; CHAFFEY et al., 2002).

Após ser formada, a nova célula do xilema passará por um processo de diferenciação através de uma sucessão de processos que podem ser divididos em alguns passos principais: a) expansão celular (alongamento e aumento da largura radial); b) síntese e deposição da parede celular (celulose, hemicelulose, proteínas da parede celular); c) lignificação (biossíntese e deposição de lignina); d) morte celular programada (LARSON, 1994, CHAFFEY, 1999).

A maturação das células do xilema está associada à morte das células dos vasos e das fibras por meio de um processo interno de suicídio celular. Desta forma, quando a lignificação é completada, os elementos condutores do xilema sofrem morte celular programada (PCD), levando a perda de todo conteúdo celular, resultando então na

formação de uma unidade única de vasos condutores do xilema (FUKUDA, 1996; ROBERTS; MC CANN, 2000; GROOVER; JONES, 1999).

A xilogênese representa, portanto, um exemplo da diferenciação celular em uma forma extremamente complexa. Este processo é controlado por uma enorme variedade de fatores exógenos (fotoperíodo e temperatura) e endógenos (hormônios vegetais) e pela interação entre eles. A presença de famílias multigênicas, a extrema plasticidade do metabolismo envolvido, e a expressão coordenada de muitos genes estruturais envolvidos em todas as etapas da xilogênese proporcionam essa enorme complexidade, dificultando o entendimento sobre o processo de formação da madeira (PLOMION et al., 2001).

2.7.3 Análise da expressão gênica

Apesar da reconhecida importância econômica, muito pouco se sabe sobre os fatores genéticos e os processos celulares, moleculares e bioquímicos envolvidos na formação da madeira (HIGUCHI, 1997; CARPITA & McCANN, 2000; PLOMION et al., 2001).

A utilização de técnicas moleculares tem permitido a identificação e caracterização de algumas “chaves moleculares” envolvidas no processo de formação da madeira (HIGUCHI, 1997; SAVIDGE et al., 2000), assim como envolvidos na lignificação (WHETTEN et al., 1998), síntese de celulose (ARIOLI et al., 2000), hemicelulose (REID, 2000), proteínas da parede celular (ZHANG et al., 2000) e pectinas (REID, 2000).

A grande maioria dos estudos utiliza plantas anuais, por exemplo, *Zinnia elegans* e *Arabidopsis thaliana*, como modelo. Em *Zinnia*, células em suspensão foram estimuladas a produzir elementos traqueários. Dessa forma, proteínas e genes puderam ser identificados em diferentes estádios do desenvolvimento desses elementos traqueários (FUKUDA, 1997).

Graças ao mecanismo único de transdiferenciação, através do qual células diferenciadas podem iniciar um novo programa de diferenciação, Demura e colaboradores (2002) construíram um perfil compreensivo da expressão gênica durante a transdiferenciação de células do mesófilo em células de xilema, usando um sistema

de cultura *in vitro* de *Zinnia*. Neste modelo, células fotossintéticas do mesófilo isoladas de folhas de *Zinnia* se transdiferenciam em células do xilema através de um processo morfogênico caracterizado por estágios como a formação da parede secundária e a morte celular programada. Através da construção de bibliotecas de cDNA foi possível a obtenção de mais de 8.000 ESTs de *Zinnia*. A posterior análise de micro-arranjos permitiu estabelecer um padrão de expressão gênica em diferentes estádios, a identificação de genes não descritos previamente e que codificam proteínas que podem estar relacionadas ao processo de transdiferenciação, assim como proteínas quinases e fatores de transcrição.

Arabidopsis thaliana também tem sido considerado um ótimo sistema para a análise de genes envolvidos na formação da parede celular e madeira, através da utilização de mutantes (REGAN et al., 1999; ZHONG et al., 2001; ROBERT et al., 2005). Estima-se que cerca de 15% do genoma em *Arabidopsis* seja dedicado à biogênese e modificação da parede celular e destes, apenas pequenos grupos têm sido caracterizados (CARPITA et al., 2001). Além disso, mais de 600 genes estão envolvidos na síntese e reciclagem de polissacarídeos e quase 25% destes estão envolvidos na reciclagem de pectinas. A utilização de plantas de *Arabidopsis* em estudos de diferenciação do xilema é estimulada pelo fato de que mesmo sendo herbáceas e, portanto, não apresentarem crescimento secundário, tais plantas são induzidas a formar xilema secundário através da remoção de suas inflorescências (CHAFFEY et al., 2002; OH et al., 2003). Em 2002, Chaffey e colaboradores 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*. Já, Samuga e Joshi (2002) demonstraram que genes relacionados com o desenvolvimento da parede celular secundária, como celulose sintase (CESA), possuem 91% de similaridade. Por meio de estudos realizados em xilema de *Arabidopsis*, através da técnica de micro-arranjos, OH et al. (2003) observaram que uma série de fatores de transcrição estava superexpresso no xilema secundário, quando comparado ao xilema primário. Esses resultados sugerem que o crescimento secundário nas plantas deve-se a regulação da expressão gênica e não à presença de genes estruturais diferenciados nas plantas lenhosas, tornando possível a

indução do crescimento secundário em espécies não lenhosas. Da mesma forma, KO et al. (2004) encontraram 700 genes diferencialmente expressos durante a transição do xilema primário para xilema secundário em *Arabidopsis*.

Apesar dos estudos de expressão gênica para estudar a formação da madeira através da utilização de plantas modelo, como *Arabidopsis*, e sistemas de culturas de células, como *Zinnia*, terem identificado muitos genes associados com o processo de diferenciação do xilema, estes estudos representam uma primeira etapa para compreender profundamente os mecanismos moleculares e genéticos que controlam o crescimento secundário. Dessa forma, evidências mais detalhadas e mais profundas são necessárias, pois embora os sistemas que utilizam herbáceas e sistemas de culturas de células sejam usados para estudar o processo de diferenciação do xilema, a total compreensão do crescimento secundário requer um sistema de estudos em árvores propriamente ditas (LI et al., 2006).

Entre as espécies fornecedoras de madeira, *Populus* spp. e *Eucalyptus* spp são os principais candidatos para a realização de estudos envolvendo a análise de genes e proteínas relacionados à formação e qualidade da madeira, pelo fato de apresentarem: facilidade de propagação vegetativa (PLOMION et al., 2001); possibilidade de transformação genética (FILLATTI et al., 1987; GONZÁLEZ et al., 2002); pequeno tamanho do genoma (5×10^8 e 6×10^8 , respectivamente, pouco menor que *Arabidopsis*); a disponibilidade de mapas genéticos (GRATTAPAGLIA et al., 1998; GION et al., 2000) e por último, no caso de *Populus* por apresentar o genoma totalmente seqüenciado (TUSKAN et al., 2006). Entretanto, apesar da vasta gama de genes provavelmente relacionados com o processo de formação da madeira, o estudo em árvores teve início apenas no final dos anos 90 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).

Sterky et al. (1998) foram os pioneiros ao relatar cerca de 6.000 ESTs de *Populus* spp de tecidos formadores de madeira, em bibliotecas de cDNA preparadas a partir do câmbio e de tecidos do xilema em desenvolvimento. Seguindo a mesma linha, Hertzberger et al. (2001) estudaram o padrão de expressão de 2.995 genes durante os estágios sucessivos de maturação das células do xilema em *Populus*. Nesse estudo,

amostras de tecidos foram coletadas de cinco diferentes regiões (células meristemáticas, células em expansão, células em alongação, células em formação da parede secundária e células maduras), de modo a cobrir toda a sequência de desenvolvimento. A análise de micro-arranjos demonstrou que mais de 500 genes são diferencialmente expressos durante o desenvolvimento secundário do xilema.

A partir destes trabalhos, e principalmente, a medida que aumentaram as sequências de ESTs de espécies arbóreas disponíveis em bancos de dados públicos, diversos estudos de análise de expressão gênica no processo de formação da madeira, para identificar genes relacionados à sua formação em estádios específicos do processo de diferenciação do xilema, têm sido realizados utilizando-se *Populus* spp como modelo (SCHRADER et al., 2004, ASPEBORG et al., 2005; MOREAU et al., 2005; GEISLER-LEE et al., 2006).

Com o objetivo de identificar genes relacionados a transição do crescimento primário para o crescimento secundário, van Raemdonck e colaboradores (2005) coletaram amostras em duas posições do caule de árvores de *Populus* com 6 meses de idade. Uma amostra do ápice, cujo sistema vascular consiste basicamente de xilema e floema primários e uma amostra da base, que é caracterizada pela presença de crescimento secundário. Utilizando a técnica de cDNA-AFLP, os autores identificaram diversos genes potencialmente envolvidos na transição progressiva do crescimento primário para o crescimento secundário em *Populus*.

Atualmente, um ótimo modelo que vem sendo utilizado para estudar alterações na expressão gênica em espécies arbóreas é a utilização da indução da formação de madeira de reação. Normalmente induzida a partir do estímulo de inclinação da árvore, a madeira de reação é produzida como um mecanismo do tronco na tentativa de manter ou corrigir sua posição original, e caracteriza-se pelo baixo conteúdo de lignina, elevado conteúdo de celulose e menor ângulo das microfibrilas (PLOMION et al., 2001). Nas angiospermas, a madeira de reação é conhecida como madeira de tração e nas gimnospermas é chamada de madeira de compressão (TIMELL, 1986). Dessa forma, diversos estudos vêm sendo conduzidos em *Populus* utilizando esse modelo, encontrando diversas alterações nos conteúdos de celulose e lignina (PILATE et al.,

2004; DEJARDIN et al., 2004; ANDERSSON-GUNNERÅS et al., 2006; BHANDARI et al., 2006).

Em espécies de *Eucalyptus*, os primeiros estudos envolvendo a análise da expressão gênica no processo de xilogênese foram publicados apenas recentemente. PAUX et al. (2004), foram os precursores, tendo construído uma biblioteca subtrativa (SSH) de xilema usando cDNAs de xilema em desenvolvimento como controle e folhas como população-*driver*. Nesse estudo, foi identificado um total de 224 genes específicos do xilema, sendo que os genes para CesAs, genes de lignificação, XETs e expansinas foram significativamente mais expressos. Recentemente, Foucart et al. (2006) utilizaram a mesma técnica (SSH) para identificar genes diferencialmente expressos entre o xilema e o floema durante a xilogênese em *Eucalyptus*. Os autores ampliaram, portanto, os estudos iniciais, com o desenvolvimento de bibliotecas subtrativas para tecidos de xilema *versus* floema, identificando 87 genes preferencialmente expressos no xilema e que codificam proteínas envolvidas na formação da parede celular secundária, regulação hormonal e morte celular programada.

Utilizando a técnica de cDNA-AFLP, Ranik e Myburg et al (2006), estudaram o padrão de expressão dos genes relacionados a xilogênese em árvores de 6 anos de idade de *Eucalyptus*. Os autores coletaram amostras de tecidos em estágios diferentes do desenvolvimento da madeira (xilema imaturo, xilema maduro, floema e casca) em duas posições do caule, de modo a permitir uma visão global da abundância e variabilidade dos transcritos ao longo do caule. Um total de 71 transcritos diferencialmente expressos foi caracterizado, incluindo diversos genes relacionados à diferenciação celular, transdução de sinais e biossíntese de parede celular.

Kirst et al (2005) estudaram a variação na abundância de mRNAs no xilema em desenvolvimento em uma progênie híbrida de *Eucalyptus* usando micro-arranjos. As mudanças quantitativas na abundância da expressão gênica foram mapeadas em dois mapas de ligação de árvores individuais, tendo sido possível identificar um QTL de expressão gênica que explica até 70% da variação no nível de transcrição. Anteriormente, os mesmos autores, utilizando a mesma população e a técnica de micro-arranjos, identificaram uma correlação entre a baixa expressão dos genes envolvidos

na biossíntese de lignina e vias de metilação, associadas ao rápido crescimento do tronco. Além disso, foram identificados dois locos genéticos responsáveis pela regulação dos níveis de transcrição de genes relacionados a biossíntese de lignina que se co-localizam com QTLs de crescimento, sugerindo assim, que as mesmas regiões genômicas podem estar regulando essas características (KIRST et al., 2004).

Também em 2005, Paux e colaboradores utilizando o modelo de indução de madeira de reação analisaram os perfis de expressão na formação da madeira de reação no xilema de eucalipto. Os autores estudaram alguns genes em sistema de macro-arranjos sob diferentes tempos de indução e identificaram 196 genes diferencialmente regulados, sendo que os genes relacionados à biossíntese de celulose (CesA) apresentaram maior expressão em conjunto com a formação da camada G e os genes relacionados à lignina apresentaram menor expressão, em relação a madeira normal.

A identificação dos genes e dos mecanismos envolvidos na formação da madeira, associada às tecnologias de seleção assistida e transformação genética, representa uma excelente oportunidade para aumentar a eficiência, reduzir o tempo e o custo necessários para a seleção de genótipos superiores de modo a agregar cada vez mais fatores que elevam a qualidade da madeira e atendem as diversas necessidades da indústria. Dessa forma, a grande maioria dos estudos em eucalipto atualmente concentra-se na caracterização e isolamento de genes relacionados ao processo de formação da madeira, particularmente os envolvidos na biossíntese de lignina e de celulose (VALÉRIO et al., 2003; PAUX et al. 2005; KIRST et al., 2004).

3 MATERIAL E MÉTODOS

3.1 Material vegetal

Plantas de *Eucalyptus grandis* variedade M33 provenientes de sementes de meios-irmãos, cultivadas em São Miguel Arcanjo, Estado de São Paulo, foram fornecidas pela empresa Suzano Papel e Celulose. Essas plantas foram mantidas em vasos plásticos com volume de 3,5 litros, contendo uma mistura de vermiculita média e substrato orgânico (Plantmax) na proporção de 1:1 e distribuídas em casa de vegetação provida de luz suplementar para perfazer um total de 16 horas de luz (fotoperíodo). Diariamente foram fertirrigadas com solução nutritiva completa de Hoagland e Arnon (1950).

Ao completar seis meses de idade, 40 plantas tiveram suas folhas, raízes e cascas removidas, seus caules foram cortados (Figura 4), sendo imediatamente congelados em nitrogênio líquido e mantidos em freezer a -80°C . Por fim, todas as amostras das 40 plantas foram reunidas em um *bulk* e armazenadas em freezer (-80°C) até o momento da sua utilização para a extração do RNA total.

Figura 4 – Material vegetal utilizado no experimento. (A) Plantas de *Eucalyptus grandis* com seis meses (barra = 6 cm). (B) Caules de *Eucalyptus grandis* com seis meses sem a casca e cortados em segmentos, prontos para serem macerados (barra = 2 cm)

3.2 Extração do RNA total

Para evitar a degradação do RNA por RNases, todas as soluções e reagentes foram preparados com água previamente tratada com DEPC (dietilpirocarbonato) na concentração de 1% (v/v), por 16 horas a 37°C, seguida de autoclavagem (120°C, 15 psi, 20 minutos). Com a mesma finalidade, todo o material plástico utilizado foi tratado através de uma sequência de lavagens com SDS 2% (p/v) por 2 horas a 70°C, álcool comercial por 20 minutos, e H₂O₂ 3% (v/v), por mais 20 minutos. O material foi enxaguado duas vezes com água DEPC 0,01% (v/v) por 20 minutos antes de ser autoclavado. A vidraria e as porcelanas utilizadas foram esterilizadas a seco, a 200°C, por 6 horas.

A extração do RNA total das plantas de 6 meses de idade foi realizada seguindo o protocolo de extração baseado em fenol descrito por SALZMAN et al. (1999) com duas pequenas modificações: i) solubilização da amostra (2 g) em 20 mL do tampão de extração (o protocolo original recomenda a solubilização de 2 g de amostra em 10 mL do tampão de extração, porém testes anteriores realizados no laboratório indicaram melhor rendimento de extração quando solubilizadas em maior volume de tampão); ii) utilização do dobro da quantidade do agente antioxidante PVP-solúvel recomendado (o PVP atua de forma a impedir a oxidação dos compostos fenólicos na amostra, minimizando sua ligação aos ácidos nucleicos). De acordo com os testes de RT-PCR, descritos no item 3.4, havia um forte indicativo da presença de compostos fenólicos, inibidores de reação.

Desta forma, 2 g do material vegetal foi macerado em nitrogênio líquido com auxílio de almofariz e pistilo, transferido para tubos de centrifuga (Nalgene) contendo 20 mL do tampão de extração (Anexo A), seguido de agitação vigorosa, por 1 minuto, em vórtex. Em seguida, adicionou-se igual volume de CIA (24:1 v/v) e agitou-se por 20 minutos, antes da centrifugação a 16.000 x g, por 10 minutos, a 4°C. A fase superior foi transferida para novo tubo, repetindo-se esse passo mais duas vezes, antes da adição, ao sobrenadante final de 2 volumes de etanol absoluto e 0,1 volume de NaCl 5 M, para precipitação por 16 horas a -20°C. Após centrifugação (16.000 x g), por 10 minutos a 4°C, o precipitado foi ressuspensionado em 10 mL de água tratada com DEPC e novamente centrifugado por 5 minutos a 20.000 x g. O sobrenadante foi transferido para

novo tubo, ao qual adicionou-se 1 volume de CIA:Fenol (1:1 v/v). Após 10 minutos de agitação à temperatura ambiente, a amostra foi centrifugada a 13.000 x g por 10 minutos. A fase superior foi transferida para novo tubo, repetindo-se esse passo mais uma vez. Depois da centrifugação, adicionou-se 2 volumes de etanol absoluto e 0,1 volume de NaCl 5 M, ao sobrenadante, para precipitação adicional por 16 horas a -20°C. O precipitado resultante da centrifugação (16.000 x g) por 15 minutos a 4°C, foi ressuspenso em 500 µL de água tratada com DEPC. Após uma centrifugação adicional nas mesmas condições, o sobrenadante foi transferido para tubo *ependorf* de 1,5 mL, e o volume da amostra ajustado para 1 mL com água tratada com DEPC, ao qual adicionou-se 333 µL de LiCl 8 M para precipitação do RNA a 4°C, por 3 horas. O precipitado, resultante da centrifugação a 12.000 x g, por 20 minutos a 4°C, foi lavado com 400 µL de etanol 80% (v/v), novamente centrifugado a 12.000 x g, por 10 minutos a 4°C, e ressuspenso em 100 µL de água tratada com DEPC.

A quantificação das amostras de RNA foi realizada por espectrofotometria, sendo a concentração calculada, para uma cubeta de 10 mm, através da seguinte fórmula:

$$[\text{RNA}] = \frac{A_{260} \times Fc \times Fd}{1000}$$

- Onde:
- A₂₆₀**: absorbância da amostra sob o comprimento de onda de 260 nm;
 - Fc**: fator de conversão, no qual 1 unidade de absorbância a 260 nm corresponde a 40 µg RNA mL⁻¹ (SAMBROOK et al., 1989);
 - Fd**: fator de diluição da amostra de leitura, no caso, 200 (4 µL de amostra em 800 µL de água tratada com DEPC);
 - 1000**: fator de correção para obtenção da concentração em µg µL⁻¹.

A qualidade da amostra foi verificada com base na relação A_{260nm}/A_{280nm}, sendo que valores entre 1,7 e 2,0 indicam boa qualidade do RNA extraído. Uma amostra de RNA em tampão de corrida (Anexo B) foi ainda submetida à eletroforese sob corrente de 75 V por 45 minutos em gel de agarose 1% (p/v) com tampão TAE 1X (Anexo C) (SAMBROOK et al., 1989), para verificação da qualidade e também da quantidade.

O RNA total obtido foi separado em alíquotas de 50 µg, tanto para o RT-PCR teste, quanto para a construção da biblioteca SAGE. O RNA das alíquotas foi precipitado com 0,1 volume de acetato de sódio 2,5 M e 2 volumes de etanol 100% e armazenado em freezer à -80°C até o momento de sua utilização na construção da biblioteca SAGE.

3.3 Purificação do mRNA a partir do RNA total para o RT-PCR teste

O mRNA foi isolado com auxílio do *Dynabeads*[®] mRNA Purification Kit (Dyna), conforme recomendações do fabricante. A metodologia desse *kit* baseia-se no pareamento de bases entre resíduos poli A na extremidade 3' da maioria dos mRNAs e resíduos *oligo* dT covalentemente ligados à superfície dos *Dynabeads Oligo* (dT)₂₅. Dessa maneira, outros tipos de RNAs desprovidos de resíduos poli A não hibridizam aos *Dynabeads Oligo* (dT)₂₅ e são prontamente eliminados.

Ao RNA total, numa diluição de 50 µg em 100 µL de água MilliQ tratada com DEPC, acrescentou-se 100 µL de tampão de ligação (*Binding Buffer*) (Anexo D). A amostra foi aquecida a 65°C por 2 minutos para o rompimento das estruturas secundárias. Enquanto isso, 200 µL (1 mg) de *Dynabeads Oligo* (dT)₂₅ foram transferidos do estoque para um tubo *ependorf* de 1,5 mL de capacidade. Os *Oligo* (dT)₂₅ foram lavados uma vez com auxílio do *Dyna* MPC[®]-S (concentrador magnético de partículas). Essa lavagem foi feita removendo-se o tubo do *Dyna* MPC[®]-S, ressuspendendo os *Oligo* (dT)₂₅ em 100 µL de tampão de ligação (Anexo D) e recolocando o tubo no *Dyna* MPC[®]-S. Depois de 30 segundos, o sobrenadante foi removido e o tubo contendo os *Oligo* (dT)₂₅ limpos, retirado do *Dyna* MPC[®]-S.

Aos *Oligo* (dT)₂₅ adicionou-se 100 µL do tampão de ligação e a diluição contendo o RNA total. A homogeneização foi feita por rotação, durante 5 minutos, à temperatura ambiente, no agitador *sample mixer* (Dyna), que acompanha o *kit*.

Em seguida, a amostra homogeneizada foi colocada no *Dyna* MPC[®]-S e o sobrenadante removido após 30 segundos. Foram feitas duas lavagens com 200 µL de tampão de lavagem B (*Washing Buffer B*) (Anexo E), com auxílio do *Dyna* MPC[®]-S, da mesma forma descrita anteriormente, removendo-se completamente o sobrenadante entre cada lavagem.

Para a eluição do mRNA adicionou-se 15 µL de Tris-HCl 10 mM as esferas que foram aquecidas a 65°C por 2 minutos. Em seguida, o tubo foi colocado imediatamente no *Dynal MPC®-S*, o mRNA eluído foi transferido para um novo tubo, livre de RNase, e seguiu para a realização da reação de RT-PCR teste.

3.4 Reação de RT-PCR teste

Como a construção de bibliotecas SAGE baseia-se na síntese de cDNA, foram realizados testes simples de RT-PCR nas amostras de RNA obtidas, a partir das plantas de 6 meses de idade, a fim de verificar o seu potencial para uso na construção da biblioteca SAGE. Para isso, foi realizada uma reação de RT-PCR, que promoveu em uma única reação a síntese e amplificação do cDNA do gene de eucalipto que codifica a *UDP-glucose pyrophosphorylase*. Essa reação foi realizada com auxílio do *Kit SuperScript™ One-Step RT-PCR with Platinum® Taq* (Invitrogen), a partir do mRNA de plantas de 6 meses de idade e os *primers forward* e *reverse* específicos para a sequência codificadora desse gene (previamente desenvolvidos no laboratório).

Para um volume final de reação de 25 µL foram adicionados 12,5 µL de 2X *Reaction Mix* contendo 0,4 mM de cada dNTP e 2,4 mM de MgSO₄, 10 pg-1 µg de mRNA; 10 µM de cada *primer (forward e reverse)*; 1 µL de RT/ Platinum® Taq Mix e, água MiliQ autoclavada suficiente para completar o volume da reação. O programa utilizado no termociclador (*Applied Biosystem - GeneAmp®PCR System 9700*) para a realização da reação foi: 50°C por 1 minuto (síntese de cDNA), 94°C por 2 minutos (pré-desnaturação), 38 ciclos de 94°C por 45 segundos (desnaturação), 55°C por 1 minuto (anelamento do *primer*), 72°C por 2 minutos e 30 segundos (extensão), seguidos de uma extensão final de 5 minutos a 72°C.

A identificação do produto de amplificação, obtido na reação de RT-PCR, foi feita por meio de eletroforese em gel de agarose 1% (p/v) com tampão TBE 0,5X (Anexo F), submetido a corrente de 75 V por 30 minutos. As amostras foram acrescidas de solução *Dye IV* (Anexo G) (SAMBROOK et al., 1989) antes de serem aplicadas no gel. As bandas foram visualizadas sob luz ultravioleta por meio de coloração com brometo de etídio (0,01 ng µL⁻¹). O marcador de peso molecular 1 kb DNA *Ladder* (Invitrogen) foi utilizado para a verificação do tamanho do fragmento amplificado (Figura 5).

A partir do resultado apresentado pelo teste de RT-PCR, o RNA total extraído foi utilizado para a construção da biblioteca SAGE.

Figura 5 - Gel de eletroforese do produto do teste de RT-PCR realizado com a amostra de RNA total das plantas de 6 meses de idade usando *primers* para a *UDP-glucose pyrophosphorylase* de *E. grandis*. (P) Padrão de DNA (1Kb); (A) Amostra

3.5 Construção da biblioteca SAGE

A construção da biblioteca SAGE das plantas de 6 meses de idade foi realizada de acordo com o manual de instruções do *kit* I-SAGE™ – versão E da Invitrogen, o qual é baseado na metodologia de Análise Serial da Expressão Gênica (SAGE) originalmente desenvolvida por VELCULESCU *et al.* (1995), como descrito nos itens seguintes.

3.5.1 Isolamento do mRNA e síntese de cDNA

Um total de 50 µg de RNA total obtido das plantas de 6 meses de idade foi diretamente utilizado na construção da biblioteca SAGE. O primeiro passo da metodologia consistiu no isolamento do mRNA das amostras utilizando partículas esféricas magnéticas associadas à resíduos oligo(dT) (metodologia com mesmo princípio da descrita no item 3.3). 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 tubo. Removeu-se o sobrenadante e adicionou-se 500 µL do tampão de lise/ligação (*Lysis/binding buffer*) (Anexo H). Enquanto isso, as amostras de mRNA foram preparadas através da solubilização do precipitado de RNA total em 1000 µL do tampão de lise/ligação (Anexo H). Após o preparo das amostras, o tubo contendo as partículas magnéticas equilibradas no tampão de lise/ligação (Anexo H) 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 mRNA 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 mRNA que foram realizadas duas vezes com o tampão de lavagem A (*Washing buffer A*) (Anexo I), e uma última vez com o tampão de lavagem B (*Washing buffer B*) (Anexo J). Em seguida, a amostra foi preparada para a síntese da primeira fita das moléculas de cDNA através de 4 passos de lavagens sucessivas com o tampão de síntese da primeira fita (1X *First strand buffer*) (Anexo K). No quarto passo da lavagem, o sobrenadante somente foi removido no momento em que a mistura da reação para a síntese da primeira fita de cDNA (Anexo L) já estava pronta para adição do mRNA.

