

Universidade Federal do Rio Grande do Sul
Centro de Biotecnologia
Programa de Pós-graduação em Biologia Celular e Molecular

**AVALIAÇÃO DA EXPRESSÃO GÊNICA DIFERENCIAL ENTRE FOLHAS E
TECIDOS VASCULARES DE *EUCALYPTUS GRANDIS***

MARINA TAGLIARO JAHNS

**Dissertação submetida ao Programa
de Pós- Graduação em Biologia
Celular e Molecular do Centro de
Biotecnologia da UFRGS como
requisito parcial para obtenção do
grau de Mestre.**

Orientador: Giancarlo Pasquali
Co-orientador: Tarso Benigno Ledur Kist

Porto Alegre
Março de 2008

INSTITUIÇÃO E FONTES FINANCIADORAS

Este trabalho foi desenvolvido no Laboratório de Biologia Molecular Vegetal do Centro Biotecnologia da Universidade Federal do Rio Grande do Sul (UFRGS).

Fontes financiadoras deste trabalho: financiamento a projeto de pesquisa - Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/MCT); bolsa de mestrado - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Projeto GENOLYPTUS.

Financiamento do Projeto GENOLYPTUS: Fundo Setorial Verde-Amarelo do Ministério da Ciência e Tecnologia (MCT) e as empresas de celulose e papel (1) Aracruz Celulose S.A., (2) Grupo Raiz, (3) Ferro-Gusa Carajás S.A., (4) Celulose Nipo-Brasileira S.A. – CENIBRA, (5) International Paper do Brasil Ltda., (6) Jarí Celulose S.A., (7) Klabin S.A., (8) Lwarcel Celulose e Papel LTDA., (9) Veracel Celulose S.A., (10) Votorantim Celulose e Papel S.A., (11) Zanini Florestal S.A., (12) Suzano-Bahia Sul Papel e Celulose S.A. e (13) Vallourec & Mannesmann do Brasil S.A.

“Não basta saber, é preferível saber aplicar. Não é o bastante querer, é preciso saber querer.”

Johann Wolfgang von Goethe

AGRADECIMENTO

A todas as pessoas que participaram da minha vida e tornaram esse trabalho possível.

Sumário

<u>RESUMO.....</u>	<u>6</u>
<u>ABSTRACT.....</u>	<u>8</u>
<u>CAPÍTULO I</u>	<u>10</u>
<u>1-INTRODUÇÃO E OBJETIVOS.....</u>	<u>10</u>
<u>2-REVISÃO BIBLIOGRÁFICA.....</u>	<u>14</u>
<u>2.1. A Indústria de Celulose e Papel & o Eucalipto.....</u>	<u>14</u>
<u>2.2. Eucalyptus grandis.....</u>	<u>16</u>
<u>2.3. A Gênese da Madeira.....</u>	<u>17</u>
<u>2.4. Genômica Funcional.....</u>	<u>20</u>
<u>2.5. Genes Xilema-Específicos e a Engenharia Genética.....</u>	<u>22</u>
<u>CAPÍTULO II.....</u>	<u>28</u>
<u>Differential expression in Eucalyptus grandis.....</u>	<u>28</u>
<u>Abstract.....</u>	<u>30</u>
<u>Background.....</u>	<u>32</u>
<u>Results and Discussion</u>	<u>35</u>
<u>Conclusion.....</u>	<u>55</u>
<u>Methods.....</u>	<u>55</u>
<u>References.....</u>	<u>59</u>
<u>CAPÍTULO III.....</u>	<u>81</u>
<u>O PROJETO GENOLYPTUS E OS EXPERIMENTOS DE MICROARRANJOS DE DNA: CONSIDERAÇÕES EXTRAS, RESULTADOS E DISCUSSÃO.....</u>	<u>81</u>
<u>Da coleta de amostras ao envio dos RNAs para os ensaios de microarranjo de DNA</u>	<u>81</u>
<u>Do experimento de microarranjo ao processamento dos dados brutos</u>	<u>84</u>
<u>Das análises preliminares aos resultados finais.....</u>	<u>85</u>
<u>CONSIDERAÇÕES FINAIS.....</u>	<u>99</u>
<u>REFERÊNCIAS BIBLIOGRÁFICAS (Capítulos I e III).....</u>	<u>99</u>

RESUMO

No presente trabalho está descrita a análise de experimentos de microarranjos de DNA realizados pelos pesquisadores do Projeto GENOLYPTUS em conjunto com a empresa NimbleGen Systems Inc. (Reykjavik, Iceland), assim como a validação dos resultados gerados por essa análise, com a finalidade de encontrar genes diferencialmente expressos entre folhas e xilema de *E. grandis* para futuro estudo.

O processo de análise dos microarranjos de DNA envolveu a busca por um *software* com programas estatísticos adequados ao estudo de grande quantidade de dados com mínima probabilidade de erro. A comprovação dos resultados gerados pelo *software* escolhido foi realizada com o uso da técnica da reação em cadeia da DNA-polimerase quantitativa (em tempo real) precedida de transcrição reversa (qRT-PCR). Alguns dos genes mais diferencialmente expressos foram selecionados para esta etapa, juntamente com alguns genes com expressão não tão alta. Os genes mais expressos nas folhas que foram escolhidos codificam para; (i) uma proteína semelhante à metalotioneína do tipo 3 (MT-3); (ii) uma glicolato-oxidase; (iii) uma catalase; (iv) uma fosforribulocinase precursora de cloroplasto (PRK); (v) um fator de transcrição MYB do tipo MYB142 e, (vi) uma proteína tipo “dedo-de-zinco”. Os genes mais expressos no xilema escolhidos foram os que codificam (i e ii) duas proteínas expressas em *Arabidopsis thaliana*; (iii) uma proteína hipotética expressa em *Nicotiana benthamiana*; (iv) uma descarboxilase de UDP-glicuronato do tipo 2, (v) uma quitinase do tipo ELP; (vi) uma celulose-sintase tipo 3; (vii) um fator de transcrição MYB; (viii) uma cafeoil-CoA-3-O-metiltransferase (CCoAOMT) e; (ix) uma proteína rica em prolina híbrida do tipo 2 (HyPRP2).

Os resultados das qRT-PCRs demonstraram que o experimento de microarranjo e seus resultados foram consistentes e válidos, embora os primeiros tenham demonstrado valores de expressão freqüentemente mais altos. Acreditamos que tanto a análise dos microarranjos de DNA quanto a equivalência das amostras (duplicatas) biológicas estudadas foram totalmente sustentadas pelos resultados da qRT-PCR, pois foram robustas o suficiente para estimar um grande número de genes simultaneamente e indicar aqueles mais importantes como candidatos para futuras análises.

ABSTRACT

In the present work it is described the analysis of DNA microarray experiments developed by GENOLYPTUS researchers together with Nimblegen Systems Inc. (Reykjavik, Iceland), as well as the validation of the generated results with the aim to find differentially expressed genes between leaves and xylem tissue of *E. grandis* for further study.

The DNA microarray analysis process comprised the search for software containing statistical programs satisfactory for studying an enormous amount of data with a very low false discovery rate. The confirmation of the generated data by the chosen software was made with the use of quantitative (real-time) reverse-transcription followed by polymerase chain reaction (qRT-PCR). Some of the most differentially expressed genes were selected for this step, along with some genes with average expression. Leaf chosen genes were those that codify (i) a metallothionein-like protein type 3 (MT-3), (ii) a glycolate oxidase, (iii) a catalase, (iv) a precursor phosphoribulokinase (PRK) from chloroplast, (v) a MYB transcription factor (MYB142), and (vi) a CONSTANS-LIKE 16 zinc-finger protein. Xylem chosen genes were those codifying (i and ii) two expressed proteins from *Arabidopsis thaliana*, (iii) a hypothetical protein expressed in *Nicotiana benthamiana*, (iv) a putative UDP-glucuronate decarboxylase type 2, (v) a chitinase type ELP (ectopic deposition of lignin in pith), (vi) a cellulose synthase type 3, (vii) a MYB transcription factor, (viii) a caffeoyl-CoA 3-O-methyltransferase (CCoAOMT), and (ix) a hybrid proline-rich protein type 2 (HyPRP2).

Our qRT-PCR results proved the full consistency and validation of the microarray experiments, although the relative expression ratios of the first were often higher. We believe that our microarray analysis as well as the equivalence of biological samples (duplicates) were fully supported by

the qRT-PCR findings since they were robust enough to evaluate a massive number of genes and able to point out important candidate genes.

CAPÍTULO I

1- INTRODUÇÃO E OBJETIVOS

O Brasil ocupa o primeiro lugar em área plantada para exploração comercial de espécies arbóreas do gênero *Eucalyptus* no mundo. Em nosso país, o principal uso da madeira do eucalipto está na indústria de celulose e papel e, em segundo lugar, na produção de carvão vegetal para a indústria siderúrgica. O eucalipto também fornece matéria-prima para a produção de aglomerados, chapas, fibras, móveis, mel e óleo para a obtenção de compostos aromáticos, além de proteger o solo contra erosão, captar dióxido de carbono, gerar energia a partir da lenha e carvão e gerar taninos para o curtimento do couro (Sociedade Brasileira de Silvicultura, 2006). As empresas brasileiras de celulose e papel têm o eucalipto como fonte de matéria-prima e possuem uma das maiores produtividades florestais do mundo, suprimindo, de forma eficiente, a demanda de matéria-prima com o fornecimento de produtos de qualidade internacional.

As empresas e os cientistas brasileiros ocupam posição de destaque internacional no que se refere à pesquisa e ao desenvolvimento em genética, melhoramento, marcadores moleculares e clonagem de *Eucalyptus*. É, portanto, natural que o Brasil viesse a sediar um projeto de grande envergadura para o estudo do genoma de *Eucalyptus*. No ano de 2002, a partir da reunião de 12 grandes empresas do setor de papel e celulose e dez instituições de pesquisa brasileiras, foi criado o projeto denominado “GENOLYPTUS – Rede Brasileira de Pesquisa do Genoma de *Eucalyptus*” no âmbito do Ministério da Ciência e Tecnologia, com recursos do Fundo Setorial Verde-Amarelo e das próprias empresas participantes.

O objetivo central do Projeto GENOLYPTUS inclui o descobrimento, o seqüenciamento, o mapeamento e a determinação da função de genes de importância econômica de espécies de *Eucalyptus*, visando a incorporação de tecnologias de genética genômica nos programas de melhoramento e produção florestal em busca de plantas de *Eucalyptus* mais produtivas, mais resistentes a moléstias e pragas e mais tolerantes a estresses abióticos. O projeto GENOLYPTUS foi organizado em oito subprojetos que englobam desde a instalação de uma rede experimental a campo e implementação de tecnologias para a avaliação da qualidade da madeira, passando pela construção de mapas genéticos e físicos para o mapeamento de locos controladores de características quantitativas (do inglês, *quantitative trait loci* ou QTLs), além da investigação da base genética e identificação de genes que conferem resistência a doenças em *Eucalyptus*.

O subprojeto intitulado "Seqüenciamento do Transcriptoma de *Eucalyptus*" teve como metas a construção de bibliotecas de expressão de *Eucalyptus*, a obtenção de pelo menos 30.000 seqüências gênicas únicas ou de 150.000 ESTs válidas, e a organização e manutenção dos estoques de clones de cDNA. Como conclusão, 13 bibliotecas de expressão foram construídas: (1) xilema de *E. grandis*; (2) xilema de *E. globulus*; (3) xilema de *E. pellita*; (4) xilema de *E. urophylla*; (5) mistura de xilemas de nove espécies e híbridos; (6) mistura de floemas de nove espécies e híbridos; (7) folhas maduras de *E. grandis*; (8) folhas jovens de *E. grandis*; (9) folhas jovens de *E. grandis* infectadas a campo pelo fungo *Puccinia*; (10) plântulas *in vitro* de *E. grandis* e (11) mistura de plântulas de *E. grandis* submetidas a 20 diferentes tratamentos; (12) raízes de diferentes espécies e híbridos de *Eucalyptus*; e (13) flores de *E. grandis*. Um total de 96.493 ESTs válidas foram depositadas nas centrais de bioinformática do Projeto GENOLYPTUS, incluindo 21.905 seqüências únicas.

Dentro do projeto GENOLYPTUS também se enquadra o sub-projeto intitulado: “Gênese da Madeira em *Eucalyptus*: Genes, Funções, Regulação e Expressão Transgênica”, no qual esta Dissertação insere-se. Trata-se de uma avaliação experimental dos clones de cDNA produzidos e seqüenciados com o objetivo de transformá-los em conhecimento científico e, também, produtos biotecnológicos para aplicação imediata no melhoramento genético de *Eucalyptus* e outros vegetais por meio da transgênese. Este estudo está direcionado, principalmente, aos genes cujas expressões afetem, direta ou indiretamente, a formação da madeira. Para esta finalidade, foram planejados e analisados experimentos de microarranjos de DNA.