As partículas magnéticas ligadas às moléculas de mRNA 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 M) à reação e o tubo foi novamente colocado em banho-maria a 37°C por mais 1 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 N), 15 µL de dNTP *mix* (Anexo O), 5 µL de *E. coli* DNA ligase (Anexo P), 20 µL de *E. coli* DNA polimerase (Anexo Q) e 5 µL de *E. coli* RNase H (Anexo R). 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 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 S) 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 com 750 μL do tampão de lavagem D (*Washing buffer D*) (Anexo T). Posicionou-se o tubo no suporte magnético para remoção do tampão D e adição de 200 μL de 1X tampão 4 (Anexo U). A amostra foi então transferida para um novo tubo e armazenada a 4°C por 16 horas antes de seguir para a próxima reação.

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

A amostra de cDNA solubilizada em 1X tampão 4 foi precipitada com o auxílio do suporte magnético e ressuspensa na mistura de reação de digestão contendo 172 μL de LoTE (Anexo V), 2 μL de 100X BSA (Anexo X), 20 μL de 10X *buffer* 4 e 6 μL de *Nla* III (Anexo Z). 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 (Anexo S) para inativar a enzima, seguida de quatro lavagens com 750 μL do tampão de lavagem D (Anexo T).

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 ligase 1X (Anexo AA) 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 os 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 (Anexo V), 2 μL do tampão para ligase 10X concentrado, 1,5 μL do adaptador A (20 ng μL^{-1}) no tubo A e 1,5 μL do adaptador B (20 ng μL^{-1}) 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 AB).

As amostras foram então incubadas por 2 horas a 16°C e após este período foram lavadas 4 vezes com 500 μL do tampão de lavagem D (Anexo T). Posteriormente, as amostras foram lavadas 2 vezes com 200 μL do tampão 4 1X concentrado (Anexo U), 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 AC).

3.5.3 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 (Anexo V), 20 µL do tampão 4 10X concentrado e 4 µL de BSA 100X concentrado (Anexo X). 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 1 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 AD) 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 sequências curtas de cDNA + adaptadores, a partir das quais serão produzidas as *dítags* ou etiquetas duplas de cDNA que serão futuramente utilizadas na anotação dos genes encontrados. Assim, foi adicionado às amostras um volume de fenol:clorofórmio (1:1 v/v) 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 *musse*l glycogen e 1000 µL de etanol 100%. As amostras foram precipitadas em gelo seco por 10 minutos. O precipitado recuperado após a centrifugação a 16.000 x g por 40 minutos a 4°C foi lavado duas vezes com etanol 70% (v/v). Removeu-se todo o etanol dos tubos, os precipitados foram ressuspensos em 10 µL de LoTE (Anexo V) e incubados a 37°C por 3 minutos para garantir a solubilização.

Em cada um dos tubos (A e B) adicionou-se 5 µL do tampão *Klenow* (Anexo AE), 1 µL de BSA 100X (Anexo X), 2,5 µL da mistura de dNTP (Anexo O), 30,5 µL de água DEPC e 1,0 µL da polimerase *Klenow* (Anexo AF). 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 sequências de cDNA em dímeros. Posteriormente, foi realizada uma nova etapa de extração e precipitação dos dímeros com fenol:clorofórmio

(1:1 v/v) e acetato de amônia 7,5 M + 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 precipitado final nos dois tubos (amostra e controle) contendo os dímeros de cDNA + adaptadores foram ressuspensos em 1,5 µL de LoTE (Anexo V) 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 os tubos foram mantidos *overnight* a 16°C.

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

Após o período de incubação, as misturas foram diretamente utilizadas nas reações de otimização de PCR, após a adição de 14 µL de LoTE (Anexo V). As reações de otimização foram realizadas com 3 diferentes diluições da amostra a fim de determinar a diluição que produz quantidade suficiente de amplicons para prosseguimento do 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-SAGE*TM – versão E (Invitrogen), que contém concentração de 50-250 ng em 50 µL⁻¹ 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 de acordo com a Tabela 1.

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

Reagentes	Tubos					
	1	2	3	4	5	6
10X Tampão BV (Anexo AG)	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 <i>mix</i> (Anexo O)	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 ⁻¹)	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
<i>Platinum</i> [®] <i>Taq</i> 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

Notas: 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 µL) foram separados em gel de poliacrilamida 12% (Anexo AH), corados com brometo de etídeo (0,5 µg mL⁻¹) 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 1/20 (segunda coluna no gel) (Figura 6).

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

3.5.5 Amplificação em larga escala e purificação dos dímeros de cDNA

Foram realizadas 200 reações de PCR com volume final de 50 μL cada uma contendo: 5 μL de tampão BV 10X concentrado (Anexo AG), 3 μL de DMSO, 7,5 μL de dNTP *mix* (Anexo O), 2 μL de cada *primer* DTP-1 e DTP-2 (175 ng μL^{-1} em água DEPC, cada *primer*), 24 μL de água DEPC, 0,5 μL de *Platinum®Taq* DNA polimerase e 6 μL da amostra diluída 1/20. O programa de PCR utilizado foi o mesmo descrito anteriormente (Notas da Tabela 1). 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 v/v) e acetato de amônia 7,5 M + etanol 100% conforme descrito no item 3.5.3. O precipitado resultante foi ressuspenso em 300 μL de LoTE (Anexo V) 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%(Anexo AH), por 4 horas a 130 Volts. Durante a etapa de formação 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 cada molécula de cDNA) e baseando-se neste princípio selecionou-se no gel a banda de aproximadamente 100 pb para posterior polimerização. No gel foi 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. (1 – 9) Amostras de dímeros de cDNA; (P) Padrão de DNA. Todas as bandas de 100 pb (círculo) foram excisadas do gel para geração das *ditags*

As bandas correspondentes a 100 pb foram recortadas do gel utilizando um bisturi e colocadas em tubos de 0,7 mL (furados no fundo) encaixados em tubos de 1,7 mL para centrifugação a $14.000 \times g$ 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 então dissolvidos em 150 μ L de LoTE (Anexo V) e acetato de amônia 7,5 M na proporção 125:25 (v/v).

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

3.5.6 Digestão dos dímeros de cDNA com *Nla* III

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

O primeiro passo nesta fase consistiu da 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 (Anexo X), 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 e após este período, o volume de cada tubo foi ajustado para 200 μL com LoTE (Anexo V). Em seguida, procedeu-se a extração e precipitação das *ditags* utilizando fenol:clorofórmio (1:1 v/v) e acetato de amônia 7,5 M + etanol 100%, respectivamente, conforme descrito no item 3.5.3. Os precipitados dos três tubos foram ressuspensos em LoTE (Anexo V) e combinados em um único tubo com volume final de 32 μL .

As *ditags* de 28 pb foram separadas dos produtos não digeridos em gel de poliacrilamida 12% (Anexo AH), por 4 horas a 130 Volts. Para isto, foram adicionados 7 μL do tampão de amostra TBE (Anexo AI) ao tubo contendo 32 μL da solução de *ditags* e todo volume da solução foi aplicado em três colunas no gel. Após a eletroforese, o gel foi corado com brometo de etídio ($0,5 \mu\text{g mL}^{-1}$) e visualizado sob luz UV permitindo assim a excisão das bandas de 28 pb (Figura 8).

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

Os fragmentos de gel foram quebrados em pequenos pedaços, usando o mesmo procedimento já descrito para a purificação dos dímeros de cDNA (item 3.5.5), e homogêneos em uma solução contendo 150 μL de LoTE + acetato de amônia 7,5 M (125:25 v/v). O tubo contendo os fragmentos em solução foi incubado a 37°C *overnight*.

3.5.7 Polimerização/Concatenação e clonagem das *ditags*

A concatenação das *ditags* resulta na formação de sequências maiores de cDNA que são utilizadas na clonagem de forma 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 submetidas à reação de ligação descrita a seguir.

A solução eluída do gel foi purificada em colunas S.N.A.P.TM (segundo normas do fabricante, Invitrogen) e o volume resultante dividido em volumes de 200 µL, que foram distribuídos 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 em velocidade máxima por 30 minutos a 4°C e os precipitados resultantes lavados com etanol 70% e ressuspensos em um único tubo com 7,75 µL de LoTE (Anexo V).

Adicionou-se a solução contendo 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 da 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 à 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*. (A) Concatâmeros da amostra; (P) Padrão de DNA. 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 no item 3.5.5. Cada fração eluída do gel foi então solubilizada em dois tubos contendo 200 μL de LoTE + acetato de amônia 7,5 M (5:1 v/v) 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 (segundo normas do fabricante, Invitrogen), centrifugado por 2 minutos em 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 no item 3.5.3 e o precipitado resultante foi solubilizado em 6 μL de LoTE (Anexo V).

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 $\mu\text{g } \mu\text{L}^{-1}$) (Anexo AJ), 2,5 μL do tampão 2 10X concentrado (Anexo AK), 1,4 μL da enzima *Sph* I (Anexo AL) e 19,1 μL de água estéril. A solução foi incubada a 37°C por 30 minutos e após este período procedeu-se a extração e precipitação do vetor linearizado. O precipitado resultante foi ressuscitado em 60 μL de LoTE (Anexo V), sendo 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	-	-
pZErO [®] -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 v/v) e etanol 100%. Os *pellets* foram ressuscitados em 12 μL de LoTE (Anexo V) e utilizados na transformação das células de *E. coli* para construção da biblioteca SAGE.

Um volume de 2 μL da mistura de ligação foi utilizado para transformar 50 μL de células competentes de *E. coli One Shot TOP10* (Invitrogen). A mistura de células e vetores foi transferida para uma cuveta de 0,1 cm e procedeu-se a eletroporação com um pulso elétrico de 2,5 kV por 4 ms. Após a eletroporação, adicionou-se 250 μL do meio SOC (Anexo AM) 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 AN) e o agente de seleção ZeocinaTM (50 $\mu\text{g mL}^{-1}$).

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 os concatameros (aproximadamente 150 clones/placa). Assim, todo o volume restante foi também plaqueado e as colônias formadas foram selecionadas para o sequenciamento.

Além disso, 26 clones resistentes à Zeocina foram selecionados, tiveram seus plasmídeos extraídos, foram submetidos a reação de PCR utilizando os *primers* M13 *foward* e *reverse* e aplicados em gel de agarose 1% (p/v) para verificar o tamanho médio dos insertos que foram clonados. Através do perfil eletroforético desses clones é possível verificar que a maior parte dos fragmentos encontra-se dentro da faixa de tamanho esperada (500pb – 1Kb) (Figura 10).

Figura 10 - Perfil eletroforético dos clones testados

3.6 Sequenciamento dos polímeros de *ditags*

A extração de DNA plasmidial dos clones e o 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).

3.7 Bioinformática e análise dos dados

O estabelecimento de uma associação correta entre cada *tag* gerada e a sequência de cDNA ou EST de origem é o primeiro passo no estabelecimento de perfis de expressão gênica a partir dos dados de sequenciamento gerados nas bibliotecas SAGE. Dessa forma, para análise dos dados no presente trabalho foram utilizados 4 critérios imprescindíveis: i) seleção de sequências com qualidade de sequenciamento igual ou superior a $q \geq 20$, ii) seleção das *tags* que apresentaram número de cópias igual ou superior a 2, iii) utilização de um banco de dados de ESTs de origem específica de espécies de eucalipto, além do banco mundial de seqüências públicas (NCBI) iv) 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'. Todos os critérios adotados na geração dos dados bem como na sua análise serão discutidos nos próximos itens.

3.7.1 Seleção de sequências para análise

Com a finalidade de trabalhar com o maior número possível de *tags* a serem obtidas, e estas apresentarem uma alta confiança, antes da extração de todas as *tags* pelo software do SAGE, os dados gerados pelo sequenciamento foram avaliados quanto a sua confiabilidade através do programa *Phred* (EWING et al., 1998), em ambiente *Linux*. Este programa é capaz de ler os sinais de fluorescência dos cromatogramas gerados pelo sequenciador automático, determinar as posições ideais dos picos de cada base, comparar as localizações dos picos detectados com as dos picos idealizados, quantificar 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 atribuir valores de qualidade às bases preditas 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) (Tabela 3).

Tabela 3 - Valores de qualidade relacionados à probabilidade de erro de acordo com o programa *Phred* (EWING et al., 1998)

Qualidade	Probabilidade de erro em uma base	Acurácia
10	1 em 10	90%
20	1 em 100	99%
30	1 em 1.000	99,9%
40	1 em 10.000	99,99%
50	1 em 100.000	99,999%

No presente trabalho, foram utilizadas somente as sequências que apresentaram qualidade igual ou superior a 20, o que significa que 1 em cada 100 pb sequenciadas possa estar errada, proporcionando uma probabilidade de atribuição de bases corretas de 99%. Todas as sequências que apresentaram qualidade inferior a 20 foram sequenciadas novamente até a obtenção de sequências com qualidade adequada (Figura 11).

Figura 11 - Cromatogramas gerados por dois eventos de sequenciamento de um mesmo clone. (A) Cromatograma de baixa qualidade, onde é possível observar a ocorrência de sobreposição de picos e áreas pouco definidas; (B) Cromatograma referente ao re-sequenciamento da mesma sequência de (A), apresentando alta qualidade, sendo ainda possível observar a ocorrência de picos não sobrepostos (bem definidos)

3.7.2 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 400 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 de bibliotecas SAGE. O programa importa os arquivos de texto, identifica os sítios de restrição da enzima *Nla* III (CATG), extrai as *tags* usando os sítios reconhecidos como referência e contabiliza o número de vezes que cada *tag* aparece na biblioteca. Ao final, o programa lança um arquivo de saída ou *tag list* contendo todas sequências de 10 pb (*tags*) seguidas pelo seu nível de expressão, que representa, portanto, o número de vezes em que aquela *tag* foi sequenciada na biblioteca (Figura 12).

Report File Name = C:\Documents and Settings\Mayra\Meus documentos\SAGE\Rafa\Todas sequências 14.12.2005\rep-3.rpt
Date: 01-08-2007

Project # 1
Project File = C:\Documents and Settings\Mayra\Meus documentos\SAGE\Rafa\Todas sequências 14.12.2005\TAGBANK.sum
SearchName = TAGBANK
Anchoring Enzyme = NlaIII - CATG
Tag Length = 10
DiTag Length = 26
Total Files = 826
Total Tags = 15717
Total Duplicate Dimers = 4827
Total tags in selected Projects = 15717

Tag Abundance Report
MS Access File Name = C:\Documents and Settings\genfis\Desktop\0.MDB

Total tags after excluding tags = 15717

Count	Percent	Tag Sequence	Tag BaseFour Number
126	0,8016	AAITTCAC	64851
Genes in Class = 1 * Tags in Class = 126 Cumulative Gene Count = 1 * Cumulative Tag Count = 126			
107	0,6807	TCGTGTTTC	978687
Genes in Class = 1 * Tags in Class = 107 Cumulative Gene Count = 2 * Cumulative Tag Count = 233			
99	0,6298	CAGTTAAAG	310275
Genes in Class = 1 * Tags in Class = 99 Cumulative Gene Count = 3 * Cumulative Tag Count = 332			
76	0,4835	TTCGGTTTC	1010678
Genes in Class = 1 * Tags in Class = 76 Cumulative Gene Count = 4 * Cumulative Tag Count = 408			
73	0,4644	GGATTGCATA	671309
Genes in Class = 1 * Tags in Class = 73 Cumulative Gene Count = 5 * Cumulative Tag Count = 481			

Figura 12 - Visualização do arquivo de saída do *software* SAGE2000TM (versão 4.5) ou *Tag list*, contendo uma lista de *tags* sequenciadas na biblioteca e seu respectivo nível de expressão

3.7.3 Identificação e anotação das sequências de origem das *tags* e suas prováveis funções celulares

Todas as sequências de *tags* que apresentaram nível de expressão ≥ 2 cópias na biblioteca foram utilizadas para encontrar as sequências de cDNA que lhes deram origem, no banco de dados de ESTs do projeto Genolyptus e/ou no banco de dados públicos do NCBI.

O projeto Genolyptus (Rede Brasileira de Pesquisa do Genoma do *Eucalyptus*) foi desenvolvido 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. Esse projeto foi formado a partir do sequenciamento de aproximadamente 135.000 ESTs que dentro do Laboratório Max Feffer resultaram na construção de aproximadamente 23.292 *contigs* representativos de amostras de folhas, plântulas, raízes, xilema, floema e flores de diversas espécies de eucalipto como, *E. globulus*, *E. grandis*, *E. pellita* e *E. urophylla*. Já o banco de dados do NCBI apresentava até início de abril de 2007 cerca de 15.000 sequências de nucleotídeos e 480 sequências de proteínas para diversas espécies de eucalipto.

O contraste entre as 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), dessa forma, as *tags* serviram como base para a busca. 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 existe ou não uma cauda poli(A) 3' indicando o término da sequência. Para confirmar a identificação, cada sequência encontrada passou por alguns critérios: i) os 10 pb referentes a *tag* deveriam estar posicionados imediatamente após o sítio de restrição da enzima *Nla*III (CATG); ii) ambos deveriam estar o mais próximo da extremidade 3', uma vez que de acordo com a técnica SAGE, o RNA mensageiro deve ser clivado no sítio de restrição CATG mais próximo da cauda poliA (Figura 13). Dessa forma, todas as sequências de ESTs encontradas foram analisadas manualmente e somente aquelas

que obedeceram aos critérios acima mencionados foram utilizadas na anotação das *tags*.

```
>Contig1355ur-xy (whole contig)
5' CACTAACCATCTGCCTTTCTTCATCTTCTTTCTTCTGCTTCTACTCCGTTTCTCGTTTCGATATCGTGAAAGGAGTCCGTCGACGACA
ATGGCCGAGAAAGAGCAAGGTCTTGATCATCGGAGGGACGGGCTACATCGGCAAGTTTCATCGTCGAAGCGAGTGCAAAAGCAGGGCATCCCA
CGTTCGCGCTGGTTAGGCAGAGCACGGTCTCCGACCCCGTCAAGGGCCAGCTCGTCGAGAGCTTCAAGAACTTGGGCGTCACTCTGCTCAT
CGGTGATCTGTACGATCATGAGAGCTTGGTGAAGGCAATCAAGCAAGCCGACGTGGTGATATCGACAGTGGGGCCATGCAAATGGCGGAT
CAGACCAAGATCGTCGACGCCATTAAAGAAGCTGGCAACGTTAAGAGATTCTTTCCTTCCGAATTCGGCAACGATGTGGACAGGGTCATG
CTGTGGAGCCAGCGAAGTCTGCTTTTGAATTGAAGGCCAGATCCGCCGTGCCGTGGAGGCGGCAGGCATCCCTTACACCTACGTCCCATG
CGGCTGCTTCGCCGGCTACTTCCCTCCCAACACTGGCGCAGCCAGGAGTCACTGCTCCTCCGAAGGACAAAGTCACCGTCATGGGCGACGGA
AATGCAAAGGCAATTTTCAACAAGGAAGATGACATTGCGGCCTTCACCATCAAGGCTGTGGATGATCCGAGATCGCTGAACAAGATCCTTT
ACATCAGGCCCTCCTAAGAAGCTTTACTCATTTCAATGAGCTTGTGTCCTTGTGGGAGAAGAAAATTGGCAAGACCTCGAGAAGATTACCT
TCCTGAAGAGCAAATCCTGAAGCAAATCCAGGAGTCCCAATTTCCCATCAATGTCATATTAGCAGTGAACCATTAATCTTTGTTAAGGC
GACGGTGCCAAATTTTGAATCGAGGAGTCTTTTGGTGTGAGGCTTCTGAGCTGTACCCAGATGTGAAGTACACTACAGTGAAGAATACC
TCGAAAATTTTGTCTAAATTAAGGCCCATGCGTCTCCTGTTCTTCAAGGAGTGAGTTACCGTGACTCTGGTGGACAGTTGATATGTATTAAA
AGGCTGTACACCTAAAGAAATCAAGGTCACGGTCTTATTTAGAATTGTCTCTGATGTCATATTCTTCTTGGTCTTCTGGACCATGTATT
TGCTTTCTCTTGTGTGGTATCCATGAATTTCCAGGTTGTTGAAATTAACCTTGTCTGTCTGATATATTTTCATCTCTCAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA3'
```

Figura 13 - Sequência ilustrativa de um EST utilizado na anotação da *tag* mais abundante na biblioteca. Em negrito estão representados todos os possíveis sítios de restrição da enzima *Nla* III, em itálico e sublinhado encontra-se a sequência de 10 pb da *tag* específica utilizada no BLAST-N. Logo após o último sítio CATG e na extremidade 3' da sequência é possível observar a presença da cauda poli(A) terminal

As sequências de ESTs representativas de cada *tag* foram analisadas tanto através do programa BLAST-X (ALTSCHUL et al., 1997), quanto pelo programa *ORF Finder* (*Open Reading Frame Finder*), utilizando o banco de dados do NCBI como referência.

O BLAST-X é usado para comparar sequências de nucleotídeos traduzidas em todos os seis 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* (Figura 14). 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 de um evento ao acaso (ALTSCHUL et al., 1997). Neste trabalho, apenas alinhamentos com valores de $E \leq e-05$ foram utilizados na anotação das *tags*.

Figura 14 - Resultado ilustrativo gerado pelo BLAST-X para o EST associado a tag mais abundante

O *ORF Finder* é uma ferramenta do NCBI capaz de identificar todos os quadros abertos de leitura possíveis para uma dada sequência de nucleotídeos. A sequência de nucleotídeo traduzida em proteína possui um quadro de leitura que inicia com o códon de iniciação ATG (metionina) e se estende através de uma série de trincas representando aminoácidos até terminar em um dos três tipos de códon de terminação (TAA, TAG, TGA).

A partir dos resultados obtidos pelo BLAST-X e pelo *ORF Finder* foi possível obter a sequência completa da proteína, assim como, buscar o domínio funcional (Figura 15).

Para determinar a função das proteínas identificadas foram utilizadas diversas ferramentas computacionais de análise de sequências presentes em diferentes bancos de dados. O uso dessas ferramentas é essencial para atribuição de funções já que apenas uma pequena parcela das proteínas atualmente conhecidas apresenta suas funções determinadas experimentalmente (SCHULTZ et al., 1998). As ferramentas computacionais são capazes de reconhecer os domínios conservados das proteínas mediante a presença de padrões de aminoácidos. Os domínios das proteínas são

unidades distintas de evolução molecular e representam uma discreta unidade de estrutura molecular tridimensional. Os domínios protéicos são usualmente associados com um aspecto molecular particular, tais como, catálise e ligação e podem fornecer uma indicação da função molecular e celular da proteína (MARCHLER-BAUER et al. 2007). A ferramenta CDD (*Conserved Domain Database*) (MARCHLER-BAUER et al. 2005) presente no site do NCBI contém mais de 12.000 modelos de domínios de proteínas importados de outros bancos como Pfam (BATEMAN et al., 2002; FINN et al., 2006), SMART (SCHULTZ et al., 1998; LETUNIC et al., 2006) e COG (TATUSOV et al., 2003). Já o KEGG é um banco de dados que integra informações moleculares e sistemas de informações biológicas. Esse banco é muito útil, pois possui mapas das vias metabólicas, que representam as interações moleculares e a rede de reações para o metabolismo e outros processos celulares (KANEHISA et al., 2006).

Figura 15 - Resultado ilustrativo para análise do EST associado a *tag* mais abundante, apresentando a sequência completa da proteína gerada pelo BLAST-X, e o domínio funcional obtido pelo programa CDD

3.7.4 Classificação funcional

Para realizar a análise funcional e permitir uma visão mais dinâmica dos genes expressos na biblioteca das plantas de 6 meses de idade, os genes anotados foram divididos em categorias funcionais, seguindo a classificação originalmente proposta por RISON et al. (2000) com pequenas modificações como a inclusão das subcategorias transporte de vesículas e fatores transportadores na categoria transporte, e a inclusão da subcategoria citoesqueleto na categoria componentes estruturais (Tabela 4). 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 4 - Classificação Funcional adaptada a partir de RISON et al. (2000)

1 Metabolismo e Energia	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 as 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	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 / Destoxificação	5.3.3 Relacionada a Tradução
2.2.3 Respostas à 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 Nucléicos	6.1 Sequência Expressa
3.2 Pequenas Moléculas	6.2 Proteína de Função Hipotética
3.2.1 Canais de Íons / Porinas	6.3 Sem Homologia (Sem Registro)
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	

Finalmente, todos os dados produzidos pelo sequenciamento da biblioteca SAGE e que passaram pelas etapas de identificação, anotação, classificação funcional e análise foram armazenados na forma de “fichas eletrônicas” em um banco de dados construído localmente no Laboratório Max Feffer de Genética de Plantas (<http://genfis20.esalq.usp.br:8008/web/>) com o auxílio dos outros membros do grupo e da especialista em informática Livia M. Franceschini.

Dessa forma, para cada *tag*/gene existe uma “ficha eletrônica” onde podem ser encontradas todas as informações referentes aquela *tag*, como por exemplo: grupo a que pertence, número de cópias, resultado do BLAST-X, domínio funcional, categoria funcional e a sequência (Figura 16).

Eucalyptus Database

Laboratory: Max Effertz > Eucalyptus Database > SAOE Data Base > Search Result

USSP ESALQ
[Home](#) | [About](#) | [Contact](#)

Registered or Modified on: Tue Oct 24 14:28:22 BRST 2006 Current Date and Time is: Tue Jan 09 09:25:24 BRST 2007

Search Result

TAG: AATTTCOCAG Delete: Tag Delete? No

Registered Groups:	Cambium Leaf Plectet	Groups that TAG is registered:	Plectet
--------------------	----------------------	--------------------------------	---------

Description: refa 126

Libraries where the Tag is found: Cambium 1-6 months

Libraries Registered: Cambium 1-6 months
Rust Sensitive
Cambium 5-7 years
Rust Resistant
Cambium 1-3 years

Blast Result: gi|10764491|gb|AA022740.1|
allergenic isoflavone reductase-like protein ...
S18 q-145

Functional Domain: pfam02716, Isoflavone reductase

Category: 1.7 - SECONDARY METABOLISM

Sequence: >Contig1355ur-xy (whole contig)
CACTAACCATCTGCCTTCTTTCATCTCT
TTCTCTGCTGTACTGCGTTTCCTGGTT
TCGATATGCTGAAGACGTCCTGCAGA

[Register](#) [Alter](#) [List](#)

Status: Successful Search

< back home top

Figura 16 - Fichas “eletrônicas” do banco de dados de genômica funcional de *E. grandis* do Laboratório Max Feffer de Genética de Plantas

3.8 Correlação entre nível de transcritos e abundância de proteínas

Com o objetivo de estabelecer uma comparação entre o nível de transcritos obtidos pela técnica de SAGE e a abundância de proteínas presente nas plantas de eucalipto de 6 meses de idade, foi iniciada a análise do proteoma do mesmo material vegetal, utilizado para a construção da biblioteca SAGE. A combinação de técnicas de análise da transcrição (como a técnica de SAGE) e de análise do proteoma (como 2D-PAGE e espectrometria de massas), podem ajudar na compreensão dos mecanismos de regulação e controle envolvidos entre o gene e uma proteína, quando aplicados no estudo de uma mesma amostra.

A parte do trabalho referente ao estabelecimento do proteoma foi realizada paralelamente por outro aluno de doutorado (Alexander de Andrade) do Laboratório Max Feffer (ANDRADE, 2006). Resumidamente, as plantas tiveram suas folhas, raízes e cascas removidas, seus caules foram macerados em almofariz com nitrogênio líquido e as proteínas extraídas pelo método de extração fenólica. As proteínas foram separadas por eletroforese bidimensional em fitas IPG com gradiente de pH imobilizado linear de 4-7 na primeira dimensão e gel de poliacrilamida (12,5%) na segunda dimensão. A coloração dos géis foi realizada com coomassie G250. Foram detectados 438 *spots* e um total de 168 *spots* foram retirados do gel, digeridos com tripsina e submetidos ao sequenciamento por espectrometria de massas através do sistema LC-MS/MS.

Os espectros de massas das amostras foram analisados e processados utilizando os programas MassLynx NT BioLynx V 4.0 e ProteinLynx V 2.1 da Micromass e submetidas para sua identificação em dois bancos de dados SWISS-PROT (BAIROCH et al., 2005) e Genolyptus, através do programa BLAST (ALTSCHUL et al., 1997).

A comparação entre os transcritos obtidos pela biblioteca SAGE e as proteínas identificadas pelo sequenciamento foi realizada a partir do nível de expressão dos transcritos (número de cópias) para um determinado gene e o volume relativo médio dos *spots* de três géis de pH 4-7 analisados pelo programa Image Master 2D versão 3.10 (Amersham Pharmacia Biosciences).

4 RESULTADOS E DISCUSSÃO

4.1 Análise geral dos dados produzidos pela biblioteca SAGE

A biblioteca SAGE construída a partir de plantas de *Eucalyptus grandis* de 6 meses de idade foi produzida através do sequenciamento de 826 clones. A eficiência do sequenciamento, isto é, o número de *tags* sequenciadas por clone foi em média de 19 *tags*/clone gerando, dessa forma, um total de 15.717 *tags* sequenciadas. Para evitar a utilização de *tags* produzidas a partir de possíveis erros de sequenciamento, a busca e a identificação dos ESTs correspondentes a cada *tag* foi realizada somente para aquelas *tags* presentes em número igual ou superior a 2 cópias na biblioteca. Dessa forma, somente 9.810 *tags* foram consideradas confiáveis para realização das análises posteriores, sendo que a partir deste grupo final de *tags*, apenas 2.274 corresponderam a *tags* ou “genes” diferentes (*tags*/genes). As *tags* foram buscadas em bancos de sequências públicas de ESTs de eucalipto, como por exemplo, o banco de dados do NCBI e também entre as sequências de eucalipto produzidas pelo consórcio Genolyptus. As buscas permitiram a identificação de 43,5% das sequências analisadas, o que equivale à identificação de 989 genes. Desses 989 genes que foram identificados nos bancos de dados, 702 (71%) puderam ter uma possível função atribuída, ou seja, os ESTs correspondentes estão relacionados a uma determinada função metabólica indicada no NCBI. Todavia, 13,9%, ou seja, 138 genes, foram considerados “Sequências Expressas”, uma vez que não foram encontradas funções metabólicas específicas para seus ESTs. Já, 149 genes (15,1%) foram considerados “Proteínas de Função Hipotética”, pelo fato de seus ESTs estarem relacionados a genes no NCBI cujas funções ainda são desconhecidas (Tabela 5).

Entre os 2.274 genes, 56,5%, ou seja, 1.285 genes foram classificados como “Sem homologia” ou “Sem registro”, sendo que este grupo corresponde as *tags* para as quais não foi possível encontrar um EST de origem nos bancos públicos de eucalipto (NCBI) ou no consórcio Genolyptus (Tabela 5).

Tabela 5 - Resumo dos dados gerais produzidos pela biblioteca SAGE

Dados Gerais	Biblioteca
Total de Clones Sequenciados	826
Tamanho da Biblioteca (<i>Tags</i> Sequenciadas)	15.717
Eficiência (<i>Tags</i> /Clone)	19
<i>Tags</i> com ≥ 2 Cópias	9.810
Total de Genes com ≥ 2 Cópias	2.274
Genes Identificados	989 (43,5%)
Genes com Função Definida	702
Sequências Expressas	138
Proteínas de Função Hipotética	149
Sem Homologia (Sem Registro)	1.285 (56,5%)

Um aspecto importante que pode ser notado é a existência de um pequeno número de *tags*/genes que apresentam um alto nível de expressão e um grande número de *tags*/genes que apresentam um baixo nível de expressão. Assim, os genes que possuem até 5 cópias somam 1.907 sendo responsáveis por 83,8 % do total de genes considerados para análise. Isso representa 52 % do total de *tags* com ≥ 2 cópias (transcritos). Já os genes que possuem um maior nível de expressão, como por exemplo, os presentes em número ≥ 5 cópias somam 367 e são responsáveis por apenas 16,2 % do total de genes considerados para análise, mas representam 48 % do total de *tags* com ≥ 2 cópias (transcritos). Isso é habitual quando se utiliza a técnica de SAGE e comprova a vantagem da técnica em permitir o acesso a transcritos de baixo e alto nível de expressão ao mesmo tempo. Na figura 17 podemos ver o panorama da distribuição do número de *tags*/genes em relação ao número de cópias (transcritos)

Figura 17 - Distribuição do número de cópias para os 2274 genes com número de cópias ≥ 2

A identificação de 43,5% das sequências analisadas está em concordância com resultados obtidos por outros pesquisadores que empregaram a metodologia SAGE para análise de transcrito em diversas espécies. Os resultados de alguns desses trabalhos revelaram, por exemplo, 50% de identificação, no caso de 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), 70% de identificação a partir de células gástricas cancerígenas em humanos (LEE et al., 2003), 97% em folhas de cana-de-açúcar (CALSA; FIGUEIRA, 2007) e mais recentemente 41,5% e 35% da região cambial e de folhas de eucalipto, respectivamente (CARVALHO, 2007). A porcentagem relativamente baixa de identificação de genes através da utilização da técnica de SAGE está relacionada principalmente à pequena representatividade de ESTs de baixa frequência nos diferentes bancos de dados. Por exemplo, no início do sequenciamento de ESTs de humanos, um grande número de novos genes foram identificados (cerca de 10,4% dos ESTs sequenciados em 1996 representavam novos genes). Entretanto, conforme o avanço do sequenciamento, o número de novos genes identificados passou a diminuir, de forma que 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, J et al., 2002). Já, quando se passou a utilizar as *tags* geradas pelo método SAGE nos estudos de expressão gênica em humanos, observou-se que não era possível encontrar nos bancos de dados as respectivas sequências de cDNAs ou ESTs para um grande número dessas *tags* (de JONGE et al., 2003). Contudo, associando a pequena representatividade de ESTs de baixa frequência nos bancos de dados, o sucesso do uso de SAGE na geração de novos genes ainda não identificados em humanos (CHEN, J et al., 2002) e o fato 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), é possível notar uma diferença na sensibilidade do método utilizado na construção dos bancos de dados (EST) em relação ao método SAGE, resultando, portanto, em alguns casos, na dificuldade de identificação das SAGE *tags* em bancos de ESTs.

Na tentativa de driblar as dificuldades de identificação das SAGE *tags*, alguns pesquisadores têm utilizado bancos genômicos para geração de “*tags* virtuais” que são então comparadas as *tags* obtidas experimentalmente. Utilizando essa estratégia, Fizames et al. (2004) conseguiram identificar 57% dos genes analisados a partir da biblioteca de raízes de *Arabidopsis* e ROBINSON et al. (2004), 79% dos genes analisados a partir da biblioteca de folhas de *Arabidopsis*. Entretanto, essa abordagem é limitada em alguns casos devido à ausência de bancos genômicos para muitas das espécies de interesse, como é o caso da espécie de eucalipto utilizada no presente estudo.

Com relação aos dados gerados neste trabalho é importante ressaltar que na época em que o sequenciamento da biblioteca SAGE foi finalizado, haviam poucas informações de ESTs de eucalipto disponíveis nos bancos de dados públicos. Dessa maneira, as buscas realizadas para identificação e anotação das SAGE *tags* tiveram como base o banco de ESTs do consórcio Genolyptus, o qual foi disponibilizado ao Laboratório Max Feffer de Genética de Plantas apenas para pesquisa e cujas informações não são disponibilizadas ao público. As buscas de *tags* não identificadas, ou seja, cujos ESTs de origem não foram identificados no banco de ESTs do consórcio Genolyptus, bem como, das *tags* correspondentes aos genes importantes para a formação da madeira, só foi possível a partir de 2006 com a publicação no banco de dados públicos do NCBI de um grande número de sequências de ESTs para amostras de xilema de *Eucalyptus gunnii*. Embora todo este processo de refazer as buscas tenha demandado um enorme tempo e esforço, uma vez que estas são realizadas manualmente, uma a uma, essa experiência foi de certa forma relevante, já que foi possível confirmar a identificação e a anotação previamente realizada para um grande número de *tags* a partir de uma fonte diferente de sequências. A grande importância desta confirmação está no fato de que quando o organismo utilizado na geração do banco de referência é o mesmo a partir da qual as *tags* foram geradas, ela demonstra o poder e a eficiência da técnica de SAGE no acesso a transcritos de forma única, mesmo sendo as SAGE *tags* produzidas a partir de sequências curtas de 10 pb. Sendo assim, à medida que mais sequências genéticas de eucalipto sejam disponibilizadas em bancos de dados públicos e, principalmente, com a possível conclusão do

sequenciamento do genoma completo que já está em andamento, o número de *tags* identificadas pode aumentar de maneira considerável.

4.2 Perfil transcricional dos dados produzidos pela biblioteca SAGE

A partir dos dados produzidos pela construção da biblioteca SAGE, todos os 989 genes que puderam ser identificados nos bancos de dados foram agrupados em categorias funcionais, de modo que esses genes pudessem ser analisados sob uma perspectiva dos processos biológicos e/ou metabólicos que desempenham.

A partir da distribuição dos genes encontrados em categorias funcionais foi possível observar que 19% pertencem à categoria “Estrutura e Organização da Estrutura Celular”, 18% dos genes pertencem à categoria “Metabolismo e Energia”, 15% à categoria “Vias de Informação”, 12% à categoria “Processos Celulares”, 7% à categoria “Transporte” e 29% dos genes pertencem à categoria “Sem Função Definida” (Figura 18).

Figura 18 - Distribuição dos genes identificados a partir da biblioteca SAGE, de plantas de eucalipto com 6 meses de idade, em categorias funcionais. A categoria “Sem Função Definida” foi constituída pelas *tags* classificadas como “Sequências Expressas” e “Proteínas de Função Hipotética”, ou seja, para as quais foram encontrados os prováveis ESTs de origem mas, não foi possível identificar a provável função destas sequências através do NCBI

Dentro da categoria “Estrutura e Organização da Estrutura Celular”, os genes mais representativos são aqueles relacionados à formação da parede celular, ao citoesqueleto e a proteínas e RNA ribossomais. Na categoria “Metabolismo e Energia” a maior parte dos genes está relacionada à respiração e ao metabolismo de aminoácidos, carboidratos e lipídeos. A grande maioria dos genes encontrados na categoria “Vias de Informação” está relacionada à síntese/modificação/degradação de proteínas, o que concorda com o elevado número de genes para RNA e proteínas ribossomais. Na categoria “Processos Celulares”, um grande número de genes esteve associado a respostas de proteção/destoxificação e também transdução de sinais. Já para a categoria “Transporte”, a maior parte dos genes esteve relacionado ao transporte de vesículas e ao transporte de pequenas moléculas.

A *tag* mais expressa na biblioteca, das plantas de eucalipto com 6 meses de idade, apresentou um número de cópias equivalente a 126. Essa *tag* identificou com alta similaridade um EST referente a uma *Isoflavone reductase* (1e-145) (Tabela 6) e a *Phenylcoumaran benzylic ether reductase* ou PCBER (2e-135). Ambas enzimas são redutases dependentes de NADPH e estão envolvidas na biossíntese de lignanas e isoflavonóides (SHOJI et al., 2002). Essa foi também a *tag*/gene mais expressa em uma biblioteca SAGE da região cambial de madeira juvenil de eucalipto (CARVALHO et al., 2007). É interessante observar que alguns estudos demonstraram uma correlação positiva entre a expressão dessas enzimas redutases dependentes de NADPH e a expressão de enzimas chaves na biossíntese de lignina tais como a PAL, COMT e CCoAOMT (VINCENT et al., 2005), assim como, no crescimento secundário de árvores transgênicas de *Populus* (ISRAELSSON et al., 2003). Além disso, Shoji et al. (2002) demonstraram a expressão de PCBER em células do caule durante crescimento secundário ativo e também sua correlação com a produção de compostos secundários adicionados à parede celular em lignificação, levando a formação do cerne e promovendo uma fonte adicional de defesa contra o ataque de patógenos em tabaco. Dessa maneira, muito provavelmente a alta expressão de transcritos para redutases dependentes de NADPH na biblioteca de plantas de eucalipto com 6 meses de idade, indique a ocorrência de crescimento secundário e deposição de compostos secundários nas células do xilema dessas plantas de eucalipto.

Tabela 6 - Lista dos 50 genes (*tags*) mais expressos na biblioteca SAGE de plantas de eucalipto com 6 meses de idade. Para cada *tag* é fornecido o número de cópias na biblioteca, a identificação do gene (BLAST-X) bem como o valor *E* para cada identificação. As *tags* “Sem Homologia” são aquelas para as quais não foi possível encontrar o EST ou o cDNA de origem, já as *tags* “Sequência Expressa” representam aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente, porém apresentaram valor $E \geq e-05$

(continua)

TAG	Nº de cópias	Blast(x) – NCBI	E value
1	126	gi 10764491 allergenic isoflavone reductase-like protein	e-145
2	107	Sequência Expressa	-
3	99	Sem Homologia	-
4	76	Sem Homologia	-
5	73	gi 56311394 ADP-ribosylation factor 1	9e-100
6	63	Sequência Expressa	-
7	60	Sem Homologia	-
8	57	gi 34913256 unknown protein	1e-29
9	56	gi 120676 Glyceraldehyde-3-phosphate dehydrogenase	3e-21
10	54	Sem Homologia	-
11	48	gi 15236267 PIP3 (Plasma Membrane Intrinsic)	2e-148
12	41	Sem Homologia	-
13	37	Sequência Expressa	-
14	37	gi 30693968 lipid binding	9e-29
15	34	gi 30841338 arabinogalactan protein	5e-53
16	33	gi 23452315 beta-tubulin 1	0.0
17	32	gi 1345882 cytochrome b5	3e-60
18	32	Sem Homologia	-
19	29	gi 8134569 homocysteine methyltransferase	0.0
20	29	gi 15236989 translation initiation factor	4e-55
21	27	gi 29367605 wheat adenosylhomocysteinase-like	1e-154
22	26	gi 92880884 adenosine kinase	2e-113
23	26	gi 92898246 heavy metal transport/detoxification	1e-47
24	26	gi 1934859 caffeoyl-CoA O-methyltransferase (CCoAOMT)	3e-138
25	26	gi 47176684 copper chaperone	9e-30
26	26	gi 84579145 hypothetical protein	3e-09
27	25	gi 1566700 metallothionein-like TYPE3	2e-24
28	25	gi 38016521 fiber protein Fb34	9e-71
29	24	gi 115820 chlorophyll a-b binding protein	6e-125
30	24	gi 1838961 acyl carrier protein	1e-33
31	24	gi 1223652 metallothionein 1	3e-05
32	24	gi 34909054 tubulin beta chain	2e-139
33	24	gi 35187687 zinc-finger protein	6e-51
34	24	Sequência Expressa	-
35	24	Sequência Expressa	-
36	23	gi 18398335 RCD1 (Radical-induced Cell Death)	1e-10
37	23	gi 21536535 unknown	9e-17
38	22	gi 13924490 cytosolic glutamine synthetase (GS1)	9e-20
39	22	gi 15241245 RAN2 GTP binding	4e-73
40	22	gi 67003907 cellulose synthase (CesA1)	2e-144
41	22	gi 60656567 endo-1,4-beta-xylanase	8e-107
42	22	gi 49642980 unnamed protein product	3e-20
43	22	Sem Homologia	-

Tabela 6 - Lista dos 50 genes (*tags*) mais expressos na biblioteca SAGE de plantas de eucalipto com 6 meses de idade. Para cada *tag* é fornecido o número de cópias na biblioteca, a identificação do gene (BLAST-X) bem como o valor *E* para cada identificação. As *tags* “Sem Homologia” são aquelas para as quais não foi possível encontrar o EST ou o cDNA de origem, já as *tags* “Sequência Expressa” representam aquelas para as quais encontrou-se a sequência de EST ou cDNA correspondente, porém apresentaram valor $E \geq e-05$

(conclusão)

TAG	Nº de cópias	Blast(x) – NCBI	E value
44	20	gi 3482933 cdc2 protein kinase	3e-74
45	20	gi 13430203 auxin-repressed protein	2e-31
46	20	gi 4733946 sucrose synthase (SuSy)	4e-74
47	20	gi 78191442 S-adenosyl methionine synthase-like (SAMS)	6e-84
48	20	gi 2677828 cysteine protease	7e-165
49	20	Sem Homologia	-
50	19	gi 4336893 cinnamyl alcohol dehydrogenase (CAD)	3e-54

Para segunda *tag* mais expressa na biblioteca, com 107 cópias, não foi possível encontrar a possível função metabólica relacionada no NCBI, já para a terceira e a quarta *tags* com 99 e 76 cópias respectivamente, não foi possível sequer encontrar um EST correspondente tanto no banco de dados do NCBI quanto no banco do Genolyptus (Tabela 6).

Dentro da categoria “Metabolismo e Energia” foram encontradas diversas *tags* entre as 50 mais expressas. Entre as *tags* relacionadas com energia, uma apresentou homologia a *Glyceraldehyde-3-phosphate dehydrogenase* citosólica de *Nicotiana tabacum* (GAPDH) (56 cópias), uma apresentou homologia ao *Cytochrome b5* de *Nicotiana tabacum* (32 cópias) e outra apresentou homologia a *Chlorophyll a/b-b* do sistema antena coletor de luz de *Pinus sylvestris* (24 cópias) (Tabela 6).

As angiospermas e algas possuem duas distintas enzimas (GAPDH), uma encontrada no citosol envolvida na glicólise, forma extremamente conservada entre as espécies e que foi encontrada na biblioteca e outra localizada no cloroplasto a qual participa do ciclo de Calvin (MEYER et al., 1994).

Nos organismos eucariotos fotossintéticos que contêm 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). Dessa forma, a presença com elevada expressão da *tag* relacionada a *chlorophyll a/b* do sistema antena coletor de luz é um resultado interessante. A face externa da casca tem uma baixa permeabilidade para a difusão de gases, levando ao acúmulo de CO₂ nos espaços intercelulares, muito maior do que no ambiente. Dessa maneira, a provável fotossíntese no caule deve funcionar como um importante mecanismo para reduzir parcialmente a anaerobiose interna evitando uma possível limitação para respiração no escuro e produção de ATP. A ocorrência de cloroplastos funcionais com fotossistemas ativos nas camadas mais externas da periderme e também em tecidos mais profundos do caule tem sido reportada em diversas espécies arbóreas (PFANZ et al., 2002) podendo explicar, portanto, a presença desse transcrito na biblioteca. A presença dessa *tag* no presente estudo é corroborada pelos resultados encontrados por Carvalho et al. (2007). Além disso, proteínas do complexo coletor de luz e rubisco vêm sendo identificadas por outros pesquisadores que estudam a formação do xilema, (JUAN et al., 2006; CELEDON et al., 2007).

Entre as *tags* relacionadas com Metabolismo, foram encontradas: uma *tag* para metabolismo de nitrogênio (*Cytosolic glutamine synthetase*); três *tags* para o metabolismo de aminoácidos sendo, uma *S-adenosylmethionine synthase* (SAMS), uma *Adenosylhomocysteinase* (degradação de aminoácidos) e uma *Homocysteine methyltransferase*; duas *tags* para metabolismo de lipídeos (uma *Acyl carrier protein* e uma *Lipid binding protein*); uma *tag* para metabolismo de Nucleotídeos/Fósforo/Vitaminas/Co-Fatores (*Adenosine Kinase*) (Tabela 6).

Recentemente, Pereira et al. (2006) demonstrou que mutantes de *Arabidopsis* com uma deficiência na enzima *Adenosine Kinase* apresentaram folhas e hipocótilos com morfologia anormal, provavelmente devido a grande quantidade de pectinas de baixa metil esterificação na parede celular primária. Estes resultados indicam uma importante função para essa enzima na reciclagem de *S-adenosylmethionine* dependente de metiltransferases durante o processo de metil esterificação das pectinas e na determinação das propriedades de adesão celular.

A participação das *tags* que representaram a *Cytosolic glutamine synthetase* – GS1 (glutamina sintetase citosólica) e a *S-adenosylmethionine synthase* (SAMS) no processo de formação da madeira será discutida posteriormente.

Em meio às 50 *tags* mais expressas relacionadas com a categoria “Processos Celulares”, encontram-se: uma *tag* relacionada a divisão celular (*cdc2 protein kinase*); uma *tag* relacionada a regulação celular (RCD1 - *Radical-induced Cell Death*); quatro *tags* relacionadas a respostas de proteção/destoxificação (uma *Heavy metal transport/detoxification*, uma *Copper chaperone* e duas *Metallothioneins*) e uma *tag* relacionada a respostas à estímulos (*Auxin-repressed protein*) (Tabela 6).

A *cdc2 protein kinase* é importante durante a divisão celular do meristema cambial (HERTZBERG et al., 2001), e a existência na biblioteca de uma *tag* relacionada indica a presença de células cambiais no início do processo de diferenciação no material amostrado. Já, a proteína RCD1 está envolvida na regulação hormonal, regulação de vários genes de resposta a estresse e ao processo de morte celular programada (PCD) (AHLFORS et al., 2004).

As *Metallothioneins* (metalotioninas) identificadas em diversos organismos, estão associadas com a resposta a estresse e destoxificação de diversos íons metálicos, mais comumente, zinco e cobre. Todavia, Jacob-Wilk et al. (2006) em trabalho realizado em algodão, demonstraram que as metalotioninas podem ter um importante papel no fornecimento de zinco aos domínios de ligação ao zinco do complexo das *Cellulose synthases* - CESAs (celuloses sintases) durante as mudanças entre os estados oxidados e reduzidos. Segundo os autores, as metalotioninas são proteínas de redução, e poderiam estar envolvidas na regulação da formação, ativação, ou desativação do complexo das CESAs, fornecendo ou retirando íons de zinco aos domínios de ligação de zinco das proteínas CESAs. Dessa forma, eles propõem um modelo onde a ativação, *turnover* e degradação dos complexos das proteínas CESAs na forma dimerizada é mediada pela redução do complexo através da adição do zinco pelas metalotioninas e subsequente proteólise envolvendo uma *Cysteine protease* (cisteína protease).

Com relação as *tags* presentes na categoria “Transporte” encontram-se entre as 50 *tags* mais expressas: uma *tag* relacionada ao transporte de pequenas moléculas (*Aquaporin PIP3*), possivelmente relacionada a promoção da expansão celular mediada pela pressão de turgor (COSGROVE, 1993) e duas *tags* relacionadas ao transporte de vesículas (um *ADP-ribosylation factor 1* e um *RAN2 GTP binding*) (Tabela 6).

Já entre as 50 *tags* mais expressas relacionadas com a categoria “Estrutura e Organização da Estrutura Celular”, encontram-se: seis *tags* diretamente relacionadas à parede celular (uma *Arabinogalactan protein*, uma *Caffeoyl-CoA O-methyltransferase* – CCoAOMT, uma *Cellulose synthase* – CesA1, uma *endo-1,4-beta-xylanase*, uma *Sucrose synthase* – SuSy e uma *Cinnamyl alcohol dehydrogenase* – CAD), sendo que o papel de cada uma destas enzimas será discutido posteriormente; duas *tags* para citoesqueleto (duas *β -tubulins*) cuja importância também será discutida posteriormente (Tabela 6).

Finalmente, encontram-se presentes entre as 50 *tags* mais expressas: uma *tag* relacionada à Síntese/Degradação/Modificação de RNA (*Zinc-finger protein*); duas *tags* relacionadas à Síntese/Degradação/Modificação de Proteínas (um *Translation initiation factor* e uma *Cysteine protease*), que foram catalogadas na categoria “Vias de Informação” (Tabela 6).

É possível notar ainda que entre as 50 *tags* mais expressas, encontram-se cinco *tags* para as quais não foi possível encontrar a provável função do EST correspondente no banco de dados do NCBI, quatro *tags* para as quais o resultado encontrado no BLAST-X do NCBI gerou uma proteína de função desconhecida ou hipotética e ainda, oito *tags* para as quais não foi possível sequer encontrar o EST de origem (Tabela 6). Isso reforça ainda mais a idéia da necessidade de se aumentar a quantidade de sequências disponíveis para o acesso público.

Quando a distribuição em categorias funcionais é feita através da distribuição da abundância do número de transcritos ao invés da distribuição pelo número de genes, a representatividade dos genes relacionados aos processos acima citados pode ser melhor observada. Assim, a partir dessa nova distribuição utilizando apenas os transcritos que foram identificados nos bancos de dados e que possuem uma possível função atribuída, a categoria “Estrutura e Organização da Estrutura Celular” passa a

representar 27% dos transcritos encontrados, o mesmo percentual da categoria “Metabolismo e Energia”. As categorias “Vias de Informação” e “Processos Celulares” passam a ter 17% dos transcritos e a categoria “Transporte” passa a representar 12% dos transcritos identificados (Figura 19).

Figura 19 - Distribuição dos genes pelo número total de transcritos identificados a partir da biblioteca SAGE em categorias funcionais

Uma das vantagens da técnica SAGE é a possibilidade da quantificação do nível de expressão de vários genes. Dessa forma, uma maneira de analisar os resultados produzidos pela biblioteca SAGE é através da análise da relação entre o número de transcritos e o número de genes dentro de cada uma das categorias funcionais, criando dessa forma, sub-categorias funcionais. Assim, a relação entre o número de transcritos e o número de genes (transcritos/gene) pode nos mostrar qual sub-categoria específica predomina dentro de uma determinada categoria (Tabela 7).