Recentemente, foi aprovado o projeto para o seqüenciamento na íntegra do genoma do eucalipto junto ao *Joint Genome Institute* (JGI), ligado ao Departamento de Energia do governo dos EUA (USDOE). O projeto foi organizado por pesquisadores dos EUA, Brasil e África do Sul, e conta com participantes da Austrália, Portugal, França e Suécia. Todo o trabalho computacional de anotação e de montagem do genoma será realizado nos EUA e os dados serão públicos. A espécie selecionada foi *E. grandis*, chamado de eucalipto tropical, por ser a espécie que melhor adapta-se ao clima e solo brasileiros e a mais cultivada no mundo. Esta conquista é de extrema importância para o Brasil e para o aperfeiçoamento das tecnologias de genômica para os programas de melhoramento silviculturais.

Nesse contexto, o desenvolvimento do presente trabalho teve como principal objetivo identificar genes diferencialmente expressos em tecido vascular e folhas de *E. grandis* a partir da análise de microarranjos de DNA realizados em conjunto com a empresa NimbleGen Systems Inc. (Reykjavik, Iceland) para 21.442 ESTs do Projeto GENOLYPTUS. Além da possível descoberta de genes responsáveis por eventos fundamentais da

diferenciação celular e do desenvolvimento dos tecidos foliares e vasculares, a identificação de genes órgão- ou tecido-específicos poderá permitir o futuro isolamento de suas seqüências promotoras para uso no direcionamento da expressão transgênica em plantas.

Os objetivos específicos propostos foram:

- Analisar microarranjos de DNA utilizando ferramentas de bioinformática;
- Selecionar os genes mais diferencialmente expressos em tecido vascular e folhas de *E. grandis*;
- Analisar o padrão de expressão dos genes selecionados pela técnica de qRT-PCR;
- Comparar o nível de expressão dos genes em relação às duas técnicas utilizadas.

2- REVISÃO BIBLIOGRÁFICA

2.1. A Indústria de Celulose e Papel & o Eucalipto

O setor de florestas plantadas ocupa atualmente lugar de destaque entre os diferentes segmentos industriais nacionais no que se refere à geração de renda, arrecadação de tributos, mão-de-obra empregada e geração de divisas. A disponibilidade de madeira de plantio, mais barata e homogênea que a das florestas naturais, é uma alternativa eficiente para a demanda de biomassa lenhosa, diminuindo a devastação das florestas nativas. Uma grande quantidade de madeira é manufaturada para gerar combustível, fibras (para produção de pasta de celulose e diversos tipos de papel) e madeira para serrarias (indústria de construção civil e movelaria).

O uso das Florestas Plantadas representa 4% do Produto Interno Bruto do Brasil, sendo que desses, 35% corresponde ao setor de papel e celulose (ABRAF, 2007). O Brasil é o sétimo maior produtor de celulose mundial, ocupando o primeiro lugar quando se trata de celulose de fibra curta de mercado. Ocupa o 11º lugar como produtor de papel mundial segundo dados da Associação Brasileira de Celulose e Papel - Bracelpa (BRACELPA, 2006). O Setor Florestal Brasileiro contribui com 8,4% das exportações do país (Sociedade Brasileira de Silvicultura, 2006). Os principais mercados para exportação da celulose brasileira têm sido a Europa (49%), a Ásia (30%) e a América do Norte (19%). No segmento de papel, os principais mercados são a América Latina (55%), a Europa (17%), a América do Norte (16%), a Ásia (7%) e a África (5%). Grande parte desse sucesso deve-se à qualidade da madeira que cresce nas florestas plantadas (IPEF, 2004).

O setor brasileiro de celulose e papel é constituído por 220 empresas, localizadas em 16 estados, sendo 35 empresas exportadoras

habituais. As indústrias nacionais do setor aplicaram US\$ 12 bilhões nos últimos 10 anos para ampliar a capacidade produtiva, expandindo a produção nacional de celulose de fibra curta de mercado de 1,4 para cerca de 10,1 milhões de toneladas anuais. O segmento de celulose e papel utiliza madeira exclusivamente de florestas plantadas e teve em 2005 um faturamento de U\$ 8,9 bilhões (Sociedade Brasileira de Silvicultura, 2006).

O Brasil possui área total absoluta de 851 milhões de hectares (ha), sendo 477,7 milhões desse total correspondentes a florestas naturais e 5,6 milhões a florestas plantadas. O país conta, ainda, com 61,8 milhões de ha de unidades federais de conservação sob regime de proteção integral (45,5%) e de uso sustentável (54,5%). Assim, as florestas plantadas ocupam apenas 0,65% do território nacional e 1% do solo agropecuário. Dos 5,6 milhões de ha de florestas plantadas, 3,4 milhões de ha estão ocupados com eucalipto, 1,8 milhões de ha com *Pinus* e 326 mil ha com outras espécies, como acácia-negra, gmelina, álamo, seringueira, teca e araucária (BRACELPA, 2006; Sociedade Brasileira de Silvicultura, 2006).

A silvicultura teve início no Brasil no início do século passado, com o estabelecimento dos plantios florestais com espécies exóticas para substituição da madeira das florestas nativas de difícil reposição. As principais espécies exóticas foram os eucaliptos, introduzidos pela Companhia Paulista de Estrada de Ferro em 1904, e as coníferas (*Pinus*), pela Companhia Melhoramentos de São Paulo em 1922 (ABRAF, 2007).

O gênero *Eucalyptus* (família *Myrtaceae*) inclui mais de 700 espécies descritas que, somadas a diversos híbridos naturais e artificiais, totalizam mais de 1.500 formas distintas com grande variabilidade genética. Proveniente da Austrália, Nova Guiné, Indonésia e Filipinas, o *Eucalyptus* é o gênero arbóreo de mais ampla distribuição no planeta (Neilson, 2000; revisado em Poke *et al.*, 2005; Revista da Madeira, 2007).

O gênero foi introduzido no país em 1904 e, desde então, seu plantio passou por diversas etapas de adaptação ao ambiente, melhoramento da qualidade da madeira e sustentabilidade. Hoje, buscam-se características de produtividade, qualidade e eficiência combinadas à menor geração de impactos ambientais e melhor qualidade de vida dos trabalhadores e das comunidades das regiões dos plantios (ABRAF, 2006).

As principais espécies de *Eucalyptus* cultivadas atualmente no Brasil são o *E. grandis*, *E. citriodora*, *E. camaldulensis*, *E. saligna*, *E. urophylla*, entre outras. Além disto, foram desenvolvidos cruzamentos entre estas e muitas outras espécies, derivando-se delas diversos híbridos, como é o caso do *Eucalyptus urograndis* (*E. urophylla* x *E. grandis*). Entre as centenas de espécies e híbridos existentes, *E. grandis* é uma das espécies de maior importância comercial, utilizada como base principal em vários programas de melhoramento genético ao redor do mundo, principalmente por suas características de rápido crescimento e alta densidade da madeira (Neilson, 2000; FAO, 2007).

2.2. *Eucalyptus grandis*

O *E. grandis* é a espécie de eucalipto mais plantada fora da Austrália. Lá, possui ocorrência natural de forma descontínua e fragmentada numa longa faixa costeira (de Newcastle até Atherton). O clima na área é de temperado-quente a subtropical-moderado, com invernos suaves e chuvas abundantes e bem distribuídas. (IPEF, 2004).

No Brasil, *E. grandis* é a espécie mais importante de *Eucalyptus* utilizada na produção de papel e pasta de celulose, especialmente na forma do híbrido *E. urophylla* x *E. grandis* (mais conhecido entre os técnicos florestais como “*Urograndis*”), mas também na forma de outros híbridos. *E. grandis* apresenta um bom desenvolvimento em regiões de clima tropical e subtropical, além de conter um ótimo teor de celulose na

madeira (em torno de 48%) e possuir um bom crescimento volumétrico (Colodette *et al.*, 2004; revisado em Poke *et al.*, 2005). Estudos realizados em 1983 por Brito *et al.* no Instituto de Pesquisas e Estudos Florestais (IPEF) e no Departamento de Silvicultura da Escola Superior de Agronomia Luiz de Queiroz (ESALQ/USP) demonstraram que o *E. grandis* é uma das espécies de eucalipto que possui o maior volume cilíndrico médio, com cerca de 1.105 m³/ha. Porém, possui elevados índices de lignina na madeira (24% de lignina total; Pereira *et al.*, 2000; Colodette *et al.*, 2004).

No presente ano de 2008, em virtude de sua importância econômica, tamanho reduzido do seu genoma e valor silvicultural de seus híbridos, o *E. grandis* foi eleito a espécie vegetal arbórea a ter seu genoma completamente seqüenciado pelo JGI do USDOE. Os coordenadores e principais responsáveis pelo projeto de seqüenciamento são o Dr. Gerald. A. Tuskan (JGI, Walnut Creek, California, EUA), Dr. Alexander Myburg (Universidade de Pretória, Pretória, África do Sul) e Dr. Dario Grattapaglia (Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF), coordenador geral do Projeto GENOLYPTUS. Assim, o *E. grandis* tornar-se-á em breve a terceira espécie arbórea com seu genoma definido e o natural modelo para as futuras pesquisas científicas em *Eucalyptus*. As duas outras espécies arbóreas cujos genomas foram totalmente seqüenciados são o álamo (*Populus trichocarpa*; revisado em Brunner *et al.*, 2004; Tuskan *et al.*, 2006) e a videira (*Vitis vinifera*; Jaillon *et al.*, 2007; Velasco *et al.*, 2007).

2.3. A Gênese da Madeira

Há considerável interesse na identificação de marcadores genéticos que caracterizem os principais eventos da formação da madeira, pois esses são considerados fatores-chave na determinação das suas principais

características, influenciando seu desempenho, qualidade e valor industrial (Paux *et al.*, 2005). A madeira é composta por traqueídeos, vasos e fibras que estão envolvidos no transporte de água e no suporte mecânico da árvore (Whetten *et al.*, 2001). As propriedades da madeira variam muito entre as espécies e entre indivíduos/genótipos da mesma espécie. A variabilidade ocorre tanto entre diferentes tecidos (proporção de diferentes tipos celulares) bem como em nível celular: tamanho, forma, estrutura da parede celular, textura e composição química (Paux *et al.*, 2005).

Pouco se sabe sobre os processos celulares, moleculares e do desenvolvimento que estão por trás da formação da madeira (xilogênese). A xilogênese representa um exemplo de diferenciação celular complexo. Este processo é controlado por uma grande variedade de fatores, tanto exógenos (fotoperíodo e temperatura) quanto endógenos (fito-hormônios) e também pela interação entre eles. A madeira (xilema secundário) é formada por uma sucessão de cinco principais eventos que incluem: (i) divisão celular; (ii) expansão celular (alongação e alargamento radial); (iii) espessamento da parede celular; (iv) lignificação; e (v) morte celular programada. Todos estes estágios são controlados pela expressão coordenada de numerosos genes com funções regulatórias e estruturais, sendo a maioria ainda desconhecida (Hertzberg *et al.*, 2001; Plomion *et al.*, 2001; Yang *et al.*, 2004). O estudo de genes e mecanismos responsáveis pela formação da parede secundária, pela biossíntese de lignina e celulose e pela xilogênese teve grande progresso nos últimos anos (Beers & Zhao, 2001; Boudet *et al.*, 2003; Ohashi-Ito *et al.*, 2003; Oh *et al.*, 2003; Paux *et al.*, 2004; Prassinis *et al.*, 2005).

A formação da madeira é um processo que se origina a partir do crescimento secundário das plantas, com a atividade do câmbio vascular. O sistema vascular secundário compreende o floema secundário (casca), o

câmbio vascular (câmbio) e o xilema secundário (lenho) e é um sistema dinâmico de quatro dimensões. O câmbio é uma bainha de tecido meristemático que circunda o tronco e a raiz da árvore. O câmbio mantém-se por divisões celulares repetidas e produz linhagens celulares radiais, dando origem às células do xilema secundário em direção ao seu interior e às células do floema secundário em direção ao seu exterior. Quando uma célula se distancia do câmbio, ela diferencia-se até tornar-se uma célula madura. Assim, é possível reconhecer os estágios de diferenciação de uma célula individual partindo do ponto final voltando ao ponto inicial do câmbio. Essa característica do sistema vascular secundário torna-o ideal para o estudo de diferenciação celular. O crescimento do câmbio vascular aumenta o diâmetro e a circunferência da árvore. O controle do crescimento e diferenciação do câmbio é acompanhado por mudanças na expressão de genes importantes envolvidos com o desenvolvimento (Chaffey, 1999).