Tabela 7 – Distribuição dos genes identificados na biblioteca SAGE em categorias e subcategorias funcionais. Para cada categoria/subcategoria são fornecidos: o número de genes identificados (G), o número total de transcritos (T) e a relação transcritos/genos (T/G)

Categorias Funcionais	Genes (G)	Transcritos (T)	T/G
1 Metabolismo e Energia	180	1125	6,25
1.1 Energia (Autotrófico/Carbono/ATP)	86	458	5,32
1.2 Metabolismo de Carboidratos	31	165	5,32
1.3 Metabolismo de Nitrogênio	5	39	7,8
1.4 Metabolismo de Aminoácidos	26	162	6,23
1.5 Metabolismo de Lipídeos	16	111	6,93
1.6 Metabolismo de Nucleotídeos/Fósforo/Vitaminas/Co-Fatores	8	46	5,75
1.7 Metabolismo Secundário	8	144	18
2 Processos Celulares	120	710	5,91
2.1 Divisão Celular	11	68	6,18
2.2 Tradução de Sinais	31	121	3,90
2.3 Regulação Celular	28	139	4,96
2.4 Respostas de Proteção/Destoxificação	41	310	7,56
2.5 Respostas a Estímulos	9	72	8
3 Transporte	71	488	6,87
3.1 Transporte de Grandes Moléculas	12	93	7,75
3.2 Transporte de Pequenas Moléculas	23	156	6,78
3.3 Transporte de Vesículas	30	213	7,1
3.4 Kinesinas/Fatores Transportadores	6	26	4,33
4 Estrutura e Organização da Estrutura Celular	184	1168	6,35
4.1 Membrana Celular	13	50	3,84
4.2 Parede Celular	67	506	7,55
4.3 Citoesqueleto	25	217	8,68
4.4 RNAr e Proteínas Ribossomais	74	367	4,96
4.5 Organização dos Cromossomos	5	28	5,6
5 Vias De Informação	147	739	5,03
5.1 Síntese/Degradação/Modificação de DNA	7	24	3,43
5.2 Síntese/Degradação/Modificação de RNA	38	152	4
5.3 Síntese/Degradação/Modificação de Proteínas	102	563	5,52
6 Sem Função Definida	287	1733	6,04
6.1 Sequência Expressa	138	1002	7,26
6.2 Proteína de Função Hipotética	149	731	4,9
Total	989	5963	6,0

Observa-se que dentro da categoria “Metabolismo e Energia”, a maior parte dos genes e dos transcritos está relacionada com o fornecimento de energia, estando classificados na sub-categoria “Energia (Autotrófico/Carbono/ATP)”, seguidos de “Metabolismo de Carboidratos” e “Metabolismo de Aminoácidos”. Porém, quando se leva em conta a relação entre o número de transcritos pelo número de genes (T/G), a

sub-categoria “Metabolismo Secundário” apresenta a maior relação, seguido pela sub-categoria “Metabolismo de Nitrogênio” (Tabela 7), mostrando portanto que os genes relacionados ao metabolismo de nitrogênio e principalmente ao metabolismo secundário encontram-se mais expressos.

O grande número de genes listados na sub-categoria “Energia (Autotrófico/Carbono/ATP)” é devido a grande quantidade de genes associados a: respiração (três *tags* para *Cytochrome b5* e três *tags* para *Cytochrome c*); fotossíntese (sete *tags* para *Chlorophyll a/b-b* do sistema antena coletor de luz e duas *tags* para proteínas *Photosystem II*); glicólise (cinco *tags* para a *Glyceraldehyde-3-phosphate dehydrogenase cytosolic* e uma *tag* para *Pyruvate kinase*); ciclo dos ácidos tricarboxílicos (duas *tags* para *Malate dehydrogenase*, uma *tag* para *NADP- isocitrate dehydrogenase* e uma *tag* para a *2-oxoglutarate dehydrogenase*); fermentação alcoólica (três *tags* *Alcohol dehydrogenase*) e síntese de ATP (*Mitochondrial ATP synthase*).

Em órgãos não fotossintetizantes, os carboidratos são consumidos através da respiração para produzir energia e esqueletos de carbono necessários para o metabolismo celular e biossíntese das moléculas estruturais, incluindo os polímeros da parede celular. Entretanto, a quantidade de O₂ livre na região cambial deve ser limitada pela barreira física imposta pela casca e também devido ao próprio consumo de O₂ durante a respiração. Dessa forma, uma parte da energia necessária para o crescimento secundário poderia estar sendo fornecida pela fermentação alcoólica (KIMMERER; STRINGER, 1988). Assim, como sugerido por Carvalho et al. (2007), o possível papel da fermentação alcoólica como rota alternativa no suprimento de energia nas plantas de eucalipto pode ser indicado pela presença de três *tags* para *Alcohol dehydrogenase* (álcool desidrogenase) na biblioteca. Da mesma forma, outros estudos recentes relataram a presença de transcritos ou proteínas associadas à fermentação durante a formação do xilema e crescimento secundário (GION et al., 2005; JUAN et al., 2006; RANIK et al., 2006).

Como já era esperado, foram encontrados diversos genes importantes para o processo de formação da madeira presentes na sub-categoria “Metabolismo de Carboidratos”. As enzimas ativas no metabolismo de carboidratos (CAZymes) incluem famílias de *Glycosyl Transferases* – GT (glicosil transferases), *Glycosyl Hidrolases* – GH

(glicosil hidrolases), polissacarídeo liases e esterases, que são fundamentais não só para a síntese de açúcares e amido durante a fotossíntese, mas também são extremamente importantes para a biossíntese dos polímeros de celulose, pectina, xilanos e compostos secundários na região cambial (HENRISSAT, 1991; COUTINHO et al., 2003; GEISLER-LEE et al., 2006). Entre essas enzimas foram encontrados genes para CAZymes cujas funções já encontram-se bem estabelecidas, como por exemplo, as CESAs. No entanto, foram identificados ao mesmo tempo genes para enzimas com funções ainda não esclarecidas, mas que têm sido associadas ao processo de formação da madeira, como por exemplo, GT8D (membro da família das GT) e GH36 e GH55 (membros das GH). Entretanto, os genes referentes a essa sub-categoria serão discutidos com maior detalhe nos itens 4.3.3.1 (Biossíntese de Hemicelulose e Pectina) e 4.3.3.2 (Biossíntese de Celulose).

Da mesma forma, na sub-categoria “Metabolismo de Aminoácidos” foram encontrados três *tags* para *S-adenosylmethionine synthase* (SAMS) cuja importância será discutida no item 4.3.4 (Lignificação).

A grande relação T/G para a sub-categoria “Metabolismo Secundário” foi devido a presença de cinco *tags* referentes a *Isoflavone reductases*, sendo uma a mais expressa na biblioteca (126 cópias). Já, no caso da sub-categoria “Metabolismo de Nitrogênio”, a grande relação T/G foi devido a presença de três *tags* para *Glutamine synthetase* (GS1).

Dentro da categoria “Processos Celulares”, a maior parte dos genes e dos transcritos está relacionada com “Respostas de Proteção/Destoxificação”, “Tradução de Sinais” e “Regulação Celular” porém, a maior relação T/G está na sub-categoria “Respostas a Estímulos” seguida da sub-categoria “Respostas de Proteção/Destoxificação” e “Divisão Celular” (Tabela 7).

O elevado número de genes assim como a alta relação T/G verificada para a sub-categoria “Respostas de Proteção/Destoxificação” deve-se a presença de sete *tags* para *Metallothioneins*, sendo: quatro *tags* para o tipo 1, duas *tags* para o tipo 2 e uma *tag* para o tipo 3, que curiosamente foi a mais expressa na biblioteca. Já para a sub-categoria “Transdução de Sinais” o elevado número de genes deve-se a presença de *Protein kinases*.

Para a sub-categoria “Regulação Celular” o elevado número de transcritos é devido principalmente à presença de uma *tag* para (RCD1 - *Radical-induced Cell Death*) com alta expressão (23 cópias). Entre outros genes interessantes presentes nesta sub-categoria estão três *tags* relacionadas a proteínas 14-3-3 e uma *tag* relacionada a uma *Germin-like proteins* (GLP). As proteínas 14-3-3 constituem uma família multigênica presente na maioria dos organismos e altamente conservadas entre as espécies. Essas proteínas ocorrem em distintas isoformas e estão envolvidas na regulação dos sinais múltiplos nos eventos de transdução, atuando na ativação de uma ampla gama de proteínas alvo (FERL, 2004). Em plantas, diferentes isoformas da proteína associam-se com fatores de transcrição promovendo a ativação de vários genes ligados ao estresse (FERL, 1996). Em *Arabidopsis* foram encontradas trinta diferentes isoformas dessas proteínas (DELILLE et al., 2001). A presença dessas *tags* no material sugere que essas proteínas possam ter um possível papel na regulação da diferenciação do xilema nas plantas de eucalipto, assim como verificado por Carvalho et al. (2007). Já as *germins* e as GLPs fazem parte da superfamília de proteínas conhecidas como *cupin*, caracterizadas por duas funções enzimáticas diferentes, podendo atuar como oxalato oxidase e superóxido dismutase, sendo que ambas as atividades resultam na produção de H₂O₂ que pode ser desviado para a polimerização oxidativa dos monolignóis durante a lignificação (MATHIEU et al., 2006). Isto poderia sugerir então uma ligação entre a expressão das GLPs no caule das plantas de eucalipto e um possível papel na lignificação.

A maior relação T/G para a sub-categoria “Respostas a Estímulos” foi devido à presença de *tags* relacionadas a proteínas reprimidas ou induzidas por auxinas. Acredita-se que as auxinas atuem como mensageiros intercelulares durante os processos embrionários, desenvolvimento de meristemas e desenvolvimento vascular e, também, que a sinalização de longa distância mediada por auxinas poderia integrar de maneira simultânea a morfogênese através das plantas, podendo, portanto, determinar o padrão de diferenciação dos tecidos vasculares bem como a diferenciação de células simples para se tornarem elementos traqueários (BERLETH; SACHS 2001). Groover e colaboradores (2003) demonstraram em *Zinnia* que a resposta a auxinas nos estádios iniciais de desenvolvimento pode ser indicada pela presença de proteínas

induzidas por auxina (IAA), mais precisamente pelas *zIAA8*, e envolvidas no início da diferenciação vascular. Mais recentemente, Ko et al. (2004) revelaram que as auxinas possuem um importante papel na rota de sinalização que resulta na transição do crescimento primário para o crescimento secundário em resposta ao peso da planta. Dessa forma, a grande presença dessas auxinas na biblioteca sugere que o material está sofrendo processo de crescimento secundário.

Para a sub-categoria “Divisão Celular” a elevada relação T/G (Tabela 7) deve-se a presença de duas *tags* representando uma *cdc2 protein kinase*, importante durante a divisão celular do meristema cambial (Hertzberg et al., 2001), e quatro *tags* similares a *Translationally controlled tumor proteins* (TCTP). Embora, acredita-se que as proteínas TCTPs sejam importantes no crescimento celular e também na divisão em mamíferos (BOMMER; THIELE, 2004), seu papel exato em plantas ainda é desconhecido. Além disso, foram encontradas cinco *tags* para a proteína *Skp1*. A identificação de homólogos da *Skp1* em *C. elegans*, *A. thaliana*, *M. sativa* e *H. sapiens* indicou que essa proteína é altamente conservada evolutivamente. A *Skp1* representa uma proteína intrínseca do cinetócoro, conservada através da evolução dos eucariotos e muito provavelmente esteja diretamente envolvida na maquinaria de regulação do ciclo celular (CONNELLY; HIETER, 1996). A presença dessa grande quantidade de *tags* associada à divisão celular indicam a presença de uma quantidade significativa de células meristemáticas no início do processo de diferenciação no material amostrado.

Dentro da categoria “Transporte”, a maior parte dos genes e dos transcritos está relacionada com “Transporte de Vesículas”, porém quando verifica-se a relação T/G as sub-categorias “Transporte de Grandes Moléculas”, “Transporte de Pequenas Moléculas” e “Transporte de Vesículas” apresentam valores muito próximos (Tabela 7).

Para a sub-categoria “Transporte de Pequenas Moléculas” a elevada relação T/G foi ocasionada principalmente pela alta expressão das *Aquaporins* (aquaporinas). Foram identificadas uma *tag* representando uma *Aquaporin* PIP1, três *tags* representando uma *Aquaporin* PIP2, uma *tag* representando uma *Aquaporin* PIP3 e ainda uma *tag* representando uma *Aquaporin* TIP1. O papel das aquaporinas será discutido posteriormente.

O grande número de transcritos relacionados não só ao transporte de vesículas, mas também ao citoesqueleto na biblioteca (Tabela 7) está diretamente relacionado à biossíntese e organização dos polímeros de parede celular. Entre os genes associados ao transporte de vesículas que foram encontrados na biblioteca destacam-se os membros da família das pequenas GTPases como, o *Ras* envolvido na proliferação celular, o *Rab* e o *Arf* responsáveis pelo tráfico de membranas, o *Ran* que faz o transporte através da membrana celular e *ADP-ribosylation factors* (VERNOUD et al. 2003) além de *Annexins* (CARROL et al., 1998). Os polissacarídeos da parede celular são sintetizados no complexo de Golgi a partir dos açúcares nucleotídeos e são exportados para a matriz celular em vesículas derivadas do Golgi por meio da rede de citoesqueleto (DELMER; STONE, 1988). Além disso, os complexos protéicos de membrana envolvidos na síntese de celulose são também incorporados à membrana celular a partir da fusão de vesículas. 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).

Dentro da categoria “Estrutura e Organização da Estrutura Celular”, a maior parte dos genes e dos transcritos está relacionada com “Parede Celular” e “RNAr e Proteínas Ribossomais”, porém quando leva-se em consideração a relação T/G, a sub-categoria “Citoesqueleto” destaca-se, sendo seguida pela sub-categoria “Parede Celular” (Tabela 7).

A grande quantidade de genes relacionados com a sub-categoria “RNAr e Proteínas Ribossomais”, sugere que o material amostrado encontrava-se em processo de alta atividade de síntese protéica, o que é coerente com a grande quantidade de genes e transcritos encontrados na sub-categoria “Síntese/Degradação/Modificação de Proteínas” da categoria “Vias de Informação” (Tabela 7).

Vale a pena destacar também a presença de quatro *tags* com similaridade a *Remorins* dentro da sub-categoria “Membrana Celular”. Bariola et al. (2004) através de localização tecido-específica de *Remorins* por anticorpos monoespecíficos em tomate, observaram que a proteína *Remorin 1* estava presente principalmente nas células vasculares, meristemáticas e embriogênicas. Os autores sugerem que essas proteínas

estejam associadas com o citoesqueleto ou com a membrana celular, ajudando na determinação celular e/ou atuando como “molde” (*scaffolds*) para sinalização de defesa ou desenvolvimento. A alta frequência dessas *Remorins* na biblioteca, assim como encontrado por Carvalho et al. (2007), sugere uma provável importância dessas proteínas no desenvolvimento vascular e formação da madeira em eucalipto.

Para a sub-categoria “Citoesqueleto”, a elevada relação T/G é esperada uma vez que, como já visto, está diretamente relacionada ao grande número de transcritos associados ao transporte de vesículas. Entre os genes encontrados na biblioteca para esta sub-categoria, destacam-se os relacionados à *Tubulins* (α e β), *Actins*, *Actin depolymerizing factors* e *Profilins*.

O importante papel do citoesqueleto nos processos de espessamento da parede celular já foi bem documentado. Oda e Hasezawa (2006), observaram vesículas contendo o complexo de CESAs e materiais fibrosos da parede em células em diferenciação de *Arabidopsis*. Os microtúbulos associados com a parede celular secundária, por sua vez, devem participar no transporte dessas vesículas. As *Tubulins* α e β (tubulinas α e β) são responsáveis pelos padrões de deposição da parede celular uma vez que definem a posição e orientação das microfibrilas de celulose (CHAFFEY, 2000). De acordo com Oda e Hasezawa (2006), durante a formação da parede celular secundária a deposição e orientação das microfibrilas de celulose parece ser controlada pelo arranjo de microtúbulos. As *Actins* (actinas) são proteínas que estão envolvidas com a formação dos filamentos que compõem o citoesqueleto, sendo que em plantas existem várias isoformas de actinas envolvidas em diferentes funções como divisão celular, determinação da forma celular e deposição de parede celular (KABSCH; VANDEKERCKHOVE, 1992; HERMAN, 1993). Já as *Profilins* (profilinas) são responsáveis pela regulação da actina, interferindo durante a polimerização das mesmas, tendo sido observado em *Arabidopsis* que as profilinas possuem um importante papel na elongação e determinação da forma celular (RAMACHANDRAN et al., 2000). A construção do esqueleto de microtúbulos é indicada pela elevada expressão das tubulinas (α e β), actinas, profilinas e outros fatores de despolimerização de actinas na biblioteca, confirmando mais uma vez a alta atividade da região cambial para produção de xilema nas plantas de eucalipto com 6 meses de idade.

Com relação a sub-categoria “Parede Celular”, que é a classe de genes que mais interessa no presente trabalho, o grande número de genes e transcritos, bem como, a elevada relação T/G é devido aos inúmeros genes identificados e que estão diretamente relacionados ao processo de formação da madeira. Entre eles, *tags* para genes relacionados a *Arabinogalactan proteins* – AGPs (proteínas arabinogalactanos), similar ao gene *FLA13* de *Populus*, preferencialmente expresso no tecido xilemático em diferenciação para deposição de parede secundária (LAFARGUETTE et al., 2004). As AGPs são proteínas estruturais, com um domínio de proteoglicano e importantes na mediação das interações célula-célula durante a diferenciação vascular. Estas proteínas são encontradas nos espaços intercelulares e paredes celulares, nas membranas plasmáticas e em vesículas citoplasmáticas, sugerindo suas múltiplas funções nos vários processos associados com crescimento celular e desenvolvimento da planta (MAJEWSKA-SAWKA; NOTHNAGEL, 2000).

Embora as plantas não contenham quitina endógena, foram encontradas duas *tags* para proteínas similares a *Chitinase* (quitinase), pertencentes à família de hidrolases glicosídicas GH19. De acordo com Zhang et al. (2004), esse grupo de quitinases desempenha importantes funções na síntese da parede celular primária e secundária através de sua capacidade de ligação a oligossacarídeos ao invés de sua capacidade catalítica, sendo preferencialmente expressa durante a deposição da parede secundária em algodão. De maneira similar, Aspeborg et al. (2005) observou a expressão dessas GH19 durante a síntese da parede secundária em *Populus*, sugerindo a sua importância na formação da parede celular secundária. Dessa forma, a expressão dessas proteínas no material amostrado, sugere sua participação na formação da madeira das plantas de eucalipto, o que está de acordo com o encontrado por Carvalho et al. (2007) para árvores juvenis de eucalipto.

Muito outros genes relacionados a biossíntese de celulose, hemicelulose, pectina e lignina, encontrados na biblioteca e pertencentes a sub-categoria “Parede Celular”, serão discutidos nos itens 4.3.3.1 (Biossíntese de Hemicelulose e Pectina), 4.3.3.2 (Biossíntese de Celulose) e 4.3.4 (Lignificação).

Dentro da categoria “Vias de Informação”, a grande maioria dos genes, dos transcritos e também da relação T/G está relacionada com “Síntese/Degradação/Modificação de Proteínas”, embora também tenham sido identificados alguns genes importantes dentro da sub-categoria “Síntese/Degradação/Modificação de RNA” (Tabela 7).

Dessa forma, para “Síntese/Degradação/Modificação de RNA” foram identificados alguns fatores de transcrição, como por exemplo, membros da classe *HD-Zip III*, comparáveis ao gene *PtaHB1* de *Populus*. Este gene é regulado tanto pelo desenvolvimento quanto pela estação e está estritamente associado à formação da madeira em *Populus* (KO et al., 2006). Também foram encontrados dois fatores de transcrição mediados por etileno (*Ethylene-responsive transcriptional factor*), contendo um domínio AP2/ERF e um domínio DRM1. O domínio AP2/ERF é encontrado em fatores de transcrição específicos de plantas, sendo o domínio ERF específico para ligação ao DNA. Este domínio está representado por uma grande família no genoma de *Arabidopsis* (RIECHMANN; RATCLIFFE, 2000) e é conhecido por estar envolvido na tradução de sinais de hormônios vegetais como também em resposta a estresses bióticos e abióticos (YI et al., 2004). Já a família AP2 também pode desempenhar funções durante o desenvolvimento vegetativo uma vez que sua expressão foi verificada em hastes e folhas de *Arabidopsis* (JOFUKU et al., 1994; KIRCH et al., 2003). Recentemente, van Raemdonck et al. (2005) demonstrou que um membro da família AP2/ERF estava ativo em células do floema de *Populus*, podendo ser importante na transição do crescimento primário para o crescimento secundário nessas árvores. O domínio DRM1 é normalmente reprimido na presença de auxina e é importante no ciclo entre os estados de crescimento e dormência (STAFSTROM et al. 1998). A presença desse gene em maior expressão do que o verificado por Carvalho et al. (2007) sugere que as plantas de eucalipto utilizadas neste trabalho estão provavelmente passando por um intenso processo de transição de crescimento primário para secundário garantindo dessa forma o rápido crescimento da árvore.

O fator de transcrição do tipo MYB (*MYB transcription factor*) foi encontrado em meio aos outros fatores de transcrição presentes na biblioteca. Foram identificados três *tags* para esse fator de transcrição, sendo: uma *tag* (2 cópias) com homologia ao gene

EgMYB1 de eucalipto, que está diretamente relacionado a biossíntese de lignina (Paux et al., 2005), uma *tag* (2 cópias) com homologia ao gene *MYB1* de *Pinus* e uma *tag* (8 cópias) que apresentou homologia ao gene *MYB103* de *Arabidopsis*. O alinhamento dos ESTs referentes as três *tags*/genes mostrou que não existe homologia entre eles, sugerindo que os transcritos sejam originados de genes diferentes. Dessa forma, como o gene *MYB103* de *Arabidopsis* apresentou maior expressão do que os demais, pode-se sugerir que o mesmo desempenhe importante função durante o crescimento secundário nas plantas de eucalipto estudadas.

Outro aspecto interessante com relação aos fatores de transcrição foi a ausência de *tags*/genes com homologia à proteínas do tipo LIM. As proteínas LIM são fatores de transcrição com domínio dedo de zinco envolvidos no processo de lignificação, sendo que em recente trabalho com o gene *BnLIM* de *Brassica napus* foi demonstrado que a expressão dos genes *LIM* e *Cinnamoyl-CoA reductase* (CCR) são equivalentes (HUANG et al., 2007). Da mesma forma, antisense para o gene *Ntlim1* de *Nicotiana tabacum* levou a produção de plantas com uma redução de 20% no teor de lignina (KAWAOKA; EBINUMA, 2001). Também recentemente, Carvalho et al. (2007) encontrou duas proteínas tipo LIM análogas a LIM1 de *E. globulus* em árvores juvenis de eucalipto e, através de alinhamento, observou que elas eram similares aos genes *BnLIM* de *Brassica napus* e *Ntlim1* de *Nicotiana tabacum*. Entretanto, também não foi encontrada nenhuma *tag*/gene para CCR na biblioteca de eucalipto em questão. Esses resultados indicam que embora os fatores LIM estejam envolvidos na biossíntese de lignina, o papel dessas proteínas talvez não seja tão significativo no início do desenvolvimento, como é o caso das plantas de 6 meses de idade, mas podendo assumir notável importância com o desenvolvimento.

Finalmente, a sub-categoria “Síntese/Degradação/Modificação de Proteínas”, foi caracterizada principalmente por genes relacionados à degradação de proteínas, como *Ubiquitins* (ubiquitinas), *Proteasomes* (proteassomos), *Heat Shock proteins*, *Cysteine proteases* e *Aspartic proteases*. A importância da via Ub/proteassomo 26S e sua relação com as Vias de resposta às auxinas será discutida no item 4.3.5 Morte celular programada (PCD).

4.3 Processo de Formação da Madeira

Durante o processo de formação da madeira, conhecido como 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, sofrendo finalmente o processo de morte celular programada (PCD) (CHAFFEY, 1999).

4.3.1 Divisão celular

Como já discutido anteriormente, a sub-categoria “Divisão celular” apresentou uma alta relação T/G (Tabela 7), sendo a maior parte dos transcritos similares as *cdc2 protein kinases*, *Translationally controlled tumor proteins* (TCTP) e *Skp1* (proteína do cinetócoro requerida para progressão do ciclo celular). De maneira semelhante, um elevado número de transcritos para genes relacionados ao ciclo celular foi também encontrado em células cambiais em divisão, em *Populus* (HERTZBERG et al., 2001), *Arabidopsis* (OH et al., 2003) e *Eucalyptus* (CARVALHO et al., 2007), indicando realmente que a amostra utilizada na construção da biblioteca contém células meristemáticas em processo de divisão celular e diferenciação, que vão levar ao crescimento e formação da madeira.

4.3.2 Expansão celular

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). O processo de expansão celular envolve diversas proteínas relacionadas ao aumento da extensibilidade da parede celular (HERNÁNDEZ-NIDAL et al., 2006).

O papel das aquaporinas na expansão tem sido sugerido com base na sua forte expressão em tecidos em crescimento (LUDEVID et al., 1992; BALK; de BOER 1999). Em zonas de diferenciação, a função das aquaporinas chamadas PIP (*Plasma membrane Intrinsic Proteins*) parece ser o fornecimento de água requerido para expansão e aumento de peso fresco (COSGROVE, 1993). Embora poucos trabalhos

envolvendo a análise de expressão gênica na formação da madeira (YANG et al., 2003, RANIK et al., 2006) relatem a identificação de transcritos superexpressos para canais transportadores de água. Na biblioteca foram identificadas: uma *tag* representando uma *Aquaporin* PIP1, três *tags* representando uma *Aquaporin* PIP2, uma *tag* representando uma *Aquaporin* PIP3 e ainda uma *tag* representando uma *Aquaporin* TIP1 (*Tonoplast Intrinsic Protein*). Recentemente, Carvalho (2007) encontrou uma alta expressão para aquaporinas PIP principalmente para PIP2 e PIP3 na região cambial de árvores juvenis de eucalipto. Dessa forma, certamente o grande número de transcritos encontrados para estas proteínas na biblioteca, esteja relacionado à promoção da expansão celular mediada pela pressão de turgor.

Nas dicotiledôneas, a parede celular primária consiste basicamente de uma rede de microfibrilas de celulose ligadas a glucanos de ligação cruzada (hemiceluloses) embebidas em uma matriz de pectinas (CARPITA; GIBEAUT, 1993). A diferenciação das células do xilema nessas plantas envolve a expansão da parede celular primária antes do início da deposição da parede celular secundária. No entanto, a expansão celular só é possível com a modificação da estrutura da parede celular que é catalisada por diversas enzimas como as *Xyloglucan endotransglycosylases* (XETs), *Expansins*, *Pectinesterases* ou *Pectin methylesterases* (PMEs), *Pectate lyase* e *Endopolygalacturonases* (Revisado em MELLEROWICZ et al., 2001).

As XETs são responsáveis pela remodelagem da parede celular durante a síntese da parede primária “cortando” e “reagrupando” as cadeias de xiloglucanos (tipo de hemicelulose mais abundante na parede primária das dicotiledôneas). Embora a atividade das XETs na expansão celular esteja associada ao crescimento primário promovendo um afrouxamento da parede primária através da clivagem e transferência de cadeias de xiloglucanos, 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 xiloglucanos pré-existentes com as novas cadeias que estão sendo depositadas (BOURQUIN et al., 2002). Todavia, apesar de sua reconhecida importância, apenas 1 *tag* com baixa expressão (2 cópias)

similar à uma XET de *Populus* foi encontrada na biblioteca. O mesmo foi demonstrado por Carvalho et al. (2007) que também encontrou apenas uma *tag* para essa enzima, porém com uma maior expressão em biblioteca SAGE da região cambial de árvores juvenis de eucalipto.

As *expansins* (expansinas) são proteínas que se localizam na parede celular e tem a habilidade de dar plasticidade a rede de celulose-hemicelulose das paredes primárias (GRAY-MITSUMUNE et al, 2004). As plantas possuem duas famílias multigênicas dessa proteína, conhecidas como α -expansinas e β -expansinas, que apresentam uma conservação de sequência relativamente baixa (COSGROVE, 2000). Alguns pesquisadores têm sugerido que as expansinas podem ter diferentes papéis durante o desenvolvimento de um órgão, apresentando expressão diferencial dependendo do tecido ou estímulo imposto (HERNÁNDEZ-NIDAL et al., 2006). Na biblioteca estudada foi encontrada uma única *tag* (7 cópias) com similaridade a β -expansina de *Arabidopsis* e *Oryza*. Diferentemente, Carvalho et al (2007) encontrou apenas uma α -expansina na região cambial de árvores juvenis de eucalipto, sugerindo, portanto que na fase inicial do crescimento em eucalipto as β -expansinas possam ter maior importância.

Além da rede de celulose-hemicelulose, a matriz de pectina é também importante na regulação do afrouxamento da parede celular, ocorrendo mudanças nos polissacarídeos pécticos durante o crescimento (HERNÁNDEZ-NIDAL et al., 2006). As pectinas são secretadas em uma forma altamente metil-esterificadas. Tal esterificação mascara as cargas dos grupos carboxila e impede ligações de cálcio entre as pectinas. Uma vez na parede, os grupos ésteres são removidos pelas PMEs, revelando as cargas dos grupos carboxila e aumentando a capacidade da pectina em formar um gel rígido (TAIZ; ZEIGER, 2004). Para PMEs foram encontradas três *tags*, sendo: duas similares a *PME1* de *Arabidopsis* com 6 e 2 cópias e uma *tag* similar a *PME1* de eucalipto também com 2 cópias (Figura 19). Através do alinhamento entre os três ESTs foi possível verificar que correspondem a genes diferentes.

As *Polygalacturonases* (poligalacturonases), enzimas que despolimerizam pectinas desmetiladas pela ação das PMEs, foram representadas por uma única *tag* similar a *PG1* de *Eucalyptus globulus* também com baixa expressão (2 cópias) (Figura

19). Foram também encontradas duas *tags* para a enzima *Pectate lyase* - PL (pectato liase), sendo uma *tag* (8 cópias) similar ao gene *PL2* de *Prunus* e uma *tag* (3 cópias) similar ao gene *PL1* de *Eucalyptus globulus* (Figura 19). Através do alinhamento entre os ESTs das duas PLs também foi possível verificar que correspondem a genes diferentes.

É possível observar que entre as enzimas relacionadas à modificação da parede celular, foram encontradas diversas *tags* para aquelas associadas à modificação dos componentes ricos em pectina da parede, porém na maioria dos casos com baixa expressão. Em contraste, Andersson-Gunnerås et al. (2006) observaram uma elevada expressão de transcritos para pectinesterases e pectato liases em madeira de tensão de *Populus* durante crescimento secundário. Da mesma forma, Geisler-Lee et al. (2006) verificaram uma elevada expressão para poligalacturonases na região cambial de *Populus* sob condições normais de crescimento. Recentemente, Carvalho et al. (2007) também encontrou poucos transcritos para estas enzimas na região cambial de árvores juvenis de eucalipto. Uma possível explicação para o pequeno número de *tags* em alguns casos de enzimas relacionadas à expansão celular na biblioteca, poderia ser o a ausência de ESTs para estas enzimas disponíveis publicamente nos bancos de dados.

4.3.3 Deposição da parede celular

4.3.3.1 Biossíntese de Hemicelulose e Pectina

A expansão celular ocorre, portanto, durante a síntese da parede primária terminando com o início da deposição da parede secundária e lignificação. O metabolismo dos açúcares nucleotídeos fornece os precursores para a biossíntese dos polissacarídeos da parede celular durante o processo de formação da madeira. A síntese da maioria desses polissacarídeos é realizada a partir do precursor comum UDP-D-glicose que pode ser dirigido para a síntese de hemiceluloses e pectinas e também para a síntese de celulose. Dessa forma, uma série de reações de desidrogenação, descarboxilação e epimerização que utilizam açúcares nucleotídeos são responsáveis pela síntese de glicanos e pectinas a partir de UDP-D-glicose (REITER; VANZIN, 2001). Na Figura 20 é apresentado um resumo dessas reações

Figura 20 - Reações de interconversões de açúcares nucleotídeos e biossíntese de hemicelulose, pectina e celulose. As enzimas para os passos enzimáticos nos quais foram encontradas *tags* estão representadas na figura com setas pretas e todas as *tags* encontradas para uma enzima são apresentadas. *Tags* seguidas de um asterisco (*) sobrescrito representam isoformas de um mesmo gene. Setas cinza foram utilizadas para representar os passos enzimáticos para os quais nenhuma *tag* foi encontrada

Apesar da grande importância na estrutura da parede celular, poucos genes responsáveis pela biossíntese de pectinas foram identificados até o momento. Foi encontrado na biblioteca, apenas uma *tag* com 5 cópias associada ao gene *PttGTD8* de *Populus*, que é uma glicosil transferase pertencente ao mesmo sub-grupo do gene *QUASIMODO1* (QUA1) de *Arabidopsis*. Tem sido sugerido que o gene QUA1 está envolvido na biossíntese de pectinas devido a uma deficiência fenotípica de homogalacturonanos exibida por mutantes de *Arabidopsis* (BOUTON et al., 2002). Outra indicação de que o gene *GTD8* está envolvido na biossíntese de pectinas foi

demonstrada através de estudo em madeira de tensão em *Populus* onde a redução da expressão do gene *PttGTD8* foi positivamente correlacionada a redução no conteúdo de hemicelulose (Andersson-Gunneras et al., 2006). Aspeborg et al. (2005), demonstraram que existe uma co-expressão desse mesmo gene com os genes *CesA* relacionados a parede celular. Da mesma forma, Carvalho et al. (2007) também verificaram uma baixa expressão para esse gene na região cambial de árvores juvenis de eucalipto.

Alguns monossacarídeos das cadeias de glucanos e pectinas podem ser sintetizados através do UDP-glucuronato. Este representa o principal ponto de ramificação na biossíntese dos açúcares nucleotídeos (REITER; VANZIN, 2001), sendo considerado o intermediário central na biossíntese de muitos dos precursores dos polissacarídeos da parede celular. O UDP-glucuronato pode ser produzido a partir da UDP-D-glicose por meio da *UDP-glucose dehydrogenase* (UGDH) ou, alternativamente a partir do mio-inositol por meio da *myo-inositol oxigenase*. Todavia, nenhuma *tag* representando as enzimas chaves para a rota da mio-inositol foi encontrada na biblioteca. A predominância da UGDH foi verificada em meristemas, raízes em crescimento e tecidos vasculares de *Arabidopsis*, enquanto a rota de oxigenação da mio-inositol foi predominante em hipocótilos e cotilédones de plantas com 5 dias de desenvolvimento (SEITZ et al., 2000). Juntamente com o que foi demonstrado por Carvalho et al. (2007) que também não encontrou nenhum transcrito para enzimas da rota da mio-inositol na região cambial de árvores juvenis de eucalipto, a ausência de *tags* para essa enzima sugere uma prioridade da produção de UDP-glucuronato via conversão de UDP-D-glicose pela UGDH durante o processo de formação da madeira em eucalipto.

Duas enzimas que participam da rota de interconversões entre os açúcares nucleotídeos apresentaram uma expressão relativamente alta dentro da biblioteca, a UGDH, responsável pela produção de UDP-glucuronato e a *UDP-glucuronate decarboxylase* (UXS), responsável pela produção de UDP-xilose (Figura 20).

A UGDH é a principal enzima desta via, uma vez que regula o fluxo de UDP-D-glicose para a síntese de glicanos e pectinas funcionando no direcionamento da via como um passo limitante para a síntese dos polissacarídeos da parede celular (DALESSANDRO; NORTHCOTE, 1977). Entre as 3 isoformas da UGDH que estão

presentes em *Populus*, apenas uma é mais expressa durante a formação da parede celular secundária (HERTZBERG et al. 2001). Todavia, foi encontrada na biblioteca apenas uma *tag* para UGDH contendo 10 cópias (Figura 20), similar a UGDH de *Populus* que é preferencialmente expressa durante a diferenciação do xilema (JOHANSSON et al., 2002).

Para a UXS foram encontradas duas *tags* na biblioteca, uma *tag* com 8 cópias similar a forma solúvel do gene *AtUXS3* de *Arabidopsis* e uma *tag* com 4 cópias similar ao gene associado à membrana do Golgi *AtUXS4* de *Arabidopsis* (Figura 20) (HARPER, BAR-PELED, 2002). Carvalho et al. (2007), encontrou duas *tags* para o gene *AtUXS4* e, através de alinhamento, mostrou que na verdade eram isoformas do mesmo gene. Na biblioteca das plantas de 6 meses de eucalipto apenas uma das isoformas para o gene *AtUXS4* foi encontrada, sugerindo portanto, que em fases mais avançadas do desenvolvimento, talvez outras isoformas do gene sejam requeridas.

Um interessante resultado foi à identificação de três *tags* para a enzima *Galactinol synthase* (galactinol sintase) entre os genes relacionados ao metabolismo dos açúcares nucleotídeos (Figura 20). Todas as *tags* apresentaram alta similaridade a galactinol sintase 1 de milho, e através do alinhamento e de uma análise mais detalhada dos ESTs referentes as *tags* foi possível verificar que a *tag* de 8 cópias e a *tag* de 5 cópias são isoformas de um mesmo gene, uma vez que a *tag* de 5 cópias estava presente em ambos ESTs. A galactinol sintase é uma enzima chave na via que leva à formação de rafinose e outros oligossacarídeos (RFOs). Esta enzima catalisa a síntese de galactinol, a partir de UDP-galactose ou 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, enzima que esteve representada por uma *tag* com 4 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 RFO 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 não foram encontrados transcritos para enzimas envolvidas na síntese de amido.

De maneira geral, o nível de expressão das *tags* para as enzimas envolvidas na síntese de glicanos e pectinas foi relativamente baixo quando comparado ao nível de expressão das *tags* para as enzimas relacionadas à síntese de celulose e lignina, sugerindo novamente que o material amostrado possivelmente se encontrava em ativo crescimento secundário.

4.3.3.2 Biossíntese de Celulose

Atualmente existem dois modelos propostos para a síntese das microfibrilas de celulose, sendo um mecanismo direto e um mecanismo indireto de síntese. No mecanismo direto uma glicosiltransferase adiciona milhares de resíduos de glicose nas longas cadeias lineares de β -(1 $\rightarrow$ 4) D-glicose 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:esterol glicosil transferase possivelmente está envolvida na síntese de celulose através da síntese de um pequeno sitoesterol- β -glicosídeo que serve como *primer* para a elongação da cadeia β -1,4-glucano catalisada pelas CESAs. O produto é então clivado do *primer* por uma *Cellulase endo-1,4- β -glucanase* (EG) 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). Suportando essa idéia, uma EG foi mostrada estar preferencialmente expressa no xilema secundário de eucalipto (Paux et al., 2004). Duas diferentes *tags* referentes a ESTs de *Eucalyptus* com alta similaridade à EG foram identificadas na biblioteca. A *tag* de maior expressão (10 cópias) apresentou alta similaridade ao gene *EG2* de *Eucalyptus globulus*, ao passo que a *tag* de menor expressão (8 cópias) apresentou uma alta similaridade ao gene *EG1* de *Eucalyptus globulus* (Figura 21).

Figura 21 - Direcionamento de sacarose para síntese de celulose. As enzimas para os passos enzimáticos nos quais foram encontradas *tags* estão representadas na figura com setas pretas e todas as *tags* encontradas para uma enzima são apresentadas. *Tags* seguidas de um asterisco (*) sobrescrito representam isoformas de um mesmo gene

Além de ser utilizado como precursor na síntese de glicanos e pectinas, o UDP-D-glucose é também o precursor para a síntese das microfibrilas de celulose (Figura 20 e Figura 21).

Embora o UDP-D-glucose possa ser produzido tanto a partir de D-glucose-1-P pela *UDP-glucose pyrophosphorylase* (UGP), quanto a partir de sacarose pela *Sucrose synthase* (SUSY), o nível de expressão para a SUSY foi maior do que o nível de expressão apresentado pela UGP (Figura 21). Este resultado certamente corrobora com o modelo atualmente mais aceito para a biossíntese de celulose onde complexos protéicos de membrana formados pela associação de proteínas CESA, KORRIGAN (EG) e SUSY otimizam a síntese das microfibrilas de celulose (JOSHI et al., 2004). Neste modelo a SUSY possui a importante função de direcionar UDP-glucose diretamente para as proteínas CESA do complexo, evitando portanto, competição pelo *pool* celular de UDP-glucose e possibilitando, conseqüentemente, as elevadas taxas de síntese de celulose observadas durante o crescimento secundário (AMOR et al., 1995).

Foram identificadas na biblioteca duas *tags* similares ao gene *SUSY1* de *E. grandis* uma delas com uma elevada expressão (20 cópias) e uma delas com expressão muito baixa (2 cópias). Entretanto, através de alinhamento foi possível verificar que as *tags* representam genes diferentes e não isoformas de um mesmo gene. Recentemente foi demonstrado que os transcritos para a SUSY foram os mais abundantes transcritos para as CAZymes durante a formação da madeira normal e também de tensão em *Populus* (GEISLER-LEE et al., 2006; ANDERSSON-GUNNERÅS et al., 2006).

A alta expressão de transcritos para SUSY associada ao fato de que nenhuma *tag* referente às enzimas envolvidas na síntese e degradação de amido foi identificada na biblioteca constitui-se num forte indicativo de que a atividade da SUSY é fundamental para produzir UDP-glicose e que o *pool* desse açúcar nucleotídeo produzido está sendo direcionado para a síntese de celulose e não para a síntese de compostos de reserva. Este resultado confirma 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 em *Populus* (GEISLER-LEE et al., 2006).

No caso das CESAs, recentemente foram identificados seis genes 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 *EgCesA3* (relacionado à parede secundária) e *EgCesA6* (relacionado a parede primária) para todos os outros genes foram encontradas *tags* na biblioteca (Figura 21). No entanto, enquanto uma única *tag* foi identificada para *EgCesA2*, *EgCesA4* e *EgCesA5*, foram encontradas duas *tags* para *EgCesA1*. Através da análise do alinhamento produzido pelos ESTs de eucalipto foi possível observar que no caso da *EgCesA1* 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* relacionadas à CESA, apresentando 22 cópias, ao passo que todas as demais *tags* relacionadas à CESA apresentaram número de cópias inferior a 4. Dessa forma, os resultados observados são condizentes aos previamente reportados no sentido de que

transcritos para CESA relacionados à síntese de parede secundária são mais expressos na região cambial de *Eucalyptus grandis*. Entretanto são contrastantes ao revelar um nível de expressão bastante próximo para *EgCesA2* (relacionado à parede secundária) e *EgCesA4* e *EgCesA5* (ambos relacionados à parede primária). Além disso, o gene de maior expressão relatado por RANIK; MYBURG, 2006, foi *EgCesA3* e não *EgCesA1* como encontrado no presente trabalho. Uma hipótese que explicaria esta discordância é a divergência que existe na estrutura genética entre os indivíduos de uma população natural, que foram utilizados pelos autores e os indivíduos da população utilizada no presente trabalho que representam mudas melhoradas, desenvolvidas geneticamente para serem destinadas a um plantio comercial para a produção de papel e celulose. Resultados anteriores de RANIK et al. (2006) também suportam esta hipótese. Os autores realizaram um estudo do perfil transcricional de plantas altamente produtivas de eucalipto 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 e Myburg (2006), são similares à *EgCesA1*.

De modo geral, os níveis de expressão para KORRIGAN, *EgCesA1* e SUSY na biblioteca foram muito próximos, e também estão de acordo com o encontrado por Carvalho et al. (2007) na região cambial de árvores juvenis de *eucalipto* corroborando com o atual modelo proposto para a biossíntese de celulose que envolve não somente as proteínas CESA, mas também KORRIGAN e SUSY (JOSHI et al., 2004).

4.3.4 Lignificação

A Figura 22 apresenta um resumo da via de síntese dos fenilpropanóides, que é comum a biossíntese dos precursores da lignina, contendo as *tags* encontradas para cada enzima bem como o número de cópias de cada uma delas na biblioteca.

Figura 22 - Biossíntese e polimerização dos monolignóis. As enzimas para os passos enzimáticos nos quais foram encontradas *tags* estão representadas na figura com setas pretas e todas as *tags* encontradas para uma enzima são apresentadas. *Tags* seguidas de um asterisco (*) sobrescrito representam isoformas de um mesmo gene. Setas cinza foram utilizadas para representar os passos enzimáticos para os quais nenhuma *tag* foi encontrada

A rota de biossíntese da lignina, também conhecida como via dos fenilpropanóides inicia-se com a desaminação da fenilalanina a *cinnamic acid* pela *Phenylalanine ammonia-lyase* (PAL). Foram identificadas três *tags* para a PAL na biblioteca (Figura 22), representando dois genes distintos. A *tag* de maior expressão (8 cópias) apresentou alta similaridade ao gene *PAL* de *Eucalyptus globulus* e *PAL1* de *Populus* e de *Arabidopsis*. A *tag* que apresentou 3 cópias apresentou alta similaridade ao gene *PAL1* de *Populus* e de *Arabidopsis*, já a *tag* que apresentou 2 cópias apresentou alta similaridade ao gene e *PAL1* de *Populus* e *PAL2* de *Arabidopsis*. Por meio de alinhamento dos ESTs, foi possível identificar a ocorrência de duas isoformas produzidas por *splicing* 3' a partir de um único gene para as duas *tags* de menor expressão. Ambos *AtPAL1* and *AtPAL2* são considerados os genes mais importantes

para a síntese de lignina durante a lignificação vascular entre todos os genes PAL de *Arabidopsis* (RAES et al., 2003).

No passo seguinte, através da reação catalisada pela *Cinnamate 4-hydroxylase* (C4H), o *cinnamic acid* é convertido a *p-coumaric acid* (representando a primeira oxidação na via dos fenilpropanóides com redução de um NADPH liberado pelo citocromo P-450) que é utilizado pela *4-coumarate:coenzyme 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). Foi encontrada apenas uma *tag* com 6 cópias para a C4H na biblioteca (Figura 22) com alta similaridade ao gene *PtreC4H2* de *Populus* e que ao contrário do gene *PtreC4H1* foi demonstrado estar mais expresso no xilema do que no floema (LU et al., 2006).

Em *Populus tremuloides* foram identificados dois genes para C4H com duas isoformas cada, sendo a expressão destas isoformas relacionadas 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 aqui encontrados, sugerem que um gene similar a *PtreC4H2* seja especificamente importante na região cambial e provavelmente, na lignificação de células do xilema de *Eucalyptus grandis*, semelhantemente ao resultado encontrado por Carvalho et al. (2007) na região cambial de árvores juvenis de eucalipto.

A expressão dos genes para 4CL mostram-se essenciais para a manutenção do conteúdo total e da composição dos monômeros de lignina em tabaco (KAJITA et al., 1996). Já foram identificadas várias isoformas da enzima, as quais utilizam diferentes substratos e apresentam atividade célula e tecido específica, em várias espécies. Em *Arabidopsis* observa-se uma única cópia para 4CL e sua ativação ocorre no início do desenvolvimento da semente durante a deposição de lignina nos cotilédones e na raiz (LEE et al., 1995), já em *Populus* foram encontradas duas isoformas de 4CL, *Pt4CL1* e *Pt4CL2*, que estão relacionadas com a síntese de lignina (HU et al., 1998). Estas duas isoformas apresentam diferentes estruturas e função, mostrando afinidade por substratos específicos. A isoforma *Pt4CL1* foi relacionada com a biossíntese das sub-

unidades G e S da lignina. Já a isoforma *Pt4CL2* mostrou maior afinidade por substratos que levam a formação de outros fenilpropanóides. Hibridizações *in situ* demonstraram que a expressão dos dois genes relacionados ocorre de maneira tecido-específica. Enquanto a *Pt4CL1* se expressa em tecidos do xilema com lignificação, a *Pt4CL2* pode ser visualizada em camadas da epiderme do caule e de folhas. Estes dois genes apresentaram apenas 63% de homologia em suas seqüências apresentando expressão compartimentalizada, o que pode ser considerado uma estratégia da planta para a formação de diversos compostos a partir de uma mesma rota metabólica (ALLINA et al., 1998). Assim como para a C4H, foi encontrada apenas uma *tag* com 4 cópias para a 4CL na biblioteca (Figura 22) com alta similaridade ao gene *4CL1* de eucalipto e *Pt4CL1* de *Populus*, sugerindo também que um gene similar a *Pt4CL1* seja especificamente importante na região cambial e provavelmente, na lignificação de células do xilema de *Eucalyptus grandis*.

As enzimas PAL, C4H e 4CL estão presentes como pequenas famílias multigênicas nas 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). Foi demonstrado que a região promotora da C4H possui elementos-CIS também presentes nas enzimas PAL e 4CL, assumindo-se, portanto, que estejam sob a ação de um controle regulador semelhante. Isto é consistente com a mesma localização de RNA mensageiro e proteína destas três enzimas (RO et al., 2001). 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 eucalipto 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, e também com o que foi observado por Carvalho et al. (2007) na região cambial de árvores juvenis de eucalipto, foi observada na biblioteca uma maior expressão dos transcritos para enzimas localizadas nos passos posteriores (mais

específicos) da via do que para as enzimas das reações iniciais, principalmente para a *Caffeoyl-CoA O-methyltransferase* (CCoAOMT), a *Caffeic acid 3-O-methyltransferase* (COMT), a *Laccase* e a *S-adenosylmethionine synthetase* (SAMS).

Da mesma forma, o número de transcritos observados na biblioteca para COMT, CCoAOMT e SAMS foi também muito maior que o número de transcritos observados para a enzima CCR, para a qual não foi encontrada nenhuma *tag* e, para a enzima CAD, para a qual foi observada apenas uma *tag*. Nas angiospermas, a madeira é composta por grandes quantidades de unidades de lignina 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, era esperado que as enzimas direcionadas a síntese das unidades S e G apresentassem maior expressão do que as enzimas direcionadas à síntese de unidades H em eucalipto. Além disso, PAUX et al., (2005) demonstraram que os 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.

A CCR é a enzima que inicia a conversão das *Hidroxycinnamoyl-CoA* provenientes da via comum dos fenilpropanóides em monômeros de lignina. Em *Arabidopsis thaliana*, já foram identificados dois genes para esta mesma enzima (LAUVERGEAT et al, 2001), os genes *AtCCR1* e *AtCCR2*, que apresentaram 81,6% de homologia nas sequências de proteína e demonstraram possuir diferente afinidade por substratos. Suas expressões foram notoriamente diferenciais no tempo e no espaço. Enquanto o gene *AtCCR1* mostrou alta atividade em tecidos do caule no processo de lignificação durante o desenvolvimento, *AtCCR2* apresentou alta expressão apenas em tecidos submetidos à infecção por *Xanthomonas campestris* pv. *campestris*. Portanto, verifica-se atividade diferencial dos dois genes: enquanto *AtCCR1* mostrou-se atuante para a síntese de lignina durante o desenvolvimento normal da planta, *AtCCR2* foi ativado em processo de resposta a patógenos e resistência. 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 espécies de eucalipto, existe apenas um gene identificado para a

enzima (LACOMBE et al., 1997), porém nenhuma *tag* referente à CCR foi identificada na biblioteca, talvez devido a sua baixa expressão (Figura 22). Na realidade, foi encontrada uma única *tag* para a CCR, apresentando apenas 1 cópia, sendo que de acordo com os critérios utilizados no presente trabalho essa *tag* não foi considerada.

Para a enzima CAD foi encontrada apenas uma *tag* com elevada expressão (19 cópias) similar ao gene CAD de *Eucalyptus globulus* (Figura 22). Tanto a CAD como a CCR, parecem ter em muitas espécies uma cópia única do gene, pois o antisense dessas afeta de forma significativa o conteúdo total de lignina (BOUDET et al., 2000). É interessante observar que Carvalho et al. (2007) nos estudos envolvendo a região cambial de árvores juvenis de *eucalipto* sugeriu a existência de três genes para a CAD, onde a *tag*/gene com maior expressão foi o mesmo presente na biblioteca das plantas de 6 meses. Esse resultado sugere que esse gene CAD seja de fato o mais importante durante a lignificação nas árvores de eucalipto principalmente no início do seu desenvolvimento.

O passo determinante para o direcionamento de *p-coumaroyl-CoA* para a síntese de unidades G e S, é sua conversão ao *p-coumaroyl shikimic acid/acid quinic* pela *hydroxycinnamoyltransferase* (HCT), já que a C3H (*p-coumarate 3-hydroxylase*) não é capaz de utilizar *p-coumaroyl-CoA* como substrato (SCHOCH et al., 2001). Foi demonstrado que a enzima HCT catalisa a formação dos substratos *p-coumaroyl shikimic acid/acid quinic* e do precursor de *caffeoyl-CoA* (HOFFMANN et al., 2004). Foi identificada apenas uma *tag* com 8 cópias na biblioteca cujo EST apresentou similaridade ao gene HCT de tabaco (HOFFMANN et al., 2004).

Para a enzima C3H foram encontradas duas *tags* na biblioteca, uma *tag* com 5 cópias e outra com 4 cópias, ambas com alta similaridade ao gene *C3H1* de *Pinus* e a classe de genes P450 de *Arabidopsis*, cuja expressão é evidente em células vasculares em lignificação (NAIR et al., 2002). O alinhamento e análise dos ESTs das duas *tags*/genes demonstrou que uma das *tags* estava presente em ambos ESTs sugerindo portanto se tratar de isoformas para o mesmo gene. Carvalho et al (2007) nos estudos envolvendo a região cambial de árvores juvenis de eucalipto encontrou apenas um transcrito para a C3H, sugerindo que diferentes isoformas possam ser requeridas durante o processo de lignificação em eucalipto. De maneira alternativa, a C3H pode

também atuar diretamente sobre o *p-coumaric acid* produzindo *caffeic acid*, que pode se transformar em *ferulic acid* através da ação da COMT, ou se transformar em *caffeoyl-CoA* através da ação da 4CL (Figura 22).