A ação dos genes e das proteínas presentes no xilema secundário em diferenciação irá determinar as propriedades da madeira. A composição e a morfologia da parede celular secundária das células do xilema são os principais fatores responsáveis pela qualidade da madeira (Whetten *et al.*, 2001). A diferenciação do xilema secundário envolve mudanças seqüenciais na parede celular. Isto inclui a expansão da parede celular primária e a acomodação da parede celular secundária lignificada, seguida pela digestão de componentes da parede (Bourquin *et al.*, 2002; Matsui *et al.*, 2005). Após a célula ter parado de crescer, a parede do xilema é engrossada com depósitos de celulose e depois é fortalecido pela impregnação com polissacarídeos da matriz da parede celular secundária e lignina. As paredes resultantes fornecem os elementos traqueais com força suficiente para resistir à pressão negativa dentro dos vasos, assim como para suportar o peso do próprio tronco (revisado em Ye, 2002). A

formação da parede celular secundária é comandada pela expressão coordenada de vários genes, envolvidos especificamente na síntese e deposição de polissacarídeos, ligninas, pectinas e proteínas de parede celular (Plomion *et al.*, 2001; Whetten *et al.*, 2001; Yang *et al.*, 2004). Assim, as principais características da madeira irão variar de acordo com o tamanho, a forma e a organização das células no xilema, bem como a estrutura e a composição química da parede celular secundária.

A lignina é o maior componente da madeira depois da celulose e causa um impacto negativo no processo de polpação química. Neste, as ligninas são dissolvidas para liberar o material fibroso constituído basicamente de holocelulose (celulose e hemiceluloses). A dissolução das ligninas na polpação depende do tipo de ligações e dos grupos funcionais presentes em sua estrutura. Em geral, madeiras com maior teor de celulose e menor de hemiceluloses deverão resultar em maior rendimento de polpação. Tendo em vista que o objetivo da polpação é a remoção seletiva das ligninas e a liberação da porção fibrosa da madeira, as madeiras com baixos teores de ligninas e de extrativos propiciam melhor desempenho da polpação em termos de rendimento e consumo de produtos químicos utilizados neste processo (Colodette *et al.*, 2004). Quando as ligninas são removidas, essas propriedades também irão determinar a resistência, a densidade da polpa e a qualidade dos produtos de papel (Whetten *et al.*, 2001).

2.4. Genômica Funcional

Os primeiros estudos de genômica funcional realizados em plantas basearam-se na planta-modelo *Arabidopsis thaliana* (Schena *et al.*, 1995; Girke *et al.*, 2000; White *et al.*, 2000; Schaffer *et al.*, 2001). Recentemente, houve um aumento do interesse na genômica funcional aplicada a plantas de alto interesse comercial, como o arroz (Yamamoto &

Sasaki, 1997; Yazaki *et al.*, 2004; Rafalski, 2007), tomate (Emmanuel & Levy, 2002; Alba *et al.*, 2005), cana-de-açúcar (Vettore *et al.*, 2003; Calsa & Figueira, 2007), uva (Goes da Silva *et al.*, 2005; Waters *et al.*, 2006), pêsego (Abbott *et al.*, 2002; Trainotti *et al.*, 2006), milho (Lawrence *et al.*, 2004; Haberer *et al.*, 2005) e *Pinus* (Whetten *et al.*, 2001).

A genômica funcional oferece os meios para a identificação de um grande número de genes envolvidos na parede celular do xilema e pode oferecer uma visão compreensiva desses genes. Atualmente, está disponível a tecnologia de microarranjos como uma técnica poderosa para tais fins. Com a ajuda desta tecnologia, é possível monitorar o perfil de expressão de centenas de milhares de genes que atuam durante a lignificação e a biossíntese da parede celular secundária (Oh *et al.* 2003; Ko *et al.*, 2004). Com o sistema de análise de agrupamentos (*clusters*) para dados de expressão gerados por análises de microarranjos, aumenta-se o número de genes correlacionados com a síntese da parede celular secundária (Persson *et al.*, 2005). Elementos traqueais diferenciados de células mesofílicas isoladas de *Zinnia elegans* tornaram-se conhecidos como um sistema modelo para o estudo da regulação gênica durante a diferenciação do xilema (Fukuda, 1997). No sistema de *Z. elegans* a formação da parede secundária única aos elementos traqueais é induzida por alterações no balanço hormonal. Demura *et al.* (2002) executaram amplas análises de microarranjos de expressão gênica nesse sistema de diferenciação de células e identificaram muitos genes candidatos envolvidos na formação da parede secundária. Muitos dos genes relacionados à gênese da madeira estão sendo descobertos pelo uso de técnicas de genômica funcional em larga escala em espécies como *Eucalyptus* (Paux *et al.*, 2004; Paux *et al.*, 2005; Foucart *et al.*, 2006), *Populus* (Hertzberg *et al.*, 2001; Israelsson *et al.*, 2003; Schrader *et al.*,

2004; Aspeborg *et al.*, 2005; Prassinos *et al.*, 2005; Andersson-Gunnera *et al.*, 2006), *Pinus* (Allona *et al.*, 1998; Whetten *et al.*, 2001; Lorenz & Dean, 2002; Yang *et al.*, 2004) e até *Robinia pseudoacacia* (Falsa Acácia; Yang *et al.*, 2003).

Entretanto, mesmo com o advento de vários projetos de seqüenciamento e de genômica funcional, os mecanismos moleculares que levam ao crescimento secundário das plantas ainda precisam ser investigados. Mesmo que as principais vias estruturais já tenham sido descritas, pouco se sabe a respeito da regulação deste processo, tanto em nível celular e molecular quanto do desenvolvimento (Schrader *et al.*, 2004; Li *et al.*, 2006; Groover & Robischon, 2006).

As comparações entre genes expressos em diferentes tecidos da mesma planta e nos mesmos tecidos de diferentes espécies podem fornecer noções sobre os processos do desenvolvimento e explicações de por que algumas espécies possuem propriedades mais favoráveis como maior densidade da madeira ou maior rendimento de polpação (Poke *et al.*, 2005). Adicionalmente, a identificação de genes com expressão órgão ou tecido-específica poderia permitir o isolamento de seqüências regulatórias de DNA (promotores) responsáveis por tais padrões de expressão. Estas seqüências promotoras, assim, consistem importantes ferramentas da engenharia genética vegetal de forma a melhor modular a expressão de transgenes em plantas. Tendo estas noções em mente, microarranjos de DNA foram realizados pelos pesquisadores do Projeto GENOLYPTUS com a finalidade de estudar mais profundamente as diferenças entre tecidos e espécies de *Eucalyptus*.

2.5. Genes Xilema-Específicos e a Engenharia Genética

Uma das principais vantagens da genômica funcional é fornecer conhecimento para a aplicação no melhoramento genético de plantas. Os

resultados deste e de outros trabalhos envolvendo análise de dados de microarranjos de DNA em diferentes tecidos ou espécies de *Eucalyptus* fornecerão genes xilema-específicos candidatos a estudos de engenharia genética.

Há um grande interesse em alterar a expressão de genes específicos à síntese de vasos do xilema utilizando as ferramentas da engenharia genética. Vários estudos envolvendo a regulação da quantidade e da qualidade dos componentes de ligninas já foram realizados. Enzimas envolvidas na lignificação têm sido clonadas de diversas espécies de árvores e, em alguns casos, a expressão desses genes tem sido alterada, resultando na modificação ou redução dos níveis de lignina (Grima-Pettanti & Goffner, 1999; revisado em Chiang, 2006). Muitas outras propriedades da madeira e genes envolvidos na codificação de enzimas estruturais ou fatores de transcrição também estão sendo alterados ou são potenciais alvos para a engenharia genética (revisado em Groover, 2005; revisado em Chiang, 2006).

Promotores constitutivos e não-específicos como os codificadores do RNA 35S do Vírus do Mosaico da Couve-flor (CaMV-35S; Lawton *et al.*, 1987) e da nopalina-sintase (NOS) de *Agrobacterium tumefaciens* (An *et al.*, 1986) são amplamente utilizados em estudos visando a super-expressão gênica em plantas. Porém, a expressão do gene ocorre na maioria dos tecidos (Gittins *et al.*, 2003), sem regulação temporal e espacial, provocando um crescimento anormal da planta (Lu *et al.*, 2004). O uso de promotores com expressão específica permite o controle da expressão de genes exógenos em plantas transgênicas em tecidos específicos e num momento específico. Promotores de expressão específica têm sido estudados intensivamente nos últimos anos devido a sua importância na expressão de genes, desenvolvimento e crescimento de plantas transgênicas. O uso de promotores específicos torna

necessário conhecer elementos regulatórios xilema-específicos (Lu *et al.* 2003).

Li *et al.* (1999) isolaram um gene de cafeoil-coenzima A 3-O-metiltransferase (CCoAOMT) de *Pinus taeda* expresso especificamente no xilema secundário. Esta enzima catalisa a metilação de precursores monoméricos hidroxilados de lignina e, portanto, está envolvida nas vias de biossíntese de lignina e é de interesse para o melhoramento da produção de polpa de madeira de árvores coníferas por engenharia genética (Whetten *et al.*, 1998). A disponibilidade desse promotor de CCoAOMT de *Pinus* permitiria sua aplicação na expressão de genes de angiospermas xilema-específicos para o direcionamento da expressão transgênica de genes de interesse em gimnospermas. O uso de inibição antisense de CCoAOMT sob o controle do promotor do gene da CCoAOMT em *Pinus* transgênico poderia ajudar na diminuição da lignificação ou alterar a composição das ligninas de forma a facilitar o processo de polpação.

Lu *et al.* (2003) utilizaram o promotor de um gene codificador de proteínas ricas em glicinas (SjGRP1.8) de *Sopho japonica* L. para dirigir a expressão do gene codificador da 4-coumarato:CoA ligase (4CL1) em tabaco transgênico. O promotor GRP1.8 clonado possui expressão específica em tecidos vasculares. A enzima 4CL1 é componente-chave na via biossintética de monolignóis. Neste sistema, a atividade da enzima 4CL1 aumentou de uma a duas vezes no xilema e não foi observado aumento nas folhas de tabaco transgênico. Em relação às plantas controle, houve um aumento no conteúdo de ligninas de 25% no tronco, não ocorrendo aumento nas folhas das plantas transgênicas. Em 2004, Lu *et al.* utilizaram novamente o gene 4CL1, mas desta vez fusionado aos promotores Pto4CL1 (clonado de *Populus tomentosa*) e CaMV-35S, para avaliar a expressão quantitativa e qualitativa em tabaco transgênico. O

promotor Pto4CL1 induziu a expressão do gene 4CL1 no xilema. A indução positiva da expressão do gene 4CL1 (tanto pelo promotor 35S quanto pelo Pto4CL1) causou um aumento no conteúdo de ligninas de 25% no tronco das plantas transgênicas de tabaco em relação às plantas-controle. Não houve aumento no conteúdo de ligninas nas folhas das plantas transgênicas com o emprego do promotor de Pto4CL1, enquanto nas plantas transgênicas contendo a construção CaMV-35S-Pto4CL1, houve aumento de 20% também nas folhas.

As proteínas GRP constituem uma classe de proteínas simples estruturadas em plantas contendo domínios ricos em glicina, que variam de 20% a 70% no conteúdo de glicina (Liu *et al.*, 2003). Estudos de hibridização *in situ* e imunolocalização da expressão de GRPs indicaram que elas podem ser altamente tecido- ou órgão-específicas, sendo, na maioria dos casos, específicas aos tecidos vasculares (Keller *et al.*, 1988; Cheng *et al.*, 1996) ou raiz-específicas (Bergeron *et al.*, 1994; Matsuyama *et al.*, 1999). A localização das GRPs sugeriu um papel como componente estrutural de matrizes extracelulares (Cassab, 1998). Um estudo do promotor de um gene GRP de arroz (*Oryza sativa* L.) denominado Osgrp-2 foi realizado por Liu *et al.* (2003). Neste estudo, o gene codificador da β -glicuronidase (GUS) de *Escherichia coli* foi utilizado como gene-repórter fusionado ao promotor Osgrp-2. As seqüências responsáveis pela atividade do promotor e pela capacidade de conferir expressão vascular-específica foram identificadas por meio da análise de plantas de tabaco transgênicas utilizando análises de deleção 5' e deleções internas do promotor. Como resultado, pode-se verificar que o promotor pode ser dividido em três partes relevantes à sua atividade transcricional: uma região distal 5', uma região proximal 5' - ambas com função regulatória positiva - e, entre elas, uma região regulatória negativa. A atividade vascular-específica foi atribuída à região proximal 5', mais precisamente a

uma seqüência de 99 pares de bases (pb) que parece ter dupla função, isto é, sendo responsável por aumentar a expressão em tecidos vasculares e suprimir a expressão em outros tecidos. A busca por elementos *cis* desta seqüência em bancos de dados revelou um CCAAT-*box* nesta região, que é um sítio de ligação para fatores de transcrição da classe MYB, e que foram descritos como envolvidos na expressão específica a tecidos meristemáticos e condutores associados a feixes vasculares em tabaco transgênico (Wissenbach *et al.*, 1993).