Durante a biossíntese dos monolignóis, as reações que envolvem metilação, são requeridas em duas posições do anel aromático da unidade do monolignol. Uma ocorre na posição 3 e a outra na posição 5, sendo que a metilação na posição 3 leva à formação de unidades de lignina tipo G enquanto a dupla metilação nas posições 3 e 5 levam à formação de unidades de lignina tipo 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, acreditava-se que a COMT convertia *caffeic acid* em *ferulic acid* (DIXON, 2001), evidências posteriores demonstraram que a COMT catalisa preferencialmente a conversão de *5-Hydroxyferulic acid*, *5-Hydroxyconiferaldehyde* e *5-Hydroxyconiferyl alcohol* em *sinapic acid*, *sinapaldehyde* e *sinapyl alcohol*, respectivamente, e, portanto, sua atividade ocorre preferencialmente sobre os produtos da reação catalisada pela F5H (*Ferulic acid/coniferaldehyde 5-hydroxylase*) (PARVATHI et al., 2001) (Figura 22).

Entretanto, as *tags* para F5H e COMT apresentaram padrões de expressão distintos. Foi encontrada apenas uma *tag* com 5 cópias para a F5H enquanto para a COMT foram encontradas duas *tags* com 9 e 6 cópias respectivamente (Figura 21). 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 protéicas de COMT foram detectadas em todos os estádios avaliados (JUAN et al., 2006).

A *tag* para COMT com maior número de cópias (9), apresentou uma alta similaridade ao gene *OMT* de *Eucalyptus gunnii* (POEYDOMENGE et al., 1994) enquanto à segunda *tag* (6 cópias) apresentou similaridade com o gene *AtOMT1* de *Arabidopsis*. O alinhamento das sequências dos dois ESTs não identificou nenhuma região de identidade sugerindo que dois genes para COMT sejam preferencialmente expressos durante a formação do xilema em *Eucalyptus grandis*. A repressão de COMT em plantas transgênicas de *Populus* causou o desaparecimento quase total das

unidades siringil e aumentou o conteúdo das unidades guaiacil, porém não houve redução no conteúdo total de lignina (LAPIERRE et al., 1999). Com isto demonstrou-se a necessidade da COMT para a síntese de siringil, mas não da mesma forma para guaiacil para a qual há, aparentemente, outras enzimas que podem suprir o papel de COMT, sendo a enzima CCoAOMT uma forte candidata.

Até o fim da década de 80, o papel da CCoAOMT, estava relacionado somente com respostas de defesa da planta. Porém, Zhong et al. (2000a), estudaram o mutante *elp1* em *Arabidopsis*, deficiente para CCoAOMT e observaram a repressão da via biossintética da lignificação. Desde então, esta enzima passou a ser considerada uma enzima chave na produção dos monolignóis, atuando na metilação de *caffeoyl-CoA* gerando *feruloyl-CoA*.

Altos níveis de expressão de transcritos para CCoAOMT foram encontrados em células de *Populus* durante o processo de lignificação (ZHONG et al., 2000b). Em estudo recente, realizado sobre o perfil transcricional de células do xilema de eucalipto, 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 os quais estavam presentes dois genes para lacases e um gene para CCoAOMT (PAUX et al., 2004). Da mesma maneira, proteínas CCoAOMT foram também identificadas como uma das mais abundantes proteínas durante a diferenciação do xilema em *Populus* (VANDERMIJNSBRUGGE et al., 2000). No presente trabalho, entre os transcritos identificados para as enzimas envolvidas na rota de biossíntese de lignina, os transcritos para CCoAOMT foram os mais abundantes, apresentando um total de 37 cópias na biblioteca. 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 (11 cópias) similar a *EgCCoAOMT2* de *E.globulus* e a de maior expressão (26 cópias) similar a *EgCCoAOMT1* (Figura 21). Como a expressão para a CCoAOMT foi bem maior do que a expressão para a COMT, pode-se sugerir que a síntese das unidades S e G de lignina nas plantas de 6 meses de eucalipto estejam sendo produzidas de maneira preferencial diretamente a partir de *caffeoyl-CoA*.

O processo de conversão do *cinnamic acid* para as unidades de lignina G e S requer adição de grupos metil em posições específicas dos monolignóis, levando a um grande consumo desses grupos 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-adenosylmethionine synthase* (SAMS), é ativa em muitos processos celulares sendo consequentemente considerada uma enzima *housekeeping* (HANSON; ROJE, 2001). 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). Foram encontradas três *tags* para a SAMS na biblioteca, com similaridade ao gene *SAMS* de *Arabidopsis* perfazendo um total de 31 transcritos (Figura 22), estando de acordo com resultados anteriores, os quais 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 (grupos metil) 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 as três *tags* encontradas para SAMS sugere que 2 genes sejam preferencialmente expressos para esta enzima durante a lignificação em *Eucalyptus grandis*. Muito provavelmente a *tag* com 20 cópias e a *tag* com 6 cópias são isoformas de um mesmo gene, apresentando apenas uma mudança em uma única base da *tag*, que pode ser responsável pelo aparecimento da isoforma de maior expressão.

Da mesma forma que o proposto por Carvalho et al (2007) nos estudos envolvendo a região cambial de árvores juvenis de eucalipto, foi interessante notar que os níveis de expressão dos transcritos para SAMS e CCoAOMT estiveram muito próximos. Sugerindo, portanto, que a correlação funcional das duas enzimas possa determinar um controle transcricional comum para ambas, embora nenhuma evidência neste sentido tenha sido ainda encontrada.

Um aspecto interessante é que o intenso processo de lignificação da parede celular dos elementos de vaso e das fibras que ocorre durante o crescimento

secundário requer um alto consumo de nitrogênio (N) através da desaminação da fenilalanina pela PAL liberando amônia. Por si só, o conteúdo de N consumido nas células que estão em intenso processo de lignificação já seria suficiente para causar uma deficiência nutricional na planta, caso todo N consumido fosse suprido apenas pela assimilação direta através da redução do nitrato (CÁNTON et al., 2005). Ao contrário disso, nas células em crescimento secundário ativo 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). Como sugerido por CÁNTON et al. (2005), este mecanismo local funcionaria através do sistema GS/GOGAT. Apoiando esta hipótese, foram encontradas duas *tags* para a *Glutamine synthetase* (GS) na biblioteca, uma *tag* com 22 cópias e uma *tag* com 5 cópias (Figura 22).

Nas plantas existem duas classes distintas de GS, a citosólica (GS1) e a cloroplastidial (GS2), sendo que na maioria das plantas, várias isoformas de GS1 são codificadas por uma família de genes (CREN; HIREL, 1999). Consistente com resultados anteriores, as *tags* encontradas na biblioteca foram similares ao gene GS1 de *Populus* e não foi identificada região de identidade entre estas sequências sugerindo que nas plantas de *E. grandis* com 6 meses de idade, existam pelo menos dois genes para a GS1.

O passo final na via de biossíntese da lignina consiste da polimerização dos monolignóis, que pode ser conduzida pelas *Laccases* (lacases) e/ou pelas *Peroxidases* (peroxidases) (BAUCHER et al., 2003). As lacases são enzimas que se localizam especificamente na parede celular e possivelmente estão associadas com o processamento de substratos fenólicos associados à parede e, sua importância 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). As peroxidases promovem a oxidação de monômeros de lignina em células que seguem a formação de parede celular secundária, estando também relacionadas com a destoxificação de H₂O₂, catabolismo de auxinas e resposta ao estresse devido à oxidação por compostos fenólicos (liberados quando ocorrem ferimentos ou infecção por patógenos).

Três *tags* com um total de 21 cópias, representando três genes diferentes, foram identificadas para lacases na biblioteca com similaridade a lacases de *Populus*, e duas *tags* de baixa expressão (2 cópias cada) foram identificadas para peroxidases com similaridade a peroxidases de *Arabidopsis* (Figura 22). Considerando o número de transcritos, este resultado sugere que as lacases são relativamente 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* e por Carvalho (2007) onde apenas um transcrito com baixa expressão foi encontrado para peroxidases na região cambial de árvores juvenis de eucalipto.

Outro aspecto interessante foi a presença de proteínas dirigentes (*dirigent proteins*) na biblioteca. As proteínas dirigentes são responsáveis pelo controle espaço temporal de deposição da classe de lignina adequada. Foram identificadas como participantes na síntese de lignina, por Gang et al. (1999), tiveram seus genes clonados em várias espécies (GANG et al., 1999) e, através de imunolocalização foram identificadas nas plantas em tecidos vasculares em desenvolvimento, especialmente na parede celular do xilema, indicando sua participação na lignificação. Recentemente foi proposto que as chamadas proteínas dirigentes possam atuar mediando a polimerização dos monômeros de lignina principalmente durante a lignificação em S1 e na lamela média de células do xilema, regiões onde normalmente inicia-se a biossíntese de lignina, todavia, a verdadeira função biológica destas enzimas ainda está sendo estudada (BURLAT et al., 2001).

Confirmando esta hipótese, foram encontradas três *tags* para proteínas dirigentes na biblioteca com uma expressão relativamente alta (20 cópias) (Figura 22). Duas *tags* de menor expressão (3 e 2 cópias) representaram duas isoformas similares ao gene DIR de *Forsythia x intermedia*, e a *tag* de maior expressão (15 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

ESTs não identifica região de identidade, possivelmente dois genes DIR distintos estejam atuando na polimerização de monolignóis durante a lignificação em *E. grandis*.

4.3.5 Morte celular programada (PCD)

Uma vez que a lignificação é completada, os elementos de vaso do xilema entram no último estágio de sua diferenciação, a chamada morte celular programada (PCD), seguida pela autólise do conteúdo celular. Foram identificadas na biblioteca *tags* para muitas das proteínas que vem sendo correlacionadas à morte celular programada e autólise celular em plantas, tais como, *Cysteine proteases*, *Aspartic proteases*, *SLP2* (*subtilisin serina protease*), *Nodulins*, reguladores de apoptose *Bcl-2* e *Bax inhibitor-1* e *RCD1* (*Radical-induced Cell Death*). Da mesma forma, foi encontrado um elevado número de transcritos para a via proteassomo-ubiquitina de degradação de proteínas a qual foi recentemente correlacionada a PCD (MOREAU et al., 2005).

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 (HELLMANN; 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 bem como para genes responsivos a auxinas, na biblioteca, 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 sugerindo mais

uma vez que o material amostrado encontra-se em processo de diferenciação celular e crescimento secundário.

4.4 Correlação entre nível de transcritos e abundância de proteínas

Os caminhos que levam um gene a expressar um determinado fenótipo passam por uma série de mecanismos de regulação, assim, a comparação entre o nível de expressão dos genes, na forma de RNA mensageiro e a abundância das proteínas torna-se necessária para conhecer e elucidar os processos envolvidos na regulação da transcrição e da tradução (GION et al., 2005). A combinação de tecnologias de análise da transcrição, como micro-arranjos ou SAGE, e as técnicas 2D-PAGE e espectrometria de massas da proteômica, no estudo de uma mesma amostra pode ajudar na compreensão dos mecanismos de regulação e controle envolvidos entre o gene e uma proteína (WATSON et al., 2003).

No presente trabalho foi realizada uma comparação entre os transcritos obtidos pela biblioteca SAGE e as correspondentes proteínas identificadas por 2D-PAGE/LC-MS/MS do mesmo material vegetal (ANDRADE, 2006).

Com a construção da biblioteca foi possível a identificação de 2.274 *tags* ou genes diferentes, ao passo que através da análise proteômica foram detectados 438 *spots* e um total de 168 *spots* foram retirados do gel, digeridos com tripsina e submetidos ao sequenciamento por espectrometria de massas através do sistema LC-MS/MS, sendo possível à identificação de apenas 102 *spots* diferentes (ANDRADE, 2006). O pequeno número de proteínas identificadas em relação ao número de *tags* pode ser devido ao fato de que o processo de excisão dos *spots* do gel foi realizado manualmente, favorecendo para que os *spots* ou proteínas de maior representatividade fossem preferencialmente amostrados e dificultando a amostragem das proteínas de menor abundância. Além disso, modificações pós-traducionais, a faixa de pH (4-7) amostrada, digestão incompleta pela tripsina e quantidades insuficientes para identificação pela espectrometria de massas. também podem ser apontadas como possíveis razões para a não identificação de um número maior de *spots* (CELEDON et al., 2007).

Na tabela 8, estão apresentados os números de *spots* identificados para cada proteína e a quantidade de *tags* identificadas para os genes correspondentes.

Tabela 8 – Proteínas identificadas a partir do caule de plantas de *E. grandis* com 6 meses de idade. São fornecidos: o número de *spots* identificado para cada proteína; o volume total em porcentagem ocupado por cada proteína; o número de *tags* (transcritos) para cada gene correspondente e o número de cópias para essas *tags*

(Continua)

Proteína	Nº Spots	Vol. total (%)	Nº de tags	Nº de cópias
Energia (Autotrófico/Carbono/ATP)				
Alcohol dehydrogenase	1	0,077	3	13
Mitochondrial ATP synthase (alpha/beta)	6	3,062	2	27
Vacuolar ATP synthase (A/B)	2	0,429	3	5
Oxygen-evolving enhancer protein 1	1	0,113	-	-
Glyceraldehyde-3-phosphate dehydrogenase cytosolic	3	0,537	5	73
Glyceraldehyde-3-phosphate dehydrogenase chloroplast	2	0,405	-	-
Malate dehydrogenase	3	0,390	2	5
RuBisCO large subunit	4	0,747	-	-
Quinone oxidoreductase-like protein	1	0,163	2	5
Succinate dehydrogenase	1	0,050	-	-
NADP-dependent malic enzyme	1	0,237	1	4
Phosphoglycerate dehydrogenase	1	0,360	1	2
Metabolismo de Carboidratos				
Enolase	3	2,304	1	4
Fructokinase	3	1,002	-	-
Phosphoglycerate kinase	3	3,01	-	-
UTP-glucose-1-phosphate uridylyltransferase	3	0,635	-	-
Triosephosphate isomerase	4	0,893	-	-
2,3-bisphosphoglycerate-independent phosphoglycerate mutase	1	0,618	-	-
Metabolismo de Nitrogênio				
Glutamine synthetase cytosolic (GS1)	2	1,5	3	28
Metabolismo de Aminoácidos				
Adenosylhomocysteinase	1	0,077	2	30
5 methyltetrahydropteroyltri-glutamate homocysteine transferase	1	0,296	2	36
Glutamate dehydrogenase	2	0,124	-	-
S-adenosylmethionine synthetase (SAMS)	4	7,146	3	31
Metabolismo Secundário				
Isoflavone reductase	7	4,092	5	137
Divisão Celular				
Putative cell division control protein 48 homolog D	1	0,666	-	-

Tabela 8 – Proteínas identificadas a partir do caule de plantas de *E. grandis* com 6 meses de idade. São fornecidos: o número de *spots* identificado para cada proteína; o volume total em porcentagem ocupado por cada proteína; o número de *tags* (transcritos) para cada gene correspondente e o número de cópias para essas *tags*

(Conclusão)

Proteína	Nº <i>Spots</i>	Vol. total (%)	Nº de <i>tags</i>	Nº de cópias
Regulação Celular				
14-3-3-like protein	5	0,842	3	13
Parede Celular				
Caffeic acid 3-O-methyltransferase (COMT)	3	1,41	1	6
Caffeoyl-CoA O-methyltransferase (CCoAOMT)	1	0,241	2	37
Cinnamyl alcohol dehydrogenase (CAD)	1	0,162	1	19
Sucrose synthase (SuSy)	2	0,167	2	22
Phenylalanine ammonia-lyase (PAL)	1	0,148	3	13
Citoesqueleto				
Actin	5	3,705	7	32
Profilin	1	0,156	2	30
α -Tubulin	1	0,804	3	35
β -Tubulin	1	0,770	6	82
RNAr e Proteínas Ribossomais				
60s acidic ribosomal protein P0	1	0,072	4	21
Organização dos Cromossomos				
Histone H4	1	0,314	1	4
Síntese/Degradação/Modificação de Proteínas				
Chaperone protein dnaK	1	0,120	-	-
Heat shock 70 kDa protein	6	3,834	1	6
Heat shock protein 80	1	0,140	-	-
Heat shock protein 83	1	0,182	-	-
Luminal-binding protein	2	0,527	-	-
Proteasome subunit alpha type 2	1	0,134	1	3
Elongation factor 2	2	0,4	1	4
Eukaryotic initiation factor 4A	1	0,248	1	8
Probable protein disulfide isomerase A6 precursor	1	0,241	1	2
Proteínas de Função Hipotética				
Hypothetical protein in nit 5'region	1	0,171	-	-
Probable monodehydroascorbate reductase	1	0,090	-	-
TOTAL	102	43,74	75	737

A partir da tabela 8 pode-se perceber que algumas enzimas para as quais foram encontrados *spots* de proteínas, não foi possível a identificação de nenhuma *tag*/gene correspondente na biblioteca SAGE. Entre essas proteínas encontram-se principalmente aquelas relacionadas ao metabolismo de carboidratos, como *Fructokinase*, *Phosphoglycerate kinase*, *UTP-glucose-1-phosphate uridylyltransferase*, *Triosephosphate isomerase* e *2,3-bisphosphoglycerate-independent phosphoglycerate mutase*, que são importantes para o suprimento de boa parte do carbono que é utilizado para a biossíntese de uma série de compostos do metabolismo secundário, aminoácidos, ácidos nucleicos e ácidos graxos (PLAXTON, 1996). Possivelmente, a ausência de *tags* para essas enzimas seja devida a falta de ESTs de baixa expressão nos bancos de dados.

Entretanto é possível notar também que para diversas enzimas importantes durante o processo de formação da madeira para as quais foram identificadas *tags*/genes, também foram identificadas proteínas (*spots*), como por exemplo a *Isoflavone reductase*, COMT, CCoAOMT, CAD, SUSY, PAL, SAMS, *Glutamine synthetase cytosolic* (GS1), *Actin*, *Profilin*, α -*tubulin* e β -*tubulin* (tabela 8). A grande presença em número de *spots* ou em volume (%) dessas proteínas, extremamente importantes para o crescimento secundário, confirma os dados obtidos através da construção da biblioteca SAGE e indica mais uma vez que o material amostrado no presente estudo estava em ativo crescimento secundário.

Atualmente, diferentes trabalhos têm procurado investigar a existência de uma correlação entre mRNA e proteínas em diferentes organismos. Dependendo do tecido e da espécie, esta correlação pode ser alta para alguns genes específicos, porém, baixa quando todos os genes são considerados (ANDERSON; SEILHAMER, 1997; GYGI et al., 1999; CHEN, G. et al., 2002).

A partir da figura 23, nota-se a existência de uma correlação muito fraca ($R^2=0,1037$) entre o nível de expressão dos transcritos (número de cópias) e o respectivo volume total de proteínas encontrado, uma vez que a maior parte dos pontos encontra-se distante da linha de tendência. Isso se deve ao fato de que para algumas proteínas abundantes, que foram representadas por mais de um *spot* e que possuem um alto volume (%), como por exemplo, *Fructokinase*, *Phosphoglycerate kinase*,

Triosephosphate isomerase, entre outras, nenhuma *tag* correspondente foi encontrada na biblioteca SAGE. Por outro lado, para algumas *tags*/gene que estiveram presentes em um número elevado ou que apresentaram um alto nível de expressão, como a α -*tubulin* e β -*tubulin* e *adenosylhomocysteinase*, apenas um único *spot* e/ou pequeno volume (%) foi encontrado.

Figura 23 - Correlação entre o nível de expressão dos transcritos (número de cópias) e volume total de proteínas (%) para os 102 *spots* (48 proteínas) identificados por espectrometria de massas

Em *Saccharomyces cerevisiae* a correlação entre a quantidade de mRNA e o nível de proteína demonstrou-se baixa e insuficiente para determinar o nível de expressão de proteínas a partir da quantificação dos transcritos, encontrando-se uma correlação entre mRNA e proteínas para apenas alguns genes específicos. Da mesma maneira, estudos utilizando as técnicas de micro-arranjos e 2D-PAGE envolvendo células pulmonares que apresentavam tumor em humanos obtiveram resultados semelhantes (GYGY et al., 1999; CHEN, G. et al., 2002). Gion et al. (2005), a partir da comparação entre o volume dos *spots* para as proteínas mais abundantes separadas por 2D-PAGE e sequenciadas por espectrometria de massas, com uma biblioteca de ESTs de xilema de *Pinus pinaster*, encontraram uma correlação positiva, porém fraca entre o nível de mRNA e as proteínas presentes, demonstrando que no tecido

xilemático dessa espécie existe uma tendência entre o aumento no nível de mRNA e o aumento na abundância das proteínas correspondentes.

A baixa correlação entre o número de transcritos e as proteínas indica que a simples quantificação de um determinado transcrito de mRNA não é um bom indicador para estimar a abundância de uma determinada proteína presente no caule de plantas de *Eucalyptus grandis* e vice-versa.

Assim sendo, podemos dizer que nesse caso, o estudo do transcritoma através da construção da biblioteca SAGE e o estudo do proteoma através da técnica LC-MS/MS, foi complementar, possibilitando a identificação de transcritos e proteínas conjuntamente para muitos genes, mas permitindo também a identificação ora de transcritos, ora de proteínas, alcançando dessa forma sucesso na tentativa de traçar um panorama do padrão de expressão gênica durante a formação da madeira de eucalipto.

5 CONSIDERAÇÕES FINAIS

Em uma era de rápidos avanços, as técnicas moleculares de alta velocidade (*high-throughput*) aliadas ao uso da bioinformática, têm feito com que os estudos genômicos em grande escala sejam executados rotineiramente em diversos laboratórios de todo o mundo. Todavia, apesar do importante papel das árvores para a economia mundial e também para o ambiente, o conhecimento sobre os processos envolvidos na formação da madeira ainda são muito pequenos.

As recentes publicações sobre genômica funcional da formação de madeira têm identificado diversos genes envolvidos no processo de xilogênese utilizando diferentes tecnologias para compreender melhor os processos envolvidos tanto na formação quanto na qualidade da madeira. Embora os resultados derivados desses estudos forneçam importantes informações sobre os mecanismos moleculares, eles ainda são insuficientes para esclarecer todos os processos envolvidos na formação de madeira, uma vez que a grande maioria dos estudos concentra-se em árvores com madeira juvenil ou adulta.

No presente trabalho, o uso da técnica de SAGE permitiu estabelecer um perfil transcricional dos genes associados a xilogênese (divisão, expansão, formação da parede celular secundária, lignificação, e morte celular programada) no caule de plantas de *Eucalyptus grandis* com 6 meses de idade.

A técnica de SAGE permitiu o acesso a transcritos de baixa abundância nas células, e as SAGE *tags* de 10 pb foram suficientemente específicas para identificar genes únicos, mesmo quando ESTs originados de diferentes tecidos e/ou bancos de dados foram utilizados na identificação dos genes. Com a construção da biblioteca 43,5% dos genes puderam ser identificados e ter uma possível função atribuída. Além disso, foi possível também identificar diferentes isoformas para um gene ou mesmo localização celular específica, como verificados em algumas situações.

Um aspecto que merece destaque é que 56,5% não puderam ser identificados, ou seja, para essas *tags* não foi possível encontrar um EST de origem nos bancos públicos de eucalipto (NCBI) ou no consórcio Genolyptus, reforçando a necessidade da liberação de um número maior de sequências disponíveis para acesso público.

A biblioteca SAGE foi caracterizada por um número similar de transcritos relacionados à categoria Metabolismo e Energia (27%) e à categoria Estrutura e Organização da Estrutura Celular (27%). A produção de energia para o crescimento secundário nas células cambiais esteve principalmente relacionada a 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 nas plantas de 6 meses. 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 elevada expressão dos transcritos para *Glutamine synthetase* (GS1) na biblioteca 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 das plantas de *E. grandis* de 6 meses de idade.

A categoria “Estrutura e Organização da Estrutura Celular” foi representada principalmente por transcritos relacionados a biossíntese da parede celular e por proteínas do citoesqueleto celular. Os genes mais expressos relacionados à biossíntese da parede celular evidenciam um grande investimento em crescimento secundário nessas plantas de 6 meses de idade. Uma informação interessante foi a de que os níveis de expressão dos genes CCoAOMT e SAMS foram bastante similares sugerindo que possivelmente a SAMS participe também do *cluster* de co-regulação transcricional formado por COMT e CCoAOMT. Os resultados sugerem fortemente um papel importante para a SAMS 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 com homologia a *Dirigent proteins*, sugerindo também que estas enzimas possivelmente sejam importantes na polimerização de monolignóis nas plantas de *E. grandis* de 6 meses de idade. Já entre os genes mais expressos relacionados à proteínas do citoesqueleto celular destacam-se as *Tubulins* (α e β), *Actins*, *Actin depolymerizing factors* e *Profilins*, responsáveis por determinar a orientação da deposição das microfibrilas de celulose na parede celular.

A comparação realizada entre o volume total dos *spots* das proteínas identificadas por espectrometria de massas através do sistema LC-MS/MS e o nível de expressão dos respectivos transcritos identificados na biblioteca SAGE demonstrou que existe uma correlação muito fraca entre o nível de mRNA e as proteínas do caule das plantas de *Eucalyptus grandis* com 6 meses. Entretanto, os resultados obtidos através da construção da biblioteca SAGE e os dados do proteoma foram consistentes para diversas enzimas e proteínas relacionadas ao processo de formação da madeira, confirmando os dados obtidos através da construção da biblioteca SAGE e indicando que o material amostrado no presente estudo estava em ativo crescimento secundário.

O eucalipto caracteriza-se por ser uma árvore que apresenta um extraordinário crescimento e produção de biomassa. A grande expressão dos genes relacionados a biossíntese da parede celular, proteínas do citoesqueleto celular, bem como de outros genes relacionados ao processo de formação da madeira, confirmam a alta atividade da região cambial para produção de xilema secundário nas plantas de eucalipto com 6 meses de idade garantido e sustentando dessa forma o rápido crescimento da árvore.

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.
- AHLFORS, R.; LANG, S.; OVERMYER, K.; JASPERS, P.; BROSCHE, M.; TAURIANEN, A.; KOLLIST, H.; TUOMINEN, H.; BELLES-BOIX, E.; PIIPPO, M.; INZE, D.; PALVA, E.T.; KANGASJARVI, J. *Arabidopsis* RADICAL-INDUCED CELL DEATH1 belongs to the WWE protein-protein interaction domain protein family and modulates abscisic acid, ethylene, and methyl jasmonate responses. **The Plant Cell**, Rockville, v.16, n. 7, p. 1925-37, 2004.
- 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.
- 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.; ZHANG, Z.; MILLER, W.E. 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.
- ANDERSON, L.; SEILHAMER, J. A comparison of selected mRNA and protein abundance in human liver. **Electrophoresis**, Weinheim, v. 18, p. 533-537, 1997.
- 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 Queiroz”, Universidade de São Paulo, Piracicaba, 2006.

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

ARIOLI, T.; BURN, J.E.; WILLIAMSON, R.E. Molecular biology of cellulose biosynthesis. In JAIN, S.M.; MINOCHA, S.C. (Ed.), **Molecular Biology of Wood Plants Vol 1**, Dordrecht: Kluwer Academic Publishers, 2000. p. 205-226.

ASPEBORG, H.; SCHRADER, J.; COUTINHO, P.M.; STAM, M.; KALLAS A.; DJERBI S.; NILSSON P.; DENMAN S.; AMINI B.; STERKY F.; MASTER E.; SANDBERG G.; MELLEROWICZ E.; SUNDBERG B.; HENRISSAT B.; TEERI T.T. Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. **Plant Physiology**, Lancaster, v. 137, p. 983-997, 2005.

ASSIS, T.F. Melhoramento genético do eucalipto. **Informe Agropecuário**, Belo Horizonte, v. 189, p. 32-51, 1996.

ASSIS, T.F.; MAFIA, R.G. Hibridação e clonagem. In: BORÉM, A (Ed.). **Biotecnologia Florestal**. Viçosa: Universidade Federal de Viçosa, 2007, cap. 5, p. 93-121.

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.

BAIROCH, A.; APWEILER, R.; WU, C.H.; BARKER, W.C.; BOECKMANN, B.; FERRO, S.; GASTEIGER E.; HUANG, H.; LOPEZ, R.; MAGRANE, M.; MARTIN, M.J.; NATALE, D.A.; O'DONOVAN, C.; REDASCHI, N.; YEH, L. S. The Universal Protein Resource (UniProt). **Nucleic acids research**, London, v. 1, 33, p.154-9, 2005.

BALK, P.A.; de BOER, A.D. Rapid stalk elongation in tulip (*Tulipa gesneriana* L. cv. Apeldoorn) and the combined action of cold-induced invertase and the water-channel protein gamma TIP. **Planta**, Berlin, v. 209, p. 346–354, 1999.

BATEMAN, A.; BIRNEY, E.; CERRUTI, L.; DURBIN, R.; ETWILLER, L.; EDDY, S.R.; GRIFFITHS-JONES, S.; HOWE, K.L.; MARSHALL, M.; SONNHAMMER, E.L.L. The Pfam protein families database. **Nucleic Acids Research**, London, v. 30, p. 276-280, 2002.

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

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.

BHANDARI, S.; FUJINO, E.; THAMMANAGOWDA, S.; ZHANG, D.; XU, F.; JOSHI, C.P. Xylem-specific and tension stress-responsive coexpression of KORRIGAN endoglucanase and three secondary wall-associated cellulose synthase genes in aspen trees. **Planta**, Berlin, v. 224, n. 4, p. 828-837, 2006.

BIDLACK, J.; MALONE, M.; BENSON, R. Molecular Structure and Component Integration of Secondary Cell Walls in Plants. **Proceedings of the Oklahoma Academy of Science**, Edmond, v. 72, p. 51-56, 1992.

BINNECK, E. As Ômicas: Integrando a bioinformação. **Biotecnologia Ciência & Desenvolvimento**, Brasília, v. 32, p. 28-37, jan./jun. 2004.

BOMMER, U.A.; THIELE, B.J. The translationally controlled tumour protein (TCTP)17. **The International Journal of Biochemistry & Cell Biology**, Exeter, v. 36, n. 3, p. 379-85, 2004.

BOUDET, A.M. Lignins and lignification: Selected issues. **Plant physiology and biochemistry**, Paris, v. 38, p. 81-96, 2000.

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. **The Plant Cell**, Rockville, v. 14, p. 3073-3088, 2002.

BOUTON, S.; LEBOEUF, E.; MOUILLE, G.; LEYDECKER, M.; TALBOTEC, J.; GRANIER, F.; LAHAYE, M.; HÖFTE, H.; TRUONG, H. *QUASIMODO1* Encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in *Arabidopsis*. **The Plant Cell**, Rockville, v. 14, p. 2577-2590, 2002.

BRACELPA. **Desempenho do Setor em 2006 e Projeção para 2007**. Disponível em: <<http://www.bracelpa.com.br>>. Acesso em: 10 fev. 2007.

BRASIL se prepara para o mercado de redução de emissões. **Revista da Madeira**, Curitiba, v. 15, n. 93, nov. 2005. Disponível em: <<http://www.remade.com.br/revista>>. Acesso em: 19 jan. 2007.

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.

BRITO, J.O.; BARRICHELO, L.E.G. Correlações entre características físicas e químicas da madeira e a produção de carvão vegetal: Densidade e teor de lignina da madeira de eucalipto. **IPEF**, Piracicaba, n. 14, p. 9-20, 1977.

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.

BROOKER, M. I. H. A new classification of genus *Eucalyptus* L'Her. (Myrtaceae). **Australian System Botany**, Sidney, v.13, p.79-148, 2000.

BROWN, G.R.; GILL, G.P.; KUNTZ, R.J.; LANGLEY, C.H.; NEALE, D.B. Nucleotide diversity and linkage disequilibrium in loblolly pine. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 101, p. 15255-15260, 2004.

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.

CALSA, T. Jr.; FIGUEIRA, A. Serial analysis of gene expression in sugarcane (*Saccharum* spp.) leaves revealed alternative C₄ metabolism and putative antisense transcripts. **Plant Molecular Biology**, Dordrecht, v. 63, n. 6, p. 745-762, 2007.

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.; TIERNEY, M.; CAMPBELL, M. Molecular biology of the plant cell wall: searching for the genes that define structure, architecture and dynamics. **Plant Molecular Biology**, Dordrecht, v. 47, p. 1-5, 2001.

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, n. 1, p. 1-30, 1993.

CARPITA, N.C.; McCANN, M. The Cell Wall. In: BUCHANAN, B.B.; GRUISSEM, W.; JONES, R.L. (Eds). **Biochemistry & Molecular Biology of Plants**. Rockville: American Society of Plant Physiologists, 2000. chap. 2, p. 52-108.

CARROL, A.D.; MOYEN, C.; VAN KESTEREN, P.; TOOKE, F.; BATTEY, N.; BROWNLEE, C. Ca, annexins, and GTP modulate exocytosis from maize root cap protoplasts. **The Plant Cell**, Rockville, v. 10, p. 1267-1276, 1998.

CARVALHO, M.C.C.G. **Caracterização de perfis transcricionais de folhas e da região cambial de *Eucalyptus grandis* usando o SAGE**. 2007. 153p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2007.

CARVALHO, M.C.C.G.; CALDAS, D.G.G.; CARNEIRO, R.T.; MOON, D.H.; SALVATIERRA, G.R.; FRANCESCHINI, L.M.; ANDRADE, A.; CELEDON, P.F.; ODA, S and LABATE, C.A. (2007). SAGE transcript profiling of *Eucalyptus grandis* juvenile cambial region. **New Phytologist**, London, 2007 (Submitted).

CELEDON, P.A.F.; ANDRADE, A.; MEIRELES, K.G.X.; CARVALHO, M.C.C.G.; CALDAS, D.G.G.; MOON, D.H.; CARNEIRO, R.T.; FRANCESCHINI, L.M.; ODA, S. LABATE, C.A. Proteomic analysis of the cambial region in juvenile *Eucalyptus grandis* at three ages. **Proteomics**, Weinheim, 2007. In press.

CHAFFEY, N. Cambium: old challenges-new opportunities. **Trees: Struture and Function**, Berlin, v. 13, p. 138-151, 1999.

CHAFFEY, N. Microfibril orientation in wood cells: new angles on an old topic. **Trends in plant science**, Oxford, v. 5, p. 360-362, 2000.

CHAFFEY, N. Why is there so little research into the cell biology of the secondary vascular system of trees? **New Phytologist**, London, v. 153, n. 2, p. 213-223, 2002.

CHAFFEY, N.; BARLOW, P.; BARNETT, J. A cytoskeletal basis for wood formation in angiosperm trees: the involvement of microfilaments. **Planta**, Berlin, v. 210, p. 890-896, 2000.

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, apr, 2002.

CHEN, G.; TAREK G. GHARIB, T. G.; HUANG, C. C.; TAYLOR, J. M. G.; MISEK, D. E.; KARDIA, S. L. R.; GIORDANO, T. J.; IANNETTONI, M. D.; ORRINGER, M. B.; HANASH, S. M.; BEER, D. G. Discordant Protein and mRNA Expression in Lung Adenocarcinomas **Molecular & Cellular Proteomics**, Bethesda, p.304-314, 2002.

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.

CHRISTOPHER, M.E.; MIRANDA, M.; MAJOR, I.T.; CONSTABEL, C.P. Gene expression profiling of systemically wound-induced defenses in hybrid poplar. **Planta**, Berlin, v. 219, p. 936-947, 2004.

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.

CONNELLY, C.; HIETER, P. Budding Yeast *SKP1* Encodes an Evolutionarily Conserved Kinetochoe Protein Required for Cell Cycle Progression. **Cell**, Cambridge, v. 86, n. 2, p. 275-285, 1996.

COOKE, R.; RAYNAL, M.; LAUDIE, M. Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 non-redundant ESTs. **The Plant Journal**, Oxford, v. 9, p. 101-124, 1996.

COSGROVE, D.J. Water uptake by growing cells: an assessment of the controlling roles of wall relaxation, solute uptake, and hydraulic conductance. **International journal of plant sciences**, Chicago, v. 154, p. 10-21, 1993.

COSGROVE, D.J. Expansive growth of plant cell walls. **Plant physiology and biochemistry**, Paris, v. 38, p. 129-134, 2000.

COSGROVE, D.J. Growth of the plant cell wall. **Nature reviews. Molecular cell biology**, London, v. 6, p. 850-861, 2005.

COSTA, P.; PIONNEAU, C.; BAUW, G.; DUBOS, C.; BAHRMAN, N.; KREMER, A.; FRIGERIO, J. M.; PLOMION, C. Separation and characterization of needles and xylem maritime pine proteins. **Electrophoresis**, Weinheim, v. 20, p. 1098-1108, 1999.

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.

CRESSWELL, R.D.; BOURLAY, M.; FLANCLET, A. Vegetative propagation of *Eucalyptus*. In: BONGA, J.M.; DURZAN, D.J. (Ed.). **Tissue Culture in Forestry**, Dordrecht: Martinus Nijhoff, 1985. cap. 6, p.45-72.

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

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.

DEJARDIN, A.; LEPLE, J.C.; LESAGE-DESCAUSES, M.C.; COSTA, G.; PILATE, G. Expressed sequence tags from *Populus* spp. Wood tissues – a comparative analysis from multiple libraries. **Plant biology (Stuttgart, Germany)**, Stuttgart, v. 6, p. 55-64, 2004.

DELILLE, J.M.; SEHNKE, P.C.; FERL, R.J. The *Arabidopsis* 14-3-3 family of signaling regulators. **Plant physiology**, Lancaster, v. 126, p. 35-38, 2001.

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.99, p. 15794-15799, 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.

DONSON, J.; FANG, Y.; ESPIRITU-SANTO, G.; XING, W.; SALAZAR, A.; MIYAMOTO, S.; ARMENDAREZ, V.; VOLKMUTH, W. Comprehensive gene expression analysis by transcript profiling. **Plant Molecular Biology**, Dordrecht, v. 48, p. 75-97, 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.

ELDRIDGE, K.; DAVIDSON, J.; HARWOOD, C., VAN WYK, G. **Domestication and breeding**. Oxford: Clarendon Press, 1993. 308p.

ELDRIDGE, K.G.; DAVIDSON, J.; HARWOOD, C.; VAN WYK, G. **Eucalypt domestication and breeding**. Oxford: Oxford Science Publications, 1994. 288 p.

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.

FAO. **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.

FENN, J.B.; MANN, M.; MENG, C.K.; WONG, S.F.; WHITEHOUSE, C.M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. **Science**, Washington, v. 246, p. 64-71, 1989.

FERL R.J. 14-3-3 protein: regulation of signal-induced events. **Physiologia Plantarum**, Lund, v. 120, p 173-178, 2004.

FERL, R.J. 14-3-3 proteins and signal transduction. **Annual Review Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 47, p. 49-73, 1996.

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

FERREIRA, M. Escolha de espécies de eucalipto. **Circular Técnica IPEF**, Piracicaba, v. 47, p. 1-30, 1979.

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

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.

FINN, R.D.; MISTRY, J.; SCHUSTER-BÖCKLER, B.; GRIFFITHS-JONES, S.; HOLLICH, V.; LASSMANN, T.; MOXON, S.; MARSHALL, M.; KHANNA, A.; DURBIN, R.; EDDY, S.R.; SONNHAMMER, E.L.L.; BATEMAN, A. Pfam: clans, web tools and services. **Nucleic Acids Research**, London, v. 34, p. D247-D251, 2006.

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**, Lancaster, v. 134, p. 67-80, 2004.

FLEISCHMANN, R.D.; ADAMS, M.D.; WHITE, O.; CLAYTON, R.A.; KIRKNESS, E.F.; KERLAVAGE, A.R.; BULT, C.J.; TOMB, J.F.; DOUGHERTY, B.A.; MERRICK, J.M.; ET AL. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. **Science**, Washington, v. 269, n. 5223, p. 496-512, 1995.

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.

FRY, S.C. Primary cell wall metabolism: tracking the careers of the wall polymers in living plant cells. **New Phytologist**, London, v. 161, p. 641-675, 2004.

FUKUDA, H. Xylogenesis: Initiation, progression and cell death. **Annual review of plant physiology and plant molecular biology**, Palo Alto, v. 47, p. 299-325, 1996.

FUKUDA, H. Tracheary element differentiation. **The Plant Cell**, Rockville, v. 9, p. 1147-1156, 1997.

GANG, D.R.; COSTA, M.A.; FUJITA, M.; DINKOVA-KOSTOVA, A.T.; WANG, H.; BURLAT, V.; MARTIN, W.; SARKANEN, S.; DAVIN, L.B.; LEWIS, N.G. Regiochemical control of monolignol radical coupling: a new paradigm for lignin and lignan biosynthesis. **Chemistry and Biology**, London, v. 6, p. 143-151, 1999.

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.

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.M.; LALANNE, C.; LE PROVOST, G.; FERRY-DUMAZET, H.; PAIVA, J.; CHAUMEIL, P.; FRIGERIO, J.M.; BRACH, J.; BARRE, A.; 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.

GION, J.M.; RECH, P.; PETTENATI, J.G.; VERHAEGEN, D.; PLOMION, C. Mapping candidates genes in *Eucalyptus* with emphasis on lignification genes. **Molecular breeding: new strategies in plant improvement**, Dordrecht, v. 6, p.441-449, 2000.

GOLD – Genomes on Line Database v 2.0. Disponível em: <<http://www.genomesonline.org/>>. Acesso em: 21 mar. 2007.

GONZÁLEZ, E.R. **Transformação genética de *Eucalyptus grandis* e do híbrido *E.grandis* X *E.urophylla* via *Agrobacterium***. 2002. 93p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2002.

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

GÖRG, A.; WEISS, W.; DUNN, M.J. Current two dimensional electrophoresis technology for proteomics. **Proteomics**, Weinheim, v.4, p.3665-3685, 2004.

GRATTAPAGLIA, D. Mapas genéticos e seleção assistida por marcadores moleculares. In: BORÉM, A (Ed.). **Biotecnologia Florestal**. Viçosa: Universidade Federal de Viçosa, 2007, cap. 9, p. 201-230.

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.

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

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

GRAY-MITSUMUNE, M.; MELLEROWICZ, E.J.; ABE, H.; SCHRADER, J.; WINZELL, A.; STERKY, F.; BLOMQUIST, K.; MCQUEEN-MASON, S.; TEERI, T.T.; SUNDBERG, B. Expansins abundant in secondary xylem belong to subgroup A of the α -expansin gene family. **Plant Physiology**, Lancaster, v. 135, p. 1552–1564, 2004.

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

GROOVER, A.T.; PATTISHALL, A.; JONES, A.M. IAA8 expression during vascular cell differentiation. **Plant Molecular Biology**, Dordrecht, v. 51, n. 3, p. 427-435, 2003.

GYGI, S.P.; ROCHON, Y.; FRANZA, B.R.; AEBERSOLD, R. Correlation between protein and mRNA abundance in yeast. **Molecular cell biology**, London v.19, p.1720–1730, 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.

HARPER, A.D.; BAR-PELED, M. Biosynthesis of UDP-Xylose. Cloning and Characterization of a Novel Arabidopsis Gene Family, *UXS*, Encoding Soluble and Putative Membrane-Bound UDP-Glucuronic Acid Decarboxylase Isoforms. **Plant Physiology**, Lancaster, v. 130, p. 2188-2198, 2002.

HAYGREEN, J.G.; BOWYER, J.L. **Forest products and wood science**: an introduction. 3rd.ed. Ames: Iowa State University Press, 1996. 484 p.

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

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

HERMAN, I.M. Actin isoforms. **Current opinion in cell biology**, Philadelphia, v. 5, p. 48-55, 1993.

HERNÁNDEZ-NIDAL, J.; LABRADOR, E.; MARTÍN, I.; JIMÉNEZ, T.; DOPICO, B. Transcriptional profiling of cell wall protein genes in chickpea embryonic axes during germination and growth. **Plant physiology and biochemistry**, Paris, v. 44, p. 684–692, 2006.

HERTZBERG, M.; ASPEBORG, H.; SCHRADER, J.; ANDERSSON, A.; ERLANDSSON, R.; BLOMQUIST, 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.

HIGUCHI, T. **Biochemistry and Molecular Biology of Wood**. New York: Springer-Verlag, 1997. 362 p.

HO, C.K.; CHANG, S.H.; TSAY, J.Y.; TSAI, C.J.; CHIANG, V.L.; CHEN, Z.Z. *Agrobacterium tumefaciens* mediated transformation of *Eucalyptus calmadulensis* and production of transgenic plants. **Plant Cell Reports**, Berlin, v. 17, n. 9, p.675-680, 1998.

HOAGLAND, D.R.; ARNON, D.I. **The water culture of growing plants without soil**, California: Berkley, 1950. 347p

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. **The Plant Cell**, Rockville, v. 16, p. 1446-1465, 2004.

HU, W.; KAWAOKA, A.; TSAI, C.; LUNG, J.; OSAKABE, K.; EBINUMA, H.; CHIANG, V.L. Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 95, p. 5407-5412, 1998.

HUANG, R.Z.; WU, X.L.; JIANG, C.Z.; HU, Z.H.; HE, H.Y.; LIOU, Z.H.; XU, Y.H.; JIN, W.; ZHENG, T.; CHEN, J.Q. Isolation, characterization and expression of a gene from *Brassica napus* encoding a LIM-domain protein. **Plant Physiology and Biochemistry**, Paris, 2007, *In press*. <doi:10.1016/j.plaphy.2006.10.023>

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.

JACOB-WILK, D.; KUREK, I.; HOGAN, P.; DELMER, D.P. The cotton fiber zinc-binding domain of cellulose synthase A1 from *Gossypium hirsutum* displays rapid turnover in vitro and in vivo. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 103, n. 32, p. 12191–12196, 2006.

JANSEN, R.C.; NAP, J.P. Genetical genomics: the added value from segregation. **Trends in Genetics**, Cambridge, v. 17, p. 388-391, 2001.

JOFUKU, K.D.; DEN BOER, B.G.W.; VAN MONTAGU, M.; OKAMURO, J.K. Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. **The Plant Cell**, Rockville, v. 6, p. 1211–1225, 1994.

JOHANSSON, H.; STERKY, F.; AMINI, B.; LUNDEBERG, J.; KLECZKOWSKI, L.A. Molecular cloning and characterization of a cDNA encoding poplar UDP-glucose dehydrogenase, a key gene of hemicellulose/pectin formation. **Biochimica et Biophysica Acta**, Amsterdam, v. 1576, p. 53-58, 2002.

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.

JOUANIN, L.; GOUJON, T.; DE NADAI, V.; MARTIN, M.T.; MILA, I.; VALLET, C.; POLLET, B.; YOSHINAGA, A.; CHABBERT, B.; PETIT-CONIL, M.; LAPIERRE, C. Lignification in transgenic *Poplars* with extremely reduced caffeic acid O-methyltransferase activity. **Plant Physiology**, Lancaster, v.123, p.1363-1373, 2000.

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, H-J.G.; NI, W. Lignification of plant cell walls: Impact of genetic manipulation. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 95, p. 12742-12743, 1998.

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**, New York, v. 52, p. 553-567, 2003.

KABSCH, W.; VANDEKERCKHOVE, J. Structure and function of actin. **Annual review of biophysics and biomolecular structure**, California, v. 21, p. 49-76, 1992.

KAGEYAMA, P.Y. **Variação genética em progênies de uma população de *Eucalyptus grandis* Hill ex Maiden**. 1980. 71 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1980.

KAJITA, S.; KATAYAMA, Y.; OMORI, S. Alterations in the biosynthesis of lignin in transgenic plants with chimeric genes for 4-Coumarate:Coenzyme A ligase. **Plant & cell physiology**, Kyoto, v. 37, n. 7, p. 957-965, 1996.

KANEHISA, M.; GOTO, S.; HATTORI, M.; AOKI-KINOSHITA, K.F.; ITOH, M.; KAWASHIMA, S.; KATAYAMA, T.; ARAKI, M.; AND HIRAKAWA, M. From genomics to chemical genomics: new developments in KEGG. **Nucleic Acids Research**, London, v. 34, p. D354-357, 2006.

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.

KARAS, M.; BACHMANN, D.; BAHR, U.; HILLENKAMP, F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. **Internacional Journal of Mass Spectrometry and Ion Processes**, Amsterdam, v. 78, p. 53-56, 1987.

KAWAOKA, A.; EBINUMA, H. Transcriptional control of lignin biosynthesis by tobacco LIM protein. **Phytochemistry**, Oxford, v. 57, p. 1149–1157, 2001.

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, 1988.

KIRCH, T.; SIMON, R.; GRÜNEWALD, M.; WERR, W. The *DORNROSCHE/ENHANCER OF SHOOT REGENERATION1* gene of Arabidopsis acts in the control of meristem cell fate and lateral organ development. **The Plant Cell**, Rockville, v. 15, p. 694–705, 2003.

KIRST, M. Genômica florestal: novas abordagens, desafios e perspectivas. In: BORÉM, A (Ed.). **Biotecnologia Florestal**. Viçosa: Universidade Federal de Viçosa, 2007, cap. 10, p. 231-252.

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

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

KO, J.; PRASSINOS, C.; HAN, K. Developmental and seasonal expression of *PtaHB1*, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). **The New Phytologist**, London, v.169, p. 469-478, 2006.

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.

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.

LAFARGUETTE, F.; LEPLÉ, J.C.; DEJARDIN, A.; LAURANS, F.; COSTA, G.; LESAGE-DESCAUSES, M.C.; PILATE, G. Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. **The New Phytologist**, London, v. 164, n. 1, p. 107-121, 2004.

LAPIERRE, C.; POLLET, B.; PETIT-CONIL, M.; TOVAL, G.; ROMERO, J.; PILATE, G.; LEPLÉ, J.C.; BOERJAN, W.; FERRET, V.; DE-NADAI, V.; JOUANIN, L. Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. **Plant Physiology**, Lancaster, v. 119, p. 153-163, 1999.

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.

LAU, D.; RODRÍGUES, M.A.D., LAU, E.Y.; ALFENAS, A.C. Resistência a doenças: aplicações em patossistemas florestais. In: BOREM, A (Ed.). **Biotecnologia Florestal**. Viçosa: Universidade Federal de Viçosa, 2007, cap. 7, p. 143-173.

LAUVERGEAT, V.; LACOMME, C.; LACOMBE, E.; LASSERRE, E.; ROBY, D.; GRIMAPETTENATI, J. Two cinnamoyl-CoA reductase (CCR) genes from *Arabidopsis thaliana* are differentially expressed during development and in response to infection with pathogenic bacteria. **Phytochemistry**, Oxford, v. 57, p. 1187-1195, 2001.

LEE, D.; ELLARD, M.; WANNER, L.A.; DAVIS, K.R.; DOUGLAS, C.J.; The *Arabidopsis thaliana* 4-coumarate:CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA. **Plant Molecular Biology**, Dordrecht, v. 28, n. 5, p. 871-884, 1995.

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, 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.

LEITE, N.B. Avanços da silvicultura brasileira são significativos. **Visão Agrícola**, Piracicaba, n. 5, p.12-18, jul-dez, 2005.

LETUNIC, I.; COPLEY, R.R.; PILS, B.; PINKERT, S.; SCHULTZ, J.; BORK, P. SMART 5: domains in the context of genomes and networks. **Nucleic Acids Research**, London, v. 34, p. D257-D260, 2006.

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.

LOPEZ, M. F. Proteome Analysis I. Gene Products Are Where the Biological Action Is. **Journal of chromatography**, Amsterdam, v. 722, n. 1, p. 191-202, 1999.

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.

LUDEVID, D.; HÖFTE, H.; HIMELBLAU, E.; CHRISPEELS, M.J. The expression pattern of the tonoplast intrinsic protein g-TIP in *Arabidopsis thaliana* is correlated with cell enlargement. **Plant Physiology**, Lancaster, v. 100, p. 1633–1639, 1992.

McCANN, M.C.; ROBERTS, K. Architecture of the primary cell wall. In: LLOYD, C.W. (Eds). **The Cytosqueletal Basis of Plant Growth and Form**, London, 1992, p. 109-129.

MAJEWSKA-SAWKA, A.; NOTHNAGEL, E.A. The multiple roles of arabinogalactan proteins in plant development. **Plant Physiology**, Lancaster, v.122, n. 1, p. 3-9, 2000.

MANN, M.; HENDRICKSON, R.C.A. Analysis of proteins and proteomes by mass spectrometry. **Annual review of biochemistry**, Palo Alto, v. 70, p. 437-473, 2001.

MARCHLER-BAUER, A.; ANDERSON, J.B.; CHERUKURI, P.F.; DEWEESE-SCOTT, C.; GEER, L.Y.; GWADZ, M.; HE, S.; HURWITZ, D.I.; JACKSON, J.D.; KE, Z.; LANCZYCKI, C.J.; LIEBERT, C.A.; LIU, C.; LU, F.; MARCHLER, G.H.; MULLOKANDOV, M.; SHOEMAKER, B.A.; SIMONYAN, V.; SONG, J.S.; THIESSEN, P.A.; YAMASHITA, R.A.; YIN, J.J.; ZHANG, D.; BRYANT, S.H. CDD: a Conserved Domain Database for protein classification. **Nucleic Acids Research**, London, v. 33, p. D192-D196, 2005.

MARCHLER-BAUER, A.; ANDERSON, J.B.; DERBYSHIRE, M.K.; DEWEESE-SCOTT, C.; GONZALES, N.R.; GWADZ, M.; HAO, L.; HE, S.; HURWITZ, D.I.; JACKSON, J.D.; KE, Z.; KRYLOV, D.; LANCZYCKI, C.J.; LIEBERT, C.A.; LIU, C.; LU, F.; LU, S.; MARCHLER, G.H.; MULLOKANDOV, M.; SONG, J.S.; THANKI, N.; YAMASHITA, R.A.; YIN, J.J.; ZHANG, D.; BRYANT, S.H. CDD: a conserved domain database for interactive domain family analysis. **Nucleic Acids Research**, London, v. 35, p. D237-240, 2007.