Modular a síntese de celulose poderá ter um impacto direto em vários aspectos do desenvolvimento e do crescimento de uma planta. Um profundo entendimento dos processos de biossíntese de celulose é de extrema importância para futuros experimentos que possuem como alvo o melhoramento genético da produção de celulose em plantas importantes economicamente (Doblin *et al.*, 2002). As endoglicanases KORRIGAN (KOR) parecem ter um importante papel na biossíntese de celulose em *Arabidopsis* (Joshi *et al.*, 2004), sendo necessárias para a formação da parede nas células das plantas em crescimento e sendo mais expressas durante o desenvolvimento do xilema. Bhandari *et al.* (2006) estudaram um clone de cDNA de KOR do xilema de *Populus tremuloides* (PtrKOR) e verificaram que o gene é positivamente regulado no xilema e apresenta resposta a estresses juntamente com três outros genes (PtrCesA1, PtrCesA2 e PtrCesA3) codificadores de celulose-sintases (CesA) associadas à parede celular secundária de *Populus*. A utilização da engenharia genética para alterar os níveis de expressão de KOR (com alteração coordenada ou não dos níveis de CesA) poderá resultar em plantas transgênicas que produzam fibras e madeira de melhor qualidade.

A caracterização dos elementos regulatórios xilema-específicos é um importante passo antes da aplicação de técnicas de transgênese. Os elementos flanqueadores a montante e a jusante de dois genes de *Pinus*

com expressão preferencial no xilema foram estudados por No *et al.* (2000). Os genes PtX3H6 e PtX14A9 codificam componentes do núcleo protéico de arabinogalactano-proteínas (AGPs), um grupo de proteoglicanas ou glicoproteínas envolvidas no desenvolvimento da planta e que freqüentemente exibem expressão tecido-específica (Nothnagel, 1997). As seqüências flanqueadoras 3' e 5' foram testadas em combinação com o promotor CaMV-35S e o terminador NOS em tabaco e álamo transgênicos. A conclusão foi de que tanto a seqüência 3' quanto a seqüência 5' dos dois genes são necessárias para a expressão correta de ambos. No entanto, as seqüências 3' possuem diferentes funções para cada gene. Em PtX3H6, esta região atua como inibidor enquanto em PtX14A9, ela atua como ativador, aumentando os níveis de expressão do gene de forma tecido-específica. Deve-se levar em consideração que mesmo em angiospermas, a expressão de um gene-repórter não é sempre igual a da espécie da qual o promotor foi obtido (Stalberg *et al.*, 1998)

Muitos genes utilizados na transgênese para a melhoria da madeira já foram identificados e muitos ainda o serão a partir dos projetos genoma de árvores. O uso de *A. thaliana* como modelo genético foi combinado com sua habilidade em produzir tecidos lenhosos para o isolamento de genes expressos em tecidos vasculares (Beers & Zhao, 2001). No trabalho de Beers e Zhao (2001), foram encontrados três genes xilema-específicos: *XCP1*, *XCP2* e *XSP1*. Enquanto os genes *XCP1* e *XCP2* parecem codificar cisteíno-proteases semelhantes à papaína, *XSP1* parece codificar uma serino-protease semelhante à subtilisina. Novos promotores que atuam especificamente em tecidos vasculares podem ser identificados a partir de cDNAs por intermédio de seqüências anotadas dos genomas de *Arabidopsis* (Kubo *et al.*, 2005; Persson *et al.*, 2005; Zhao *et al.*, 2005), *Populus* (Brunner *et al.*, 2004) e *Pinus* (Yang *et al.*, 2004). Assim, surge

uma importante fonte de genes e promotores úteis para a modificação da madeira em espécies de árvores economicamente importantes.

Atualmente observa-se um grande esforço das empresas e instituições de pesquisa direcionado ao melhoramento da madeira com a finalidade de se obter um maior rendimento a partir de menor quantidade de matéria-prima. A aplicação segura e cuidadosa da biotecnologia (cultivo assistido por marcadores, engenharia genética e propagação *in vitro*) nas práticas florestais deve ajudar a desenvolver árvores geneticamente superiores em um período de apenas algumas décadas (Plomion *et al.*, 2001).

CAPÍTULO II

Differential expression in *Eucalyptus grandis*

(Trabalho a ser submetido ao periódico *BMC Genomics*)

Marina Tagliaro Jahns¹, Fernanda Macedo Bastolla¹, Giancarlo Pasquali^{1,*}

¹Programa de Pós-graduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil;

* Corresponding author.

Giancarlo Pasquali^{1,*}, pasquali@cbiot.ufrgs.br

C.P. 15.005, CEP 91501-970. Tel.: +55 51 3308 9410

Abstract

Paper and pulp applications depend on the wood quality and the industrial process. For these main reasons, there is a need to increase *Eucalyptus* genomic resources in order to select candidate genes involved in the genetic control of wood properties. Researchers of the Brazilian Research Network on the *Eucalyptus* Genome (The Genolyptus Project) are engaged in the discovery, sequencing and mapping of economically important genes and the determination of their activities in different species of *Eucalyptus*, aiming the incorporation of new genomic technologies into the advanced eucalypt breeding programs and forest production. As part of the Genolyptus Project we have designed and analyzed DNA microarrays in order to identify differentially expressed genes between leaf and xylem tissues of *E. grandis*. This study was made with the objective to standardize the analysis of DNA microarrays aiming a future focused study on *Eucalyptus* wood properties. Quantitative reverse-transcription polymerase chain reaction (qRT-PCR) results proved the full validation of the microarray experiment, as well as its analysis. Results obtained by qRT-PCR were consistent with the microarray results for the genes analyzed, although the relative expression ratios of the first were often higher. Similar discrepancies between methods were reported previously, suggesting that ratios calculated using DNA arrays are often underestimated. We believe that our microarray analysis and biological sampling were fully supported by the qRT-PCR findings because it was robust enough to evaluate a massive number of genes and to point out important candidate genes.

Keywords: *Eucalyptus grandis*, microarray analysis, differential expression, qRT-PCR.

Background

The Planted Forest sector is, today, one of the most important for the Brazilian national economy extensively contributing to income generation, tribute collection, employment and generation of currencies. The use of planted forests represents 4% of the Gross Domestic Product of Brazil, in which 35% corresponds to the pulp and paper sector [1]. Brazil is the seventh major pulp and paper producer worldwide, occupying the first place considering short-fiber cellulose, totally derived from *Eucalyptus*. Considering paper as final product, Brazil is the eleventh worldwide producer [2].

The total area of Brazilian planted forests corresponds to 5.3 million hectares (ha), in which 3.5 million ha account for *Eucalyptus* forests and 1.8 million ha account for *Pinus*. Brazil has 220 paper and pulp enterprises located in 16 Federative States. In Brazil, 100% of cellulose pulp & paper are derived from planted forests, mostly eucalypts or pines. Exported wood-derived products reached figures of US\$ 4.0 billion in 2006. In the same year, Brazilian pulp production reached 11.1 million tons and the paper production was 8.8 million tons. Eucalypt forests are located mainly in the Southeast region, primarily in the States of Minas Gerais and São Paulo, as well as Bahia in the Northeast region [1].

The genus *Eucalyptus* (*Myrtaceae* family) includes more than 500 described species from Australia, New Guinea, Indonesia and Philippines. It is the most widely dispersed arboreal genus in the planet, mainly used for industrial plantations, and it was introduced in Brazil in 1904 [3]. Since then its cultivation went through many environmental adaptation steps, improvement of the wood quality and sustainability. *Eucalyptus* species are among the world's main sources of biomass and represent the

principal hardwood trees used for making paper pulp [4]. The main reason for the good acceptance of the Brazilian production is the high quality of the wood used. Characteristics like productivity, quality and efficiency were combined with the lower generation of environmental impacts and better life quality of the employees and the communities from the planting area [1].

The most cultivated *Eucalyptus* species in Brazil nowadays are *E. grandis*, *E. citriodora*, *E. camaldulensis*, *E. saligna* and *E. urophylla*. Moreover, crossings between species were developed resulting in many hybrids like “*E. urograndis*” (*E. urophylla* x *E. grandis*), the most exploited *Eucalyptus* hybrid in Brazil. Amongst hundreds of species and hybrids, *E. grandis* is one of the main species used in many genetic improvement programs around the world, primarily because of its fast growth and high wood density [3, 5]. *E. grandis* is suitable for genomic analysis because it has a small genome (approximately 630 Mbp) and genetic maps and mapped QTLs are available [6, 7].

Wood quality and the industrial process will mostly define the paper and pulp applications and due to them many large efforts aiming to increase *Eucalyptus* genomic resources are being developed in order to select candidate genes involved in the genetic control of wood properties. Brazilian enterprises and scientists are internationally recognized for the genetic research and development, genetic improvement, molecular markers and *Eucalyptus* cloning. For this reason, The Genolyptus Project was created in 2002 by a group of researchers from 10 Brazilian research institutes. The project is named the Brazilian Research Network on the *Eucalyptus* Genome and it is supported by the Brazilian Ministry of Science and Technology and 12 paper and pulp enterprises of the private sector. Some of the main goals of this project are focused in the discovery, sequencing and mapping of economically important genes and the

determination of their activities in different species of *Eucalyptus*, aiming the incorporation of new genomic technologies into the advanced eucalypt breeding programs and forest production.

The subproject "Sequencing of the *Eucalyptus* Transcriptome" entitles the sequencing effort to achieve 150,000 expressed sequence tags (EST) from different expression libraries. The objective was the identification of all 30,000 estimated eucalypt genes, with special emphasis on those involved in wood formation. More than 13 libraries were constructed from different organs, tissues and whole seedlings submitted to several treatments derived from four *Eucalyptus* species. Approximately 96,500 ESTs were produced, analyzed and arranged into 34,000 clusters based on sequence homology. The putative functions of sequenced clones were attributed automatically by BLAST search. ESTs were grouped according to their putative biological function and those with "no hit", "unknown function" or "hypothetical/putative function" represented 60 percent of the total.

The so called "Woodgenes Project" or "The Genesis of Wood in *Eucalyptus*: Genes, Functions, Regulation and Transgenic Expression", started in January 2005 and its main focus is the exploration of the cDNA clone collection generated in the Genolyptus Project. Special emphasis is driven to cDNAs involved in the poorly understood wood formation process, aiming the incorporation of new genomic technologies into the advanced eucalypt breeding programs and forest production. To this end, DNA microarrays were planned, ordered and analyzed.

The information about the process of wood formation is still limited, despite the great importance of such material. The systematic analysis of genes involved in wood formation in *Eucalyptus* provides valuable insights into the molecular mechanisms involved in secondary xylem

differentiation as well as new candidate-genes for wood quality improvement [8].

Robotics and bioinformatics have provided a way to monitor the expression of thousands of genes throughout the construction and use of DNA microarrays [9]. Many genes related to xylogenesis have been discovered with the use of such tools during the analysis of large-scale gene expression in *Eucalyptus* [10, 8, 4], *Populus* [11-16], *Pinus* [17-20] and *Robinia pseudoacacia* [21]. As part of the Genolyptus Project we have designed and analyzed DNA microarrays in order to identify differentially expressed genes between leaves and xylem of *E. grandis*. This study was proposed with the objective to standardize the analysis of DNA microarrays aiming a future focused study of the *Eucalyptus* wood properties.

Results and Discussion

In order to pinpoint the most differentially expressed genes between leaf blades and vascular (xylem) tissues from *E. grandis* adult trees, a total number of nine 50mer-oligoprobes covering the length of each one of the 21,442 unique sequences derived from the Genolyptus EST collection were synthesized “on-chip” in duplicate, randomly distributed in the two blocks of each slide. Therefore, four identical chips were produced at NimbleGen Scientific Inc. with a total number of 385,956 features each. Total RNA samples derived from leaf blades and xylem tissues were obtained from two 4-years-old *E. grandis* clones, i.e., two genetically identical trees generated from one single matrix-tree. Complementary cDNAs were synthesized from the two samples of leaf RNA and the two samples of xylem RNA at Nimblegen Scientific Inc. with Cy3 as fluorescent marker. After hybridization, washings, data collection and pre-

analysis, DNA microarray data were analyzed in order to verify the consistency provided by Nimblegen. For this purpose, a simple *t*-test was used comparing various sample groups (data not shown). The reliability of the data was proved, meaning that the microarrays showed very low variation between blocks from the same slide and between replicates. In summary, all four microarrays exhibited great reproducibility and robustness.

The second analysis was performed in order to determine the most differentially expressed genes between leaves and xylem of *E. grandis*, using the SAM algorithm. The SAM algorithm is part of the MeV package (freely available from the TIGR webpage) and is an alternative to Student *t* test aiming differential expression. The method is based in the ratio of the change in the gene expression between two conditions and the standart error of the measurements. SAM identifies differentially expressed genes by the use of sets of *t* tests gene-specific. Those genes with values higher than the cutoff point are considered as potentially significant. The percentage of those genes random identified is called "false discovery rate" (FDR). The algorithm creates an interactive graphic where are the observed values versus the expected values (based on the permutation data or *d*-values). The cutoff point can be adjusted to identify lower or higher set of genes, according to the delta value. FDRs are calculated to each set of data. The delta value can be handily adjusted in the graphic and it means the vertical distance (in graphic units) from the solid line that represents the inclination degree (when the observed is equal to the expected) to each one of the two parallel dotted lines that represent the area in which genes are consider differentially expressed. It is also possible to use the fold change criterion, which will determine the number of times that a single gene is more or less expressed in relation to another [22].

Considering an at least 5-fold change differential expression calculated by the SAM algorithm, 153 genes were overexpressed in leaves and 149 genes were much highly expressed in xylem (Figure 1). In Table 1 is shown the frequency of the differentially expressed genes in each tissue analyzed in accordance to their fold change in expression.

Fold change	Number of genes overexpressed in leaves	Number of genes overexpressed in xylem
500-100	39	10
100-10	102	118
10-5	12	21
Total	153	149

Table 1. Number of genes from *E. grandis* overexpressed in leaves and xylem for each fold change classification. Most genes were classified within the group of 10- to 100-fold increase in expression.

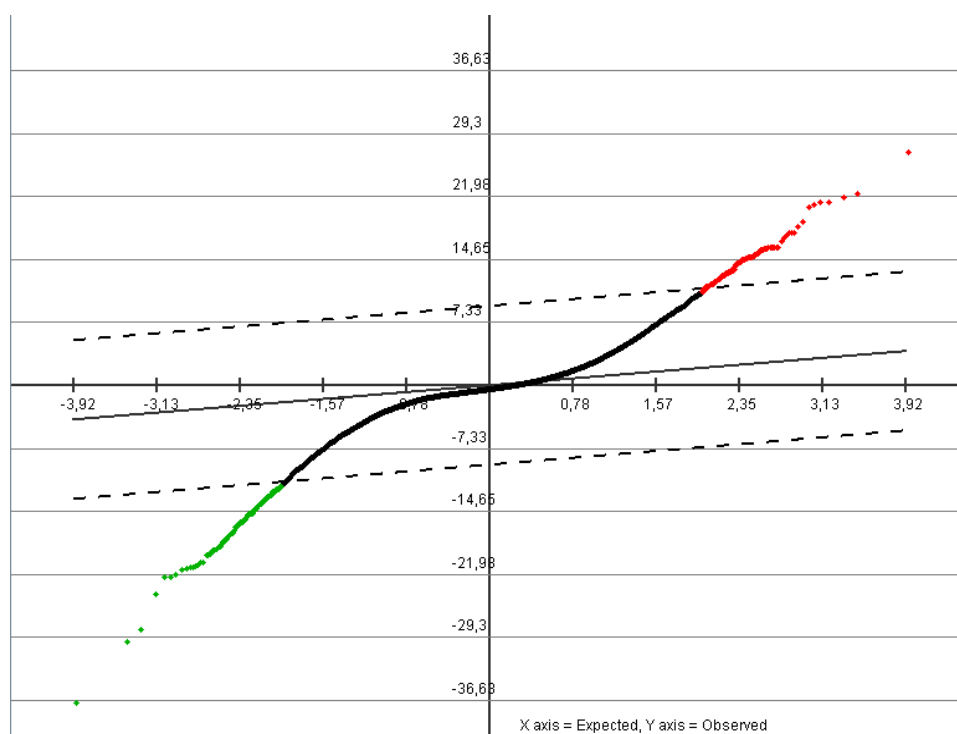


Figure 1. Graphical representation generated by the SAM algorithm. In the “x” axis are the expected values and in the “y” axis are the observed values. The black solid line represents those values that did not differ between observed and expected. In green are shown all 153 up-regulated genes in the leaves and in red are shown all 149 up-regulated genes in the xylem.

Fifteen genes from the total differentially expressed genes were chosen for validation of microarrays and further study. Their putative identities were attributed by BLASTx and primers for qRT-PCR were designed according to Table 2.

Abbreviation	Identity (BLASTx)	Amplicon (bp)	Forward/Reverse qRT-PCR Primers
MT3	Metallothionein-like protein type 3 (MT-3) [<i>Carica papaya</i> (papaya)] 6e-26	108	5' TTCAAACCTCGATCACCATGT C/ TCAGTCTCCACGAAGTCAG C
GO	glycolate oxidase [<i>Mesembryanthemum crystallinum</i>] 0.0	123	5' CAGAAGATGGCTCATCCTG A/ AGGTCCAGTGAAGCAACT T
CAT	catalase [<i>Prunus persica</i>] 0.0	125	5' AGCCTGGAGAGAGATACCG A/ CCAGTATGAAATCCAGACG CT
PRK	Phosphoribulokinase, chloroplast precursor (PRK) [<i>Mesembryanthemum crystallinum</i>] 0.0	132	5' ACCTTCATGAGGAGGCTGA C/ GTCCAGGGAATGGTAGTCG T
MYB142	MYB transcription factor MYB142 [<i>Glycine max</i>] 3e-25	124	5' GGCATTGGCTCAATATGAC A/ GTGCTTGACGTCCTCGATTA
CO	zinc finger protein CONSTANS-LIKE 16, putative [<i>Arabidopsis thaliana</i>] 5e-26	125	5' CTACGGCGAAGCTGGAGT/ CCTCGTACCGGATCTTCTTG
HyPRP2	HyPRP2 [<i>Gossypium hirsutum</i>] 9e-24	121	5' ACGCACACAAACGCACTAA T/ AAGAAGAGGAGGTTGAGG CA

UNK1	unknown protein [<i>Arabidopsis thaliana</i>] 5e-11	120	5' GGGTGGTGGCTGTATTACC T/ AGACAGCTCGCCCAAGTC
UNK2	unknown protein [<i>Arabidopsis thaliana</i>] 3e-110	147	5' CATGCAATTCCTCGTGTTCC / CGTGGTATGACTCCAGCAT C
UNK3	hypothetical protein [<i>Nicotiana benthamiana</i>] 4e-13	109	5' GGCGAAGACGTA CTTTCTC A/ TCGGACTGGACTACAACAC AA
UDP-gluc 2	putative UDP-glucuronate decarboxylase 2 [<i>Nicotiana tabacum</i>] 1e-175	146	5' GATGATGGCCGTGTTGTTA G/ TGATCTCCATCCATCAGTCG
ELP	ELP (ECTOPIC DEPOSITION OF LIGNIN IN PITH); chitinase [<i>Arabidopsis thaliana</i>] 1e-20	131	5' GTCGCTTCTGGTGGCAGT/ CCTTGGTGCAACTTCTT G
CeSA3	cellulose synthase 3 [<i>Eucalyptus grandis</i>] 0.0	124	5' CAGTGACATCCAAAGCGTC T/ ATCCGGCAACTACTCCAA C
MYB	MYB transcription factor [<i>Eucalyptus gunnii</i>] 7e-150	107	5' CTGCACAGCCTTCTTG GTA A/ AGCTTCCTCCTTATGTGCGT
CCOAMT	caffeoyl-CoA 3-O-methyltransferase [<i>Linum usitatissimum</i>] 3e-15	134	5' TCCTTATGATCAGGGTTCAT TCT/ AAACAGCAACATAATTTGG CA

Table 2. List of selected genes according to the SAM analysis of microarray results with their respective abbreviations. The identity values attributed by BLASTx are represented by their E-values. Primer sequences designed through GenScript Corporation's Real-Time PCR Primer Design software are also represented, as well as the expected amplicon lengths for each sequence.

The candidate genes selected with the microarray analysis had their expression profiles evaluated *in silico* through the SAM graph analysis (Figure 2 and Supplemental Figure 1) and by qRT-PCR. Previously we evaluated the specificity of all primers designed for qRT-PCR by BLASTn against the *Genolyptus* EST sequence collection. Every qRT-PCR primer

showed only one match, and that was to its specific target-sequence (results not shown).

The melting-curve analysis obtained after 40 cycles of amplification was also used to validate the specificity of the primers. The dissociation curves allowed us to detect PCR unspecific products and primer dimers and also let us to optimize the concentration of primers to be used in each reaction. The reaction specificity was assumed as ideal when only one single amplification peak was observed. The occurrence of more than one peak, on the other hand, would reveal the presence of unspecific products, and would require optimizations. As results from the dissociation curve analysis, we did not observed amplification products for PRK or HyPRP2 and, for this reason, they were excluded from the analysis. Quantitative RT-PCRs resulted in more than one amplification product when primers designed for ELP and Zn-Finger sequences were evaluated, proving their unspecificity. Therefore, these two candidate genes were also excluded from the analysis. For all other 11 primer pairs, just one amplification peak was observed in the dissociation curve analysis, what evidenced their high specificity to the candidate genes selected under our qRT-PCR conditions (Supplemental Figure 2).

The expression level of the 11 candidate genes selected from the previous microarray analysis were determined in six tissue (two leaf and four xylem) samples from 4-year-old field-grown *E. grandis* trees. These samples were collected from the same two trees used in the microarray analysis added by two more identical clones. After performing the qRT-analysis, we compared the data obtained for the 11 selected genes with the data from the microarray analysis, as shown in Figure 3. The results were consistent with the microarray results for the genes analyzed, except for the sequence putatively encoding the MYB transcription factor.

A discussion concerning all 11 differentially expressed genes is presented as follows.

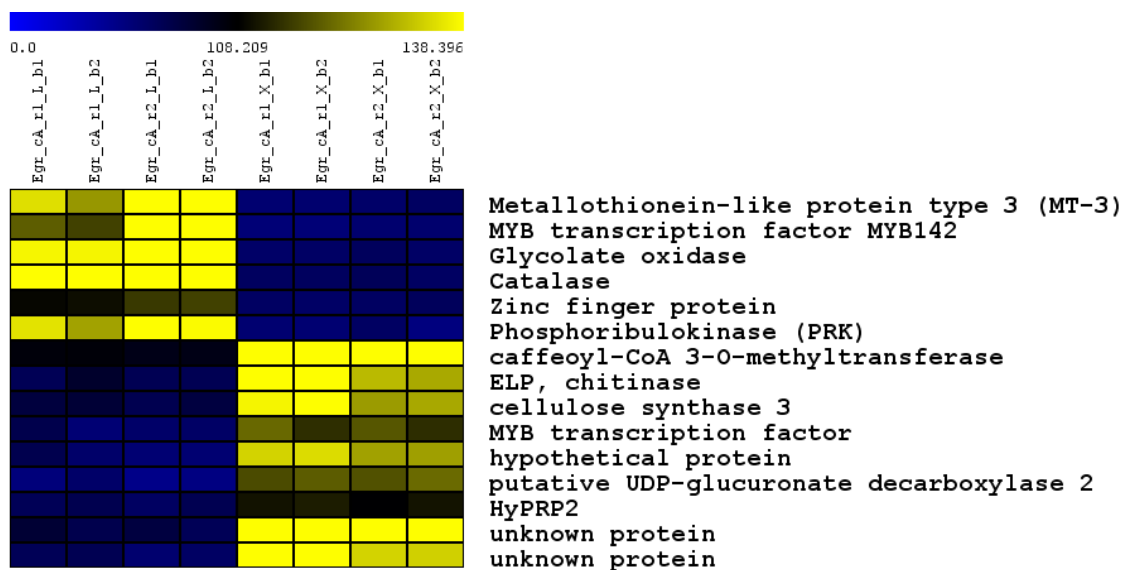


Figure 2. SAM algorithm-derived color expression pattern of 15 differentially expressed genes selected for further qRT-PCR analysis. Every line in this figure represents one single gene that has low (blue) to high (yellow) expression in leaves (left) and xylem (right). Egr_cA, *Eucalyptus grandis*; r1, clone 1; r2, clone 2; L, leaf; X, xylem; b1, block 1 (slide); b2, block 2.

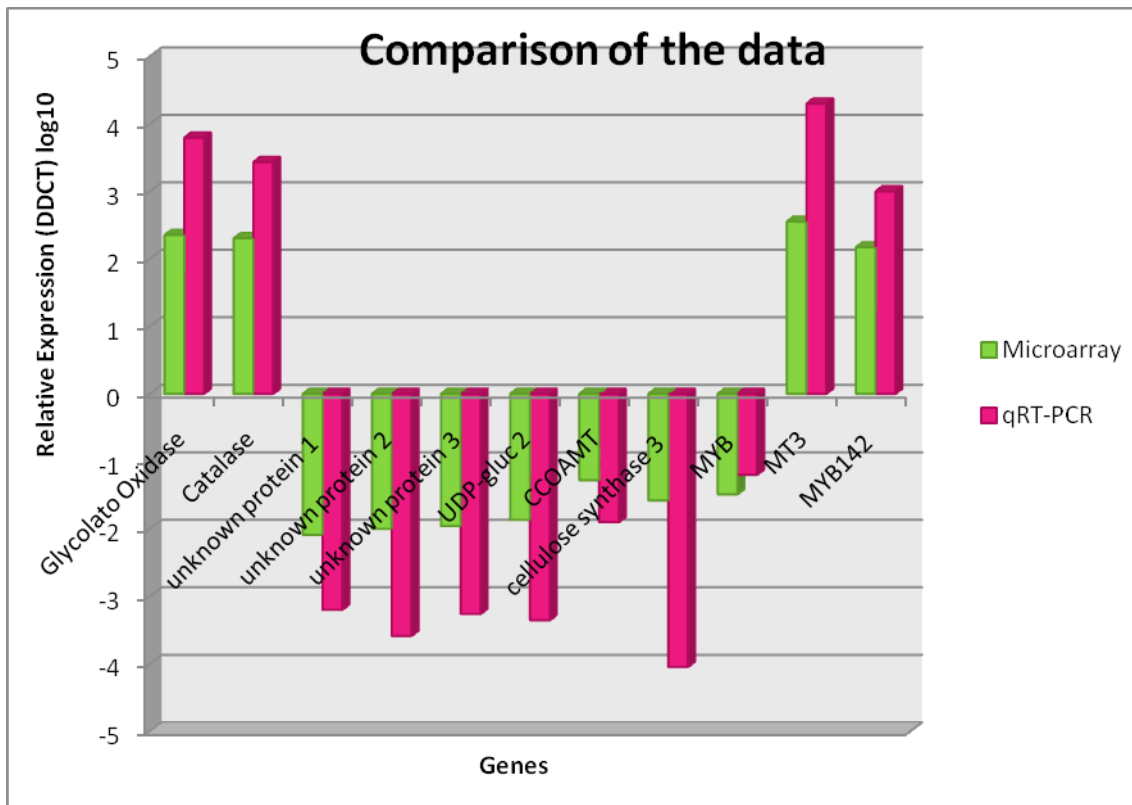


Figure 3. Comparison of data obtained for 11 selected genes by microarray and qRT-PCR analysis. The $2^{-\Delta\Delta CT}$ mean values and the microarray results for each gene are represented in a \log_{10} scale. To normalize the calculation we used the $2^{-\Delta\Delta CT}$ mean values from *E. grandis* xylem clones for the reference genes encoding ribonucleoprotein L23A, At2g28390 and EuC12 [22].

Genes up-regulated in the leaves

Type 3 metallothionein-like genes

Metallothioneins (MTs) are defined as low-molecular-weight (4–8 kDa) and cysteine (Cys)-rich proteins that can bind metals via thiol groups of their Cys residues [23]. The term MT-like is used for those putative MT sequences or proteins which protein-metal characterization was not established [24]. MT genes have been found in both animal and plant kingdoms, and the deduced amino acid sequences of plant MT-like

proteins were classified into four types in relation to the arrangement of Cys residues [25]. Plant MT-like genes are strongly believed to play a role in metal ion metabolism or detoxification and it has been demonstrated that *Arabidopsis* MT genes were induced by Cd, Cu and Zn treatment [26]. Besides detoxification, MTs could take part in regulation of gene expression and cell metabolism by donating/accepting Zn ion to/from Zn-dependent DNA binding proteins or metalloenzymes [27]. Thus, they could be involved in normal processes of growth and differentiation. MT has a putative role in redox homeostasis. It has been shown that yeast and mammalian MTs can functionally substitute superoxide-dismutases (SOD) and provide oxidative stress protection in yeast [28]. In addition, direct antioxidative role of MT as a scavenger of hydroxyl radicals or hydrogen peroxide was suggested [29]. Its up-regulation in *E. grandis* leaf blades when compared to xylem could be explained by the fact that the leaf is a photosynthetic organ more susceptible to the generation of reactive oxygen species (ROS). MT3 could also be involved in the protection from photo-oxidative stress in leaf. This presumption is supported by the fact that *A. thaliana* MT1 mRNA was up-regulated during light stress [30]. The high constitutive MT3 expression in buckwheat leaf was noticed previously [31]. Since MT is ubiquitous in virtually all living organisms and active within different tissues and organ systems of higher eukaryotes, proving its single primary role may not be possible. MTs constitutively expressed may play a background role in homeostatic mechanisms, whereas highly induced MT levels could be an adaptive response to various environmental stresses [31].

Catalase

Catalase is a central component of the enzymatic antioxidant system to control ROS and is found in a wide range of organisms from aerobic

bacteria to higher plants and animals [32, 33]. Plant catalases are encoded by a small gene family, usually three isoenzyme encoding genes in a single species, as previously described in maize [34], tobacco [35, 36], *Arabidopsis* [37], rice [38], and *E. grandis* [33]. This enzyme is a tetrameric iron porphyrin that catalyses the dismutation of hydrogen peroxide to water and oxygen and is preferentially present in peroxisomes. Photorespiration in C3 plants is an important source of hydrogen peroxide and this production is mainly counteracted by peroxisomal catalase, although other antioxidative enzymes are active in the leaf peroxisome [39].

Glycolate oxidase

Glycolate oxidase (GO), along with hydroxypyruvate reductase (HPR) and serine glyoxylate aminotransferase (SGAT) are members of the glycolate pathway. The glycolate pathway is a central part of the photorespiratory carbon metabolism in higher plants and it englobes the production of glycolate in chloroplasts during photosynthesis and its metabolism in peroxisomes [40]. These enzymes are localized in leaf peroxisomes [41] and in this organelle, GO is the first enzyme of the pathway, producing glyoxylate and hydrogen peroxide from glycolate and oxygen. GO is a flavoprotein which transfers electrons to molecular oxygen to form hydrogen peroxide. In addition to glycolate, this oxidase can oxidize L-lactate but not D-lactate, and it is insensitive to cyanide [42]. Glycolate is synthesized also by green algae [40]. It has been shown that the expression of photorespiratory enzymes such as GO and HPR is induced during greening in the light [43, 44]. Both the amount of mRNA and the efficiency of protein synthesis are increased during greening. In the study of Yamaguchi and Nishimura [45], transgenic tobacco plants with decreased GO activity by co-suppression were obtained. These

authors analyzed the relationship between GO amount and the degree of photoinhibition using those transformants, and determined the amount of GO that was required to maintain normal photosynthesis under different light conditions. Results suggested that photosynthesis was susceptible to photoinhibition with the reduction of GO activity levels to values below the threshold and that higher activities of GO are required under higher irradiation.

Phosphoribulokinase

Phosphoribulokinase (PRK) plays a critical role in regulating the flow of sugar through the Calvin cycle. PRK is unique to Calvin cycle, where it catalyses the irreversible formation of ribulose-1,5-bisphosphate using ribulose-5-phosphate as substrate. It is believed to be one of the key enzymes that are responsible for controlling the balance between assimilatory and regenerative stages [46]. It is encoded by the nuclear genome and synthesized in the cytosol as a precursor protein with an N-terminal transit peptide that directs the enzyme to the chloroplast. The activity of PRK is regulated by light, NADP/NADPH ratio, thioredoxin and Mg^{2+} ion levels [47, 48, 49]. PRK may also be regulated by the reversible dissociation of the PRK/CP12/GAPDH complex [50]. PRK expression is also regulated by light under the control of the circadian clock and down-regulated during senescence [48, 51]. *A. thaliana* PRK is encoded by a single gene and Calvin cycle PRK transcripts are especially abundant in *Arabidopsis* leaves [50]. That seems to be also the case in *E. grandis*.

Zinc finger protein CONSTANS-LIKE 16

Putterill *et al.* [52] isolated the CONSTANS (CO) gene of the model dicot *Arabidopsis*. CO has an important role in the photoperiod pathway which is one of four regulatory pathways controlling the timing of

flowering. It is known that the photoperiod pathway mediates light and temporal information from the environment into regulation of flowering time. Light is perceived by phytochromes (PHY) A to E and cryptochromes (CRY) 1 and 2, while the duration of day and night is measured by the circadian clock [53, 54]. CO acts between the circadian clock and genes controlling meristem identity [55, 56]. In *Arabidopsis*, CO belongs to a family of 17 putative transcription factors defined by two conserved domains [52, 57]. The first is a zinc finger region near the amino terminus that resembles B-boxes, which regulate protein-protein interactions in several animal transcription factors [58]. The second is a region of 43 amino acids near the C-terminus termed the CCT (CO, CO-like, TOC1) domain [57]. Studies using green fluorescent protein (GFP) fusions showed that the CCT domain is involved in nuclear localization of the CO protein. Previous analysis of CO-like genes in *Arabidopsis* showed that the family is subdivided into three broad groups. The first group comprised CO and COL1 to COL5 (two B-box genes), the second comprised COL6 to COL8 and COL16 (one B-box genes), and the third comprised COL9 to COL15 (one CO-like B-box and one more diverged zinc finger domain). *Arabidopsis* COL16 genes (Group II) have been shown to harbor a single intron located between the B-box and the CCT domains [57].

The overexpression of the gene coding for a CO protein in the leaf is consistent with the fact that leaf is an organ perceiving photoperiod, where the CO protein is responsible for sensing the day length signal.

MYB 142

MYB proteins are a superfamily of transcription factors that play regulatory roles in developmental processes and defense responses in plants. In the past decade, the R2R3-MYB genes have been extensively

studied. They were reported to be involved in many physiological and biochemical processes such as the regulation of secondary metabolism [59], the control of cell morphogenesis [60], the regulation of meristem formation and floral and seed development [61], and the control of cell cycle [62]. Some MYB proteins were also involved in various defense and stress responses [63] and in light and hormone signaling pathways [64, 65]. The work from Yanhui *et al.* [66] allowed the identification of 198 genes in the MYB superfamily from an analysis of the complete *Arabidopsis* genome sequence. Among them, 126 were R2R3-MYB, five were R1R2R3-MYB, 64 were MYB-related, and three were atypical MYB genes. These genes were shown to be expressed in a wide variety of tissues. In view of that, the occurrence of a MYB transcription factor in the leaf of *E. grandis* is predictable. In line with this information, Liang *et al.* [67] demonstrated the localization of a MYB transcription factor, AtMYB61, uniquely in guard cells of *Arabidopsis*. The results of this work support the hypothesis that AtMYB61, acting as a transcriptional regulator of stomatal function, is both necessary and sufficient to decrease stomatal aperture.

Genes up-regulated in the xylem

Putative UDP-glucuronate decarboxylase 2

UDP-glucuronate decarboxylase (UDP-GD) catalyses the conversion of UDP-glucuronate to UDP-xylose, which is an important sugar donor in the synthesis of hemicellulose and glycoproteins [68]. The UDP-GD genes have been reported in rat of the mammalian [69] and in typical dicot plants such as the pea [70] and *Arabidopsis* [71].

Extractability and recovery of cellulose from cell walls influences many industrial processes and also the utilization of biomass for energy purposes. UDP-GD produces UDP-xylose, the precursor for xylans and it

could affect the cellulose extractability when is down-regulated on cell wall structure in transgenic plants. Besides the influence of lignins on chemical processing, the importance of xylans was first realized by the pulping industry when it became apparent that delignification during the pulping process was influenced by the hemicelluloses present [72]. Bindschedler *et al.* [73] proved that the expression of the UDP-GD gene in an antisense orientation resulted in a consistent lower xylose and higher glucose content in the wall polysaccharides and promoted consequent changes to pulping properties. In addition, the pulp yield was lower and the pulp viscosity slightly higher due to the lower delignification. Concerning cellulose, the degree of polymerization was similar for all the pulps produced from control or transgenic lines. Even though hemicellulose mutants have been identified in *Arabidopsis*, this was the first example of the consequences of directed manipulation of xylans in woody plants to test for modifications in pulping properties.

MYB transcription factor

As mentioned before, plant MYB proteins were classified into three major groups: R2R3-MYB, R1R2R3-MYB, and a heterogeneous group collectively referred to as MYB-related proteins, which usually but not always contain a single MYB repeat [74-76]. The modulation of the expression of several MYB proteins leads to the regulation of the complex pathway of the phenylpropanoid metabolism and the accumulation of various end-products in specific cells, tissues or organs [77, 78], since several MYB proteins have been shown to be involved in the regulation of this metabolism [79, 80].

Many plant MYB transcription factors have been studied, like *Pinus taeda* PtMYB4, poplar PttMYB21a [81, 82], and *Eucalyptus gunnii* EgMYB2 [83]. A fourth subgroup was proposed and included AmMYB308

from *Antirrhinum majus*, ZmMYB31 and ZmMYB42 from maize, AtMYB32 and AtMYB4 from *Arabidopsis*, all of them involved in the repression of phenylpropanoid and/or monolignol biosynthesis genes [59, 84-86]. Fornale *et al.* reported in 2006 [86] the down-regulation of the maize and *A. thaliana* COMT gene by two new maize R2R3-MYB transcription factors, ZmMYB31 and ZmMYB42. The work also showed that ZmMYB31 and ZmMYB42 differentially regulate other structural genes of the lignin pathway, such as At4CL1, AtC4H, At4CCoAOMT1 and At4CAD6 resulting, in both cases, in a reduction of the total lignin content of the transgenic *A. thaliana* plants. Goicoechea *et al.* [83] reported the cloning and functional characterization of a R2R3MYB gene from *E. gunnii* named EgMYB2, which is able to bind the regulatory regions of genes encoding cinnamoyl-CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD2), the two most studied genes of the lignin biosynthetic pathway. EgMYB2 was described to be preferentially expressed in differentiating xylem tissue. The results reported strongly suggest that EgMYB2 is a positive regulator of secondary cell wall formation and lignin biosynthesis. On the other and, Legay *et al.* [87] reported the cloning and characterization of EgMYB1, a R2R3 MYB transcription factor preferentially expressed in *Eucalyptus* xylem tissues which binds, *in vitro*, the *cis*-regulatory regions of the *E. gunnii* CCR and CAD2. EgMYB1 was also described to act as a transcriptional repressor of these genes *in vivo*. EgMYB1 is preferentially expressed in the secondary xylem of stems and roots of *Eucalyptus* trees, and the corresponding protein is localized in the nucleus. EgMYB1 binds specifically the MBSIIG sites located in the promoters of lignin biosynthetic genes, and represses the transcription of these genes *in planta*. The work from Legay *et al.* [87] was the first to describe a MYB transcription factor highly and preferentially expressed in the xylem tissue of a woody angiosperm.

In *Eucalyptus* xylem, the presence of both a transcriptional activator (EgMYB2) and a repressor (EgMYB1) suggests that a balanced expression of these proteins competing for the same promoters could provide a mechanism for fine spatial and temporal control of lignin deposition, by forming either a repressing or an activating regulatory complex [87].

CCOAMT caffeoyl-CoA 3-O-methyltransferase

It is believed that two structurally distinct *O*-methyltransferases (OMT) are responsible for the methylation of lignin precursors. The caffeic acid 3-*O*-methyltransferase (COMT) methylates caffeic acid and 5-hydroxyferulic acid [88] whereas caffeoyl CoA 3-*O*-methyltransferase (CCoMT) methylates caffeoyl-CoA and 5-hydroxyferuloyl-CoA [89]. In dicotyledonous angiosperms, both COMT and CCoMT are bi-specific [90].

CCoAOMT plays an important role in lignin biosynthesis. This enzyme was initially characterized in cell suspensions of parsley and carrot challenged with a fungal elicitor [91, 92] and therefore was believed to play a role in disease resistance. Immunolocalization, tissue printing studies, and promoter β -glucuronidase (GUS) fusion analysis in various plant species have shown that the expression of the CCoAOMT gene is highly correlated with the lignifying tissues [93, 61]. CCoAOMT is encoded by two genes in poplar (*Populus trichocarpa*). In 2000, Chen *et al.* [94] described the expression pattern conferred by the two CCoAOMT promoters when fused to the *gusA*-coding sequence in transgenic poplar (*Populus tremula* X *Populus alba*). They demonstrated that both genes were expressed similarly in xylem and differently in phloem. The two CCoAOMT promoters directed reporter-gene expression in lignifying cells and in cells closely associated with lignifying cells, both during normal development and upon stress conditions known to influence lignin deposition (wounding, pathogen attack, and mechanical bending). Thus, it

is expected to find an increased CCoAOMT gene expression in *E. grandis* xylem tissue when comparing it to leaf blades.

Hybrid Proline-Rich Proteins

Hybrid proline-rich proteins (HyPRPs) are proteins that contain a signal peptide followed by an N-terminal proline-rich domain that presents different repetitive motives and a second hydrophobic C-terminal domain [95]. Such domain does not contain repetitive elements but presents a series of invariable Cys residues in a conserved distribution pattern. These proteins have been named either hybrid proline-rich proteins or hybrid glycine-rich proteins (HyGRPs). The HyPRPs may have different repetition patterns in their Pro-rich domain. However, when the second hydrophobic domain is analyzed, the HyPRPs can be subdivided in two different groups (A and B), according to the Cys residue positions in the peptide sequence [95].

The expression pattern of plant genes encoding HyPRPs is quite diverse. In rice, it has been shown that the expression of a gene which encodes a HyPRP occurs in the vascular and cortical tissues of the root cap and the elongation zone of the root. Though, the expression in leaves occurs mainly in vascular and mesophyll cells [96]. In contrast, it has been observed in maize that the expression of a gene that encodes for HyPRP is restricted to developing cortical cells. Blanco-Portales *et al.* [97] have cloned and characterized a strawberry HyPRP gene (Fahyprp). The gene expression analysis, *in situ* hybridization, immunolocalization and histochemical experiments suggested that the Fahyprp gene is related to the fruit ripening process and possibly involved in the anchoring of polyphenols (lignin and condensed tannins) to cell membranes. A role in transporting molecules from or towards the vascular tissues has been proposed for this protein [98]. It has been described that Pro-rich proteins

and hydroxy-Pro-rich proteins are preferentially located in xylem cells under a process of lignification. A putative role for these proteins as anchoring sites for lignin polymer through the interaction with the phenolic groups constituting the lignin was proposed [99]. Then, the overexpression of a gene encoding a HyPRP in the *E. grandis* xylem would it be reasonable. Even though different physiological functions have been suggested for HyPRP proteins in higher plants, including their involvement in developmental processes and responses to different stress, so far no specific function for this class of proteins has been concluded.

Cellulose Synthase 3

Cellulose is a linear homopolymer of $\beta(1-4)$ -linked glucose residues. The coordinated synthesis of glucose chains is orchestrated by specific plasma membrane-bound cellulose synthase complexes (CelS). The CelS is postulated to be composed of approximately 36 cellulose synthase (CESA) subunits. The CelS synthesizes 36 glucose chains in close proximity before they are further organized into microfibrils that are associated with other cell wall polymers. The 36 glucose chains in a microfibril are stabilized by intra- and inter-hydrogen bonding which confer great stability on microfibrils. Several elementary microfibrils come together to form macrofibrils. Many CESA isoforms appear to be involved in the cellulose biosynthetic process and at least three types of CESA isoforms appear to be necessary for the functional organization of CelS in higher plants [100]. Persson *et al.*, [101] presented evidence that the primary CelS complex in *Arabidopsis* requires three unique types of components, CESA1-, CESA3-, and CESA6-related, for activity. The work in *Arabidopsis* also showed that the CESA6-related CESAs are partially functionally redundant. This requirement is similar to the secondary CESA complex, in which CESA4, -7, and -8 are necessary [102].

Recently, six cDNAs encoding CESAs of *E. grandis* were reported and named *EgCesA1* to 6 [103]. Experiments of qRT-PCR indicated that transcripts of *EgCesA1* to 3 were abundant in tissues enriched with cells laying down secondary cell walls (xylem) but were weakly expressed in tissues undergoing primary growth (unfolding leaves). Expression of *EgCesA4* and *EgCesA5* was up-regulated in tissues rich in rapidly dividing cells undergoing primary wall synthesis, whereas *EgCesA6* was weakly expressed in all tissues analyzed. These results suggest that *Eucalyptus*, like other higher plants, expresses two contrasting groups of apparently co-regulated CESAs involved in either primary or secondary cell wall biosynthesis.

Chitinase and Ectopic Deposition of Lignin in Pith

Plant chitinases are a group of enzymes that presumably hydrolyze chitin, a biopolymer of *N*-acetylglucosamine (GlcNAc) in a β -1,4 linkage. Only a few of them have been proved to possess chitinase activities. Chitinases that are expressed in a tissue-specific manner and are not induced by pathogens or stress may have roles in normal plant growth and development. Most of the reported chitinases have been identified based on their sequence similarity to known chitinases. Plant chitinases are grouped into six different classes based on sequence similarity to tobacco chitinases [104]. The two most common classes of chitinases are class I and class II, which differ in the presence (class I) or absence (class II) of a conserved N-terminal Cys-rich lectin domain. All classes of chitinases possess some conserved amino acid residues in the catalytic domain [105]. In 2000, Zhong *et al.* [106] isolated and characterized an *Arabidopsis* ectopic deposition of lignin in pith (*elp1*) mutant with alteration of lignin deposition patterns. In this *elp1* mutant stems, lignin is deposited ectopically in patches of parenchyma cells in the pith that are

not lignified in the wild-type. The ectopic deposition of lignin in the pith of the *elp1* mutant was identified by histological staining of lignin, and it correlated with increased klason lignin content and ectopic expression of CCoAOMT, an enzyme involved in monolignol biosynthesis. In another article from 2002 [107], the same group showed that the *elp1* gene encodes a chitinase-like gene in *Arabidopsis* (*AtCTL1*) and that some pith cells in the *elp1* mutant had aberrant shapes of cells with incomplete cell walls. The mutant also overproduced ethylene, causing growth alterations in both dark- and light-grown seedlings. With these results, the authors provided direct evidence that this chitinase-like protein plays a role in normal plant growth and development in *Arabidopsis*. Such possibility may be the case for the observed chitinase-like gene in *E. grandis*.

Correlation between qRT-PCR and microarray experiments

Point out a handful of potentially interesting genes with microarray experiments and confirm those candidates by realtime RT-PCR analysis is a widely used strategy. There are several works showing the correlation between microarray experiments and quantitative RT-PCR analysis. Gachon *et al.* [108] reviewed the relevance of real-time PCR in plant studies. As an illustration, they cited the case of Rider *et al.* [109] that randomly picked 12 candidates among a pool of 185 genes previously identified by AffimetrixTM microarray experiments as being up-regulated in the seeds of the *Arabidopsis* pickel mutant compared with wild-type seeds. They confirmed the up-regulation for only 10 of those genes by real-time RT-PCR experiments.

The data from Dallas *et al.* [110] demonstrate that similar microarray scores for different genes do not necessarily mean that similar qRT-PCR scores will be obtained. For example, they discovered genes

which yielded similar average log₂ RMA scores on the microarray analysis but the average log₂ qRT-PCR scores for the same genes were substantially different. The finding that genes with similar microarray expression scores were unlikely to have similar qRT-PCR results presumably reflects the different hybridization kinetics of the probe sets for each gene. The results from this work indicates that the direction of change of gene expression levels (i.e. either up or down regulation) between subsets of interest is accurately predicted by comparison of average microarray expression scores, although the fact that they found 13–16% of non-concordance between the two techniques.

Conclusion

Our results obtained by qRT-PCR were consistent with the microarray results for the majority of the genes analyzed, although the relative expression ratios of the first were often higher. Similar discrepancies between the methods were reported previously, suggesting that the ratios calculated using DNA arrays are often underestimated [111, 10]. We believe that our microarray analysis and biological samples were fully supported by the qRT-PCR findings: it was robust enough to evaluate a massive number of genes and to point out important candidate genes for further analysis.

Methods

Plant Material and Tissue Harvesting

Mature leaves and stem scrapings were harvested from 4-year-old field-grown *E. grandis* trees (Hortoflorestal Barbanegra, Aracruz Celulose S.A., Barra do Ribeiro, RS, Brazil). In order to minimize the proportion of primary xylem mainly located in the main veins of leaves,

only leaf blades without the central vein of young and mature leaves were used for this study. To obtain the differentiating xylem tissue, we proceeded as described in Paux *et al.* [10]: the bark was removed and scrapings were taken from the exposed xylogenic tissue to avoid contamination with phloem tissue. Tissue samples were immediately frozen in liquid nitrogen after harvesting and stored at -80 °C. For the microarray experiments, leaf and xylem samples were harvested from two clones derived from one matrix. For qRT-PCR, mature leaves and xylem scrapings were collected from three clones of one *E. grandis* matrix.

Total RNA Extraction

Total RNA was extracted from 1 g of either xylem or leaves using PureLink Plant RNA Reagent Kit (Invitrogen) according to the instructions of the manufacturer. RNA purity and quality were checked by both agarose gel electrophoresis and spectrophotometry measurements at 260 and 280 nm. Polyadenylated RNA was subsequently purified from total RNA using the Oligotex mRNA Mini kit (Qiagen). The concentration of isolated mRNA was determined using spectroscopic measurements at 260 nm.

Microarray

Microarray experiments were carried out by NimbleGen Systems for 21,442 of the 21,905 ESTs of clusters and singletons proceeding from the Genolyptus Sequencing Project. Nine oligonucleotides of 50 bp were designed and synthesized to each sequence consensus. The oligonucleotides were distributed throughout the sequence and had close melting temperatures (T_m). The nine probes were duplicated adding up 385,856 features per chip and a total of 4 identical chips

were produced according to the company protocols. Complementary DNAs (cDNAs) labeled with Cy3 were synthesized and hybridizations, washing, scanning, data collection and initial data normalization were performed according to NimbleGen protocols.

Microarray Data Analysis

Gene expression data were analyzed by NimbleGen following quartile normalization and Robust Multi-Array Analysis (RMA) which consists of the following stages: elimination of background fluorescence, normalization and data transformation [112].

Statistical Analysis

Data were analyzed using the Significance Analysis of Microarrays (SAM) procedure, a validated statistical technique for identifying differentially expressed genes using microarrays [113] available on the Institute for Genomic Research (TIGR) [114].

Gene Selection for Validation

Candidate genes were selected according to the microarray analysis using SAM with the following parameters: number of permutations, 70; SO parameters, Tusher *et al.* [113] method; q values, not calculated; imputation method, K-nearest neighbors imputer; number of neighbors, 10; delta value, 9.15, and false discovery rate 0%. We also analyzed the sequences by BLASTx and discarded the without hits on X database. We therefore, sorted only sequences with higher homology and primers were projected in the most conserved regions.

Primer Design

Primers were designed using the software created by GenScript Corporation's Real-Time PCR Primer Design [115]. The primers were designed in order to be as specific as possible for the selected genes. This was done by grouping and comparing the sequences and then designing the primers on regions with major similarity. Only those that resulted in amplicons with more than 100 pb and optimal T_m at 60°C had been selected. Initially all primers were tested by qRT-PCR and amplification specificity was verified by dissociation curves analysis using ABI Prism 7500 Dissociation Curve Analysis Software (Applied Biosystems). Primer sequences are given in Table 2.

Reverse Transcription

Each mRNA sample (300 ng) was reverse transcribed to cDNA with 200 units of M-MLV RT (Promega) and their respective reagents using 10 pmol of oligo-dT_{30(V)} (SMART CDS II A, RW Genes/Clontech), 0.5 mM each deoxynucleoside triphosphate (dNTPs) in a final volume of 10 µl. Reactions were initially incubated at 70 °C for 5 min followed by 40 °C for 60 min.

qRT-PCR

The cDNAs were diluted 1:50 with nuclease-free water. Aliquots of the same cDNA sample were used with all primers set for qRT-PCR. Each PCR was performed in quadruplicate and no-template controls were included. Quantitative RT-PCRs were performed in the ABI Prism 7500 Sequence Detection System (Applied Biosystems) in a 96-well reaction plate using the parameters recommended by the manufacturer (5 min. at 94 °C; 40 cycles of 94 °C for 15 sec, 60 °C for 10 sec, 72 °C for 15 sec and 60 °C for 35 sec). Reactions were done in 20 µl volumes containing 4 pmol of each primer, 0.5 mM of dNTPs,

0.05 μ L of Platinum Taq (Invitrogen) and its reagents, 1X SYBR Green (Applied Biosystems) and 10 μ L of 1:50 cDNAs. Specificity of the amplifications was verified by melting curve analysis at the end of the PCR run using the ABI Prism Dissociation Curve Analysis Software and the threshold cycle (C_T) was determined for each primer using the ABI Prism 7500 Sequence Detection System (Applied Biosystems).

Analysis of Gene Expression

To analyze gene expression we used the $2^{-\Delta\Delta CT}$ method which is used to calculate relative changes in gene expression determined from qRT-PCR experiments [116]. In this manner, the degree of relative expression was calculated from the efficiencies of the reactions and the deviation of the C_T of the sample-target versus the control. As reference genes, we used primers specific for the ribonucleoprotein L23A, At2g28390 and EuC12 genes [22].

References

1. ABRAF: **Anuário estatístico da ABRAF: ano base 2006.** [<http://www.abraflor.org.br/estatisticas/anuario-ABRAF-2007.pdf>]
2. BRACELPA: **Associação Brasileira de Celulose e Papel: Relatório anual de desempenho 2006** [<http://www.bracelpa.org.br/bra/estatisticas/>]
3. Neilson D: **The global *Eucalyptus* resource and some solid wood-panel product developmental issues.** IUFRO Conference 2000, Launceston, Austrália.
4. 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 Phytol* 2006, **170**:739–752.

5. **FAO: Food and Agriculture Organization: Forest products and services.** [<http://www.fao.org/forestry/index.jsp>]
6. Grattapaglia D, Bradshaw Jr HD: **Nuclear DNA content of commercially important *Eucalyptus* species and hybrids.** *Can J For Res* 1994, **24**:1074–1078.
7. Myburg AA, Griffin AR, Sederoff RR, Whetten RW: **Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach.** *Theor Appl Genet* 2003, **107**:1028–1042.
8. Paux E, Carocha V, Marques C, Sousa AM, Borralho N, Sivadon P, Grima-Pettenati J: **Transcript profiling of *Eucalyptus* xylem genes during tension wood formation.** *New Phytol* 2005, **167**:89–100.
9. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**(5235):467–470.
10. Paux E, Tamasloukht M, Ladouce N, Sivadon P, Grima-Pettenati J: **Identification of genes preferentially expressed during wood formation in *Eucalyptus*.** *Plant Mol Biol* 2004, **55**:263–280.
11. Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlén M, Teeri TT, Lundberg J, Sundberg B, Nilsson P, Sandberg G: **A transcriptional roadmap to wood formation.** *P Natl Acad Sci USA* 2001, **98**(25):14732–14737.
12. Israelsson M, Eriksson ME, 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 Mol Biol* 2003, **52**:893–903.
13. Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M, Sandberga G: **A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity.** *The Plant Cell* 2004, **16**:2278–2292.

14. Aspeborg H, Schrader J, Coutinho PM, 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 TT: **Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen.** *Plant Physiol* 2005, **137**:983–997.
15. Prassinos C, Ko J, Yang J, Han K: **Transcriptome profiling of vertical stem segments provides insights into the genetic regulation of secondary growth in hybrid aspen trees.** *Plant Cell Physiol* 2005, **46**(8):1213–1225.
16. Andersson-Gunneras S, Mellerowicz EJ, Love J, Segerman B, Ohmiya Y, Coutinho PM, 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 J* 2006, **45**:144–165.
17. Allona I, Quinn M, Shoop E, Swope K, St.Cyr S, Carlis J, Riedl J, Retzel E, Campbell MM, Sederoff R, Whetten RW: **Analysis of xylem formation in pine by cDNA sequencing.** *P Natl Acad Sci USA* 1998, **95**:9693-9698.
18. Whetten R, Sun Y, Zhang Y, Sederoff R: **Functional genomics and cell wall biosynthesis in loblolly pine.** *Plant Mol Biol* 2001, **47**:275–291.
19. Lorenz WW & Dean JFD: **SAGE profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*).** *Tree Physiol* 2002, **22**:301–310.
20. Yang S, Zyl L, No E, Loopstra CA: **Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*).** *Plant Sci* 2004, **166**:1185–1195.
21. Yang J, Park S, Kamdem DP, Keathley DE, Retzel E, Paule C, Kapur V, Han K: **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 Mol Biol* 2003, **52**:935–956.

22. Bastolla FM: **Seleção e avaliação de genes de referência para estudos de expressão gênica em *Eucalyptus***. MSc thesis. Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia; 2007.
23. Kägi JH, Schaffer A: **Biochemistry of metallothionein**. *Biochemistry* 1988, **27**:8509–8515.
24. Rauser EW: **Structure and function of metal chelators produced by plants**. *Cell Biochem Biophys* 1999, **31**:19–48.
25. Cobbett C, Goldsbrough P: **Phytochelatin and metallothioneins: roles in heavy metal detoxification and homeostasis**. *Annu. Rev Plant Biol* 2002, **53**:159–182.
26. Zhou J, Goldsbrough PB: **Structure, organization and expression of the metallothionein gene family in *Arabidopsis***. *Mol Gen Genet* 1995, **248**:318–328.
27. Vallee BL: **The function of metallothionein**. *Neurochem Int* 1995, **27**:23–33.
28. Tamai KT, Gralla EB, Ellerby LM, Valentine JS, Thiele DJ: **Yeast and mammalian metallothioneins functionally substitute for yeast copper-zinc superoxide dismutase**. *Proc Natl Acad Sci* 1993, **90**:8013–8017.
29. Quesada AR, Byrnes RW, Krezoski SO, Petering DH: **Direct reaction of H₂O₂ with sulfhydryl groups in HL-60 cells: Zn-metallothionein and other sites**. *Arch Biochem Biophys* 1996, **334**(2):241–250.
30. Dunaeva M, Adamska I: **Identification of genes expressed in response to light stress in leaves of *Arabidopsis thaliana* using RNA differential display**. *Eur J Biochem* 2001, **268**:5521–5529.
31. Brkljačić JM, Samardžić JT, Timotijević GS, Maksimović VR: **Expression analysis of buckwheat (*Fagopyrum esculentum* Moench) Metallothionein-like gene (MT3) under different stress and physiological conditions**. *J. Plant Physiol* 2004, **161**:741–746.

32. Vandenberghe S, Vanderauwera S, Vuylsteke M, Rombauts S, Langebartels C, Seidlitz HK, Zabeau M, Van Montagu M, Inzé D, Van Breusegem F: **Catalase deficiency drastically affects gene expression induced by high light in *Arabidopsis thaliana*.** *The Plant J* 2004, **39**:45-58.
33. Teixeira FK, Menezes-Benavente L, Galvão VC, Margis-Pinheiro M: **Multigene families encode the major enzymes of antioxidant metabolism in *Eucalyptus grandis* L.** *Gen and Mol Biol* 2005, **28**:529-538.
34. Scandalios JG: **Subunit dissociation and recombination of catalase isozymes.** *Proc Natl Acad Sci USA* 1965, **53**:1035-1040.
35. Havir AE, McHale NA: **Biochemical and developmental characterization of multiple forms of catalase in tobacco leaves.** *Plant Physiol* 1987, **84**:450-455.
36. Willekens H, Langebartels C, Tiré C, Van Montagu M, Inzé D, Van Camp W: **Differential expression of catalase genes in *Nicotiana plumbaginifolia* (L.).** *Proc Natl Acad Sci USA* 1994, **91**:10450-10454.
37. Frugoli JA, Zhong HN, Nuccio ML, McCourt P, McPeck MA, Thomas TL, McClung CR: **Catalase is encoded by a small multigene family in *Arabidopsis thaliana* (L.) Heynh.** *Plant Physiol* 1996, **112**:327-336.
38. Iwamoto M, Maekawa M, Sato A, Higo H, Higo K: **Evolutionary relationship of plant catalase genes inferred from exon-intron structures: Isozymes divergence after the separation of monocots and dicots.** *Theor App Genet* 1998, **97**:9-19.
39. Corpas FJ, Barroso JB, del Río LA: **Peroxisomes as a source of reactive oxygen species and nitric oxide signal molecules in plant cells.** *Trends Plant Sci* 2000, **6**:145-150.
40. Tolbert NE: **Microbodies-peroxisomes and glyoxysomes.** *Annu. Rev. Plant Physiol* 1971, **22**:45-74.
41. Tolbert NE, Oeser A, Kisaki T, Hageman RH, Yamazaki RK: **Peroxisomes from spinach leaves containing enzymes related to glycolate metabolism.** *J. Biol. Chem* 1968, **243**:5179-5184.

42. Frederick SE, Gruber PJ, Tolbert NE: **The Occurrence of glycolate dehydrogenase and glycolate oxidase in green plants.** *Plant Physiol* 1973, **52**:318-323.
43. Mano S, Hayashi M, Nishimura M: **Light regulates alternative splicing of hydroxypyruvate reductase in pumpkin.** *Plant J* 1999, **17**:309-320.
44. Tsugeki R, Hara-Nishimura I, Mori H, Nishimura M: **Cloning and sequencing of cDNA for glycolate oxidase from pumpkin cotyledons and northern blot analysis.** *Plant Cell Physiol* 1993, **34**:51-57.
45. Yamaguchi K, Nishimura M: **Reduction to below threshold levels of glycolate oxidase activities in transgenic tobacco enhances photoinhibition during irradiation.** *Plant Cell Physiol* 2000, **41**(12):1397-1406.
46. Raines CA, Lloyd JC, Dyer T: **New insights into the structure and function of sedoheptulose-1,7-bisphosphatase; an important but neglected Calvin cycle enzyme.** *J Exp Bot* 1999, **50**:1-8.
47. Brandes HK, Larimer FW, Hartman FC: **The molecular pathway for the regulation of phosphoribulokinase by thioredoxin F.** *J Biol