MATHIEU, M.; LELU-WALTER, M.A.; BLERVACQ, A.S.; DAVID, H.; HAWKINS, S.; NEUTELINGS, G. Germin-like genes are expressed during somatic embryogenesis and early development of conifers. **Plant Molecular Biology**, Dordrecht, v. 61, n. 4-5, p. 615-627, 2006.

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-pathogen 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.

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

MENK, A.L.M.; ODA, S.; VENCOVSKY, R. Problemas no melhoramento genético clássico do eucalipto em função da alta intensidade de seleção. **IPEF**, Piracicaba, n. 41/42, p. 8-17, jan./dez. 1989.

MEYER, G. G.; SCHNARRENBURGER, C.; CERFF, R.; MARTIN, W. Molecular characterization of a novel, nuclear-encoded, NAD(+)-dependent glyceraldehyde-3-phosphate dehydrogenase in plastids of the gymnosperm *Pinus sylvestris*. **Plant Molecular Biology**, Boston, n. 26, v. 4, p. 1155-66, 1994.

MEYERS, B.C.; GALBRAITH, D.W.; NELSON, T.; AGRAWAL, V. Methods for Transcriptional Profiling in Plants. Be Fruitful and Replicate. **Plant Physiology**, Lancaster, v. 135, p. 637-652, 2004.

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

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.; JANSSON, 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.

MULLINS, K.V.; LLEWELLYN, D.J.; HARTNEY, V.J.; STRAUSS, S.; DENNIS, E.S. Regeneration and transformation of *Eucalyptus camaldulensis*. **Plant Cell Reports**, Berlin, v. 16, n. 11, p. 787-791, 1997.

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.

NAIR, R.B.; XIA, Q.; KARTHA, C.J.; KURYLO, E.; HIRJI, R.N.; DATLA, R.; SELVARAJ, G. *Arabidopsis* CYP98A3 mediating aromatic 3-hydroxylation. Developmental regulation of the gene, and expression in yeast. **Plant Physiology**, Lancaster, v. 130, p. 210-220, 2002.

O EUCALIPTO e suas origens. **Revista da Madeira**, Curitiba, n. 59, ano 11, set. 2001a. Disponível em: <<http://www.remade.com.br/revista>>. Acesso em: 13 jan. 2007.

O EUCALIPTO em números. **Revista da Madeira**, Curitiba, n. 59, ano 11, set. 2001b. Disponível em: <<http://www.remade.com.br/revista>>. Acesso em: 13 jan. 2007.

ODA, S.; MELLO, E.J.; SILVA, J.F.; SOUZA, I.C.G. Melhoramento florestal. In: BORÉM, A (Ed.). **Biotechnology Florestal**. Viçosa: Universidade Federal de Viçosa, 2007, cap. 6, p. 123-142.

ODA, Y.; HASEZAWA, S. Cytoskeletal organization during xylem cell differentiation. *Journal of Plant Research*, Tokyo, v. 119, n. 3, p. 167-177, 2006.

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.

PACHECO, I.A. **Polinização de *Eucalyptus saligna* Smith (Myrtaceae) por *Apis mellifera* L., 1758 (Hymenoptera, apidae)**. 1982. 87p. Dissertação (Mestrado em Entomologia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1982.

PANDEY, A.; MANN, M. Proteomics to study genes and genomes. **Nature**, London, v. 405, p. 837-845, 2000.

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.

PASQUALI, G.; ZANETTINI, M.H.B. Transgênese florestal. In: BORÉM, A (Ed.). **Biotecnologia Florestal**. Viçosa: Universidade Federal de Viçosa, 2007, cap. 14, p. 317-334.

PATTERSON, S.D.; AEBERSOLD, R. H. Proteomics: the first decade and beyond. **Nature genetics**, New York, v. 33, p. 311- 323, 2003.

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.

PEREIRA, L.A.R.; SCHOOR, S.; GOUBET, F.; DUPRET, P.; MOFFATT, B.A. Deficiency of adenosine kinase activity affects the degree of pectin methyl-esterification in cell walls of *Arabidopsis thaliana*. **Planta**, Berlin, v. 224, p. 1401-1414, 2006.

PERSPECTIVAS do setor florestal brasileiro. **Revista da Madeira**, Curitiba, n. 75, ano 13, ago. 2003. Disponível em <<http://www.remade.com.br/revista>> Acesso em 13 jan. 2007.

PFANZ, H.; ASCHAN, G.; LANGENFELD-HEYSER, R.; WITTMANN, C.; LOOSE, M. Ecology and ecophysiology of tree steams: cortical and wood photosynthesis. **Die Naturwissenschaften**, Berlin, v. 89, p. 147-162, 2002.

PILATE, G.; DEJARDIN, A.; LAURANS, F.; LEPLE, J.C. Tension wood as a model for functional genomics of wood formation. **The New phytologist**, London, v. 164, p. 63-72, 2004.

PIMENTA, A.M.C. Os desafios do proteoma. **Ciência Hoje**, Rio de Janeiro, v. 32, n. 192, p. 16-22, 2003.

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.

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.

PLOMION, C.; PIONNEAU, C.; BAILLÈRES, H. Analysis of protein expression along the normal to tension wood gradient in *Eucalyptus gunnii*. **Holzforschung**, Berlin, v.57, p.353-358, 2003.

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.

POKE, F.S.; VAILLANCOURT, R.E.; POTTS, B.M.; REID, J.B. Genomic research in *Eucalyptus*. **Genetica**, Dordrecht v. 125, 79-101, 2005.

POT, D.; CHANTRE, G.; ROZENBERG, P.; RODRIGUES, J. C.; JONE, S 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.

PRYOR, L.D.; WILLIAMS, E.R.; GUNN, B.V. A Morphometric Analysis Of *Eucalyptus-Urophylla* And Related Taxa With Descriptions Of 2 New Species. **Australian Systematic Botany**, East Melbourne, v. 8, n. 1, p. 57-70, 1995.

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.

RAMACHANDRAN, S.; CHRISTENSEN, H.E.M.; ISHIMARU, Y.; DONG, C.H.; CHAO-MING, W. Profilin plays a role in cell elongation, cell shape maintenance, and flowering in *Arabidopsis*. **Plant physiology**, Lancaster, v.124, p. 1637–1647, 2000.

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.

REGAN, S.; CHAFFEY, N.J.; SUNDBERG, B. Exploring cambial growth with *Arabidopsis* and *Populus*. **Journal of Experimental Botany**, Oxford, v. 50, n. 33, p. 3-51, 1999.

REID, J.S.G. Cementing the wall: cell wall polysaccharide synthesizing enzymes. **Current Opinion Plant Biology**, London, v. 3, p. 517-522, 2000.

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.

REITER, W.D. Biosynthesis and properties of the plant cell wall. **Current Opinion in Plant Biology**, London, v. 5, p. 536-542, 2002.

RESENDE, M.D.V. de. Melhoramento de essências florestais. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: UFV, 1999, cap. 17, p. 589-647.

RIECHMANN, J.L.; RATCLIFFE, O.J. A genomic perspective on plant transcription factors. **Current Opinion in Plant Biology**, London, v. 290, p. 2105-2110, 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.

RO, D.K.; MAH, N.; ELLIS, B.E.; DOUGLAS, C.J. Functional characterization and subcellular localization of poplar (*Populus trichocarpax**Populus deltoides*) cinnamate 4-hydroxylase. **Plant Physiology**, Lancaster, v. 126, n. 1, p. 317-329, 2001.

ROBERT, S.; BICHET, A.; GRANDJEAN, O.; KIERZKOWSKI, D.; SATIAT-JEUNEMAITRE, B.; PELLETIER, S.; HAUSER, M.T.; HOFTE, H.; VERNHETTES, S. An *Arabidopsis* endo-1,4-β-D-Glucanase involved in cellulose synthesis undergoes regulated intracellular cycling. **The Plant Cell**, Rockville, v. 17, p. 3378-3389, 2005.

ROBERTS, K.; McCANN, M.C. Xylogenesis: the birth of a corpse. **Current Opinion in Plant Biology**, London, v. 3, p. 517-522, 2000.

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.

SALVATIERRA, G.R. **Identificação de genes diferencialmente expressos em árvores de *Eucalyptus grandis* susceptíveis e resistentes à *Puccinia psidii***. 2006. 154p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2006.

SALZMAN, R.A.; FUJITA, T.; ZHU-SOLZMAN, K.; HASEGAWA, P.M.; VERNON, R.A. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. **Plant Molecular Biology Reporter**, New York, v. 17, n. 1, p. 11-17, 1999.

SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Molecular Cloning: a laboratory manual**. New York: Cold Spring Harbor Laboratory Press, 1989. 765p.

SAMPSON, J.F.; HOPPER, S.D.; JAMES, S.H. The mating system and genetic diversity of the Australian Arid Zone Mallee, *Eucalyptus rameliana*. **Australian Journal of Botany**, East Melbourne, v. 43, n. 5, p. 461-474, 1995.

SAMUGA, A.; JOSHI, C. P. A new cellulose synthase gene (*PtrCesA2*) from aspen xylem is orthologous to *Arabidopsis AtCesA7* (*irx3*) gene associated with secondary cell wall synthesis. **Gene**, Amsterdam, v. 296, p. 37-44, 2002.

SAVIDGE, R.; BARNETT, J.; NAPIER, R. (Ed.) **Cell and molecular biology of wood formation**. Oxford: BIOS Scientific Publishers Ltd. 2000. 532 p.

SCHOCH, G.; GOEPFERT, S.; MORANT, M.; HEHN, A.; MEYER, D.; ULLMANN, P.; WERCK-REICHART, 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.

SCHULTZ, J.; MILPETZ, F.; BORK, P.; PONTING, C.P. SMART, a simple modular architecture research tool: Identification of signaling domains. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 95, p. 5857-5864, 1998.

SEITZ, B.; KLOS, C.; WURM, M.; TENHAKEN, R. Matrix polysaccharide precursors in *Arabidopsis* cell walls are synthesized by alternate pathways with organ-specific expression patterns. **The Plant Journal**, Oxford, v. 21, n. 6, p. 537-546, 2000.

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, v. 55, p. 555-90, 2004.

SOCIEDADE BRASILEIRA DE SILVICULTURA. **Fatos e números do setor florestal brasileiro**. São Paulo, 2006. 109 p.

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.

STAFSTROM, J.P.; RIPLEY, B.D.; DEVITT, M.L.; DRAKE, B. Dormancy-associated gene expression in pea axillary buds. Cloning and expression of PsDRM1 and PsDRM2. **Planta**, Berlin, v. 205, p. 547-552, 1998.

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.

STEIN, J.; LIANG, P. Differential display technology: a general guide. **Cellular and Molecular Life Sciences: CMLS**, Basel, v. 59, p. 1235-1240, 2002.

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.

STRASBURGER, E. **Tratado de Botânica**. 8° ed. Barcelona: Española Ediciones Omega, 1994, 1098p.

SUGIURA, M; TAKEDA, Y. Nucleic acids. In B BUCHANAN, B.; GRUISSEM, W.; JONES, R. (Eds). **Biochemistry and Molecular Biology of Plants**, Rockville : American Society for Plant Physiologists, 2000. cap.3, p. 298-300.

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.

TAIZ, L.; ZEIGER, E. Paredes celulares: estrutura, biogênese e expansão. In: _____. **Fisiologia Vegetal**. Porto Alegre: Artmed, 2004. cap. 15, p.339-364.

TATUSOV, R.L.; FEDOROVA, N.D.; JACKSON, J.D.; JACOBS, A.R.; KIRYUTIN, B.; KOONIN, E.V.; KRYLOV, D.M.; MAZUMDER, R.; MEKHEDOV, S.R.; NIKOLSKAYA, A.N.; RAO, B.R.; SMIRNOV, S.; SVERDLOV, A.V.; VASUDEVAN, S. WOLF, Y.I.; YIN, J.J.; NATALE, D.A. The COG database: an updated version includes eukaryotes. **BMC Bioinformatics**, London, v. 4, p. 1-14, 2003.

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

TRUGILHO, P.F.; LIMA, J.T.; MORI, F.A.; LINO, A.L. Avaliação de clones de *Eucalyptus* para a produção de carvão vegetal. **Cerne**, Lavras, v. 7, n. 2, p. 104-114, 2001.

TUSKAN, A.; DIFAZIO, S.; JANSSEN, S.; BOHLMANN, J.; GRIGORIEV, I.; HELLSTEN, U.; PUTNAM, N.; RALPH, S.; ROMBAUTS, S.; SALAMOV, A.; SCHEIN, J.; STERCK, L.; AERTS, A.; BHALLERAO, R.R.; BHALLERAO, 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.; RYABOV, 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.

VALÉRIO, L.; CARTER, D.; RODRIGUES, J.C.; TOURNIER, V.; GOMINHO, J.; MARQUE, C.; BOUDET, A.M.; MAUNDERS, M.; PEREIRA, H.; TEULIÈRES, C. Down-regulation of cinnamyl Alcohol Dehydrogenase, a lignification enzyme, in *Eucalyptus camaldulensis*. **Molecular Breeding**, Dordrecht, v.12, n.2, p.157-167, 2003.

VALVERDE, R.S.; NEIVA, S.A.; SOARES, N.S. Biotecnologia e competitividade das plantações florestais. In: BORÉM, A (Ed.). **Biotecnologia Florestal**. Viçosa: Universidade Federal de Viçosa, 2007, cap. 17, p. 363-374.

van RAEMDONCK, D.; PESQUET, E.; CLOQUET, S.; BEECKMAN, H.; BOERJAN, W.; GOFFNER, D.; EL JAZIRI, M.; BAUCHER, M. Molecular changes associated with the setting up of secondary growth in aspen. **Journal of Experimental Botany**, Oxford, v. 56, n. 418, p. 2211-2227, 2005.

van WIJK, K.L. Challenges and prospects of plant proteomics. **Plant Physiology**, Lancaster, v. 116, p. 501-508, 2001.

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.

VERNOUD, V.; HORTON, A.C.; YANG, Z.; NIELSEN, E. Analysis of the small GTPase gene superfamily of Arabidopsis. **Plant Physiology**, Lancaster, v. 131, p. 1191-1208, 2003.

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.

WATSON, B.S.; ASIRVATHAM, V.S.; WANG, L.; SUMMER, L.W. Mapping the proteome of Barrel Medic (*Medicago truncatula*). **Plant Physiology**, Lancaster, v.131, p. 1104-1123, 2003.

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.

WHETTEN, R.; SEDEROFF, R. Lignin Biosynthesis. **The Plant Cell**, Rockville, v. 7, p. 1001-1013, Jul. 1995.

WHETTEN, R.W.; MACKAY, J.J.; SEDEROFF, R.R. Recent advances in understanding lignin biosynthesis. **Annual review of plant physiology and plant molecular biology**, Palo Alto, v. 48, p. 407-416, 1998.

YAMAZAKI, M.; SAITO, K. Differential display analysis of gene expression in plants. **Cellular and Molecular Life Sciences: CMLS**, Basel, v. 59, p. 1246-1255, 2002.

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.

YI, S.Y.; KIM, J.-H.; JOUNG, Y.-H.; LEE, S.; KIM, W.-T.; YU, S.H.; CHOI, D. The pepper transcription factor CaPF1 confers pathogen and freezing tolerance in *Arabidopsis*. **Plant Physiology**, Lancaster, v. 136, p. 2862–2874, 2004.

ZHANG, D.; HRMOVA, M.; WAN, C.; WU, C.; BALZEN, J.; CAI, W.; WANG, J.; DENSMORE, L.D.; FINCHER, G.B.; ZHANG, H.; HAIGLER, C.H. Members of a new group of chitinase like genes are expressed preferentially in cotton cells with secondary walls. **Plant Molecular Biology**, Dordrecht, v. 54, n. 3, p. 353–372, 2004.

ZHANG, L.; ZHOU, W.; VELCULESCU, V.E.; KERN, S.E.; HRUBAN, R.H.; HAMILTON, S.R.; VOGELSTEIN, B.; KINZLER, K.W. Gene expression profiles in normal and cancer cells. **Science**, Washington, v. 276, p. 1268–1272, 1997.

ZHANG, Y.; SEDEROFF, R.; ALLONA, I. Differential expression of genes encoding cell wall proteins in vascular tissues from vertical and bent loblolly pine trees. **Tree Physiology**, Victoria, v. 20, p. 457–466, 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.

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, 2000b.

ZHONG, R.; RIPPERGER, A.; YE, Z. Ectopic deposition of lignin in the pith of stems of two *Arabidopsis* mutants. **Plant Physiology**, Lancaster, v. 123, p. 59–69, 2000a.

ZOBEL, B.J.; JETT, J.B. **Genetics of wood production**: Springer series in wood science. Berlin: Springer, 1995, 352p.

ANEXOS

ANEXO A - Tampão de Extração de RNA (SALZMAN et al., 1999)

<i>Componentes</i>	<i>Concentração Final</i>
Tiocianato de Guanidina	4 M
Tris-HCl pH 8,0	100 mM
Citrato de Sódio pH 8,0	25 mM
N-lauril sarcosina	0,5%
*PVP solúvel	1%
*2-mercaptoetanol	0,2%
H2O DEPC 0,1%	q.s.p. volume final

(* Adicionado na hora do uso)

ANEXO B - Tampão de corrida de RNA (Sambrook et al., 1989)

<i>Componentes</i>	<i>Quantidade</i>
MOPS 10x (Anexo U)	1,5 µL
Formaldeído 37%	3,0 µL
Formamida deionizada	7,5 µL
Tampão Dye III	2,0 µL
Brometo de etídeo	1,0 µL
RNA	5,0 µg

ANEXO C - TAE 1X (Sambrook et al., 1989)

<i>Componentes</i>	<i>Quantidade</i>
Tris	4,84 g
Ácido acético glacial	1,142 mL
EDTA 0,5 M (pH 8,0)	2 mL
Água	q.s.p. 1 L

ANEXO D - Tampão de ligação (*Binding Buffer*) (conforme protocolo *Dynabeads® mRNA Purification kit* – Dynal)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl (pH 7,5)	20 mM
LiCl	1 M
EDTA	2 mM
Água MilliQ tratada com DEPC	q.s.p. volume final
Esterilização por filtração (Millipore 0,2 µm)	

ANEXO E - Tampão de lavagem B (*Washing Buffer B*) (conforme protocolo *Dynabeads® mRNA Purification kit* – Dynal)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl (pH 7,5)	10 mM
LiCl	0,15 M
EDTA	1 mM
Água MilliQ tratada com DEPC	q.s.p. volume final
Esterilização por filtração (Millipore 0,2 µm)	

ANEXO F - Tampão TBE 0,5X (Sambrook et al., 1989)

<i>Componentes</i>	<i>Quantidade</i>
Tris	5,4 g
Ácido bórico	2,75 g
EDTA pH 8.0	2 mL
Água MilliQ	q.s.p. 1 L

ANEXO G - Tampão da amostra *Dye IV* (Sambrook et al., 1989)

<i>Componentes</i>	<i>Concentração final</i>
Sacarose	40%
Azul de bromofenol	0,25%
Água MilliQ autoclavada	q.s.p. volume final

ANEXO H - Tampão de ligação (*Lysis/Binding Buffer*) (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,5	100 mM
LiCl	500 mM
EDTA	10 mM
Duodecil sulfato de lítio	1%
DTT (Ditiotreitol)	5 mM

ANEXO I - Tampão de lavagem A (*Washing Buffer A*) (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,5	10 mM
LiCl	0,15 M
EDTA	1 mM
Duodecil sulfato de lítio	0,1%
Glicogênio	10 µg mL ⁻¹

ANEXO J - Tampão de lavagem B (*Washing Buffer B*) (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,5	10 mM
LiCl	150 mM
EDTA	1 mM
Glicogênio	10 µg mL ⁻¹

ANEXO K - Tampão de síntese da primeira fita (*1X First Strand Buffer*) (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
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 ⁻¹

ANEXO L - Reação de síntese da primeira fita de cDNA (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
5X First Strand buffer	18 µL
RnaseOUT™	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 M - Reação da Transcriptase reversa (*RT SuperScript™ II*) (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen), 200 U μL^{-1} em:

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

ANEXO N - Tampão de síntese da segunda fita (*5X Second Strand Buffer*) (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

Componentes	Concentração final
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 O - *dNTP Mix* (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

Componentes	Concentração final
dATP	10 mM
dGTP	10 mM
dCTP	10 mM
dTTP	10 mM
Em água MilliQ tratada com DEPC	

ANEXO P - Enzima *E. coli DNA ligase* (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen), 10 U μL^{-1} em:

Componentes	Concentração final
Tris-HCl pH 7,4	10 mM
KCl	50 mM
EDTA	0,1 mM
DTT	1 mM
BSA	200 $\mu\text{g mL}^{-1}$
Glicerol (p/v)	50%
Triton® X-100	0,1%

ANEXO Q - Enzima *E. coli* DNA polimerase (conforme protocolo do kit I-SAGE™ – versão E da Invitrogen), 10 U μL^{-1} em:

<i>Componentes</i>	<i>Concentração final</i>
Fostato de potássio pH 7,4	50 mM
KCl	100 mM
DTT	1 mM
Glicerol (p/v)	50%

ANEXO R - Enzima *E. coli* RNase H (conforme protocolo do kit I-SAGE™ – versão E da Invitrogen), 2 U μL^{-1} em:

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,5	20 mM
KCl	100 mM
EDTA	0,1 mM
DTT	0,1 mM
BSA	50 $\mu\text{g mL}^{-1}$
Glicerol (p/v)	50%
MgCl ₂	10 mM

ANEXO S - Tampão de lavagem C (*Washing Buffer C*) (conforme protocolo do kit I-SAGE™ – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,5	50 mM
NaCl	1 M
EDTA	0,5 mM
SDS (Duodecil Sulfato de Sódio)	1%
Mussel glycogen	10 $\mu\text{g mL}^{-1}$

ANEXO T - Tampão de lavagem D (*Washing Buffer D*) (conforme protocolo do kit I-SAGE™ – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,5	5 mM
NaCl	1 M
EDTA	0,5 mM
BSA	200 $\mu\text{g mL}^{-1}$

ANEXO U - Tampão 4 (*1X Buffer 4*) (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris-acetato pH 7,9	20 mM
Acetato de magnésio	10 mM
Acetato de potássio	50 mM
DTT	1 mM

ANEXO V - Tampão LoTE (*LoTE*) (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,5	3 mM
EDTA pH 7,5	0,2 mM

ANEXO X - *100X BSA* (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
BSA	10 mg mL ⁻¹
Fosfato de potássio pH 7	20 mM
NaCl	50 mM
EDTA	0,1 mM
Glicerol (p/v)	5%

ANEXO Z - Enzima *Nla III* (conforme protocolo do *kit I-SAGE™* – versão E da Invitrogen), 10 U μL^{-1} em:

<i>Componentes</i>	<i>Concentração final</i>
KCl	200 mM
Tris-HCl pH 7,5	10 mM
EDTA	0,1 mM
DTT	1 mM
100X BSA	500 $\mu\text{g mL}^{-1}$
Glicerol (p/v)	50%

ANEXO AA - Tampão ligase (*1X Ligase buffer*) (conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,5	6 mM
NaCl	5 mM
MgCl ₂	6 mM
BSA	0,1 mg mL ⁻¹
β - mercaptoetanol	7 mM
ATP	0,1 mM
DTT	2 mM
Espermidina	1 mM

ANEXO AB - Enzima *T4 DNA ligase* (conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen), 4 U μL⁻¹ em:

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,5	10 mM
Cloreto de potássio	50 mM
DTT	1 mM
Glicerol (p/v)	50%

ANEXO AC - Enzima *BsmF I* (conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen), 2 U μL⁻¹ em:

<i>Componentes</i>	<i>Concentração final</i>
NaCl	50 mM
Tris-HCl pH 7,4	10 mM
EDTA	0,1 mM
DTT	1 mM
BSA	200 μg mL ⁻¹
Glicerol (p/v)	50%

ANEXO AD - Reação de PCR teste – Verificação da digestão com *BsmF* I (conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen)

Componentes	Tubos			
	1	2	3	4
10X BV buffer	5 µL	5 µL	5 µL	5 µL
DMSO	3 µL	3 µL	3 µL	3 µL
dNTP mix	5 µL	5 µL	5 µL	5 µL
DTP-1 ou DTP-2	2 µL	2 µL	2 µL	2 µL
UDGH anti-sense primer	2 µL	2 µL	2 µL	2 µL
Platinum®Taq DNA polimerase	0,5 µL	0,5 µL	0,5 µL	0,5 µL
Água MilliQ tratada com 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 AE - Tampão Klenow (*10X Klenow buffer*) (conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen)

Componentes	Concentração final
Tris-HCl pH 8,0	500 mM
NaCl	500 mM
MgCl ₂	100 mM

ANEXO AF - Enzima *Klenow polimerase*(conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen), 3-9 U µL⁻¹ em:

Componentes	Concentração final
Fosfato de potássio pH 7,0	50 mM
KCl	100 mM
DTT	1 mM
Glicerol (p/v)	50%

ANEXO AG - Tampão BV 10X concentrado (*10X BV Buffer*) (conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen)

Componentes	Concentração final
Sulfato de amônio	166 mM
Tris-HCl pH 8,8	670 mM
MgCl ₂	67 mM
β - mercaptoetanol	100 mM

ANEXO AH - Gel de poliacrilamida 12%

<i>Componentes</i>	<i>Quantidade</i>
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 AI - Tampão de amostra 6X TBE (6X TBE Sample Buffer) (conforme protocolo do kit I-SAGE™ – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris base	18 mM
Ácido Bórico	18 mM
EDTA (livre de ácidos)	0,4 mM
Ficoll™ tipo 400	3%
Azul de bromofenol	0,02%
Xileno cianol	0,02%

ANEXO AJ - Vetor pZErO® - 1 (conforme protocolo do kit I-SAGE™ – versão E da Invitrogen)

<i>Componente</i>	<i>Concentração final</i>
pZErO® - 1	1 µg µL ⁻¹ em tampão TE pH 7,5

ANEXO AK - Tampão 2 10X concentrado (10X Buffer 2) (conforme protocolo do kit I-SAGE™ – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Tris-HCl pH 7,9	100 mM
MgCl ₂	10 mM
NaCl	50 mM
DTT	1 mM

ANEXO AL - Enzima *Sph* I (conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen), 5 U μL^{-1} em:

<i>Componentes</i>	<i>Concentração final</i>
NaCl	100 mM
Tris-HCl pH 7,4	10 mM
EDTA	0,1 mM
DTT	1 mM
BSA	100 $\mu\text{g/mL}$
Glicerol (p/v)	50%

ANEXO AM - Meio SOC (conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
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 AN - Meio LB sólido (conforme protocolo do *kit* I-SAGE™ – versão E da Invitrogen)

<i>Componentes</i>	<i>Concentração final</i>
Triptona	1%
Extrato de levedura	0,5%
NaCl	0,5%
Agar	1,5%
pH ajustado para 7,5 e esterilização por autoclavagem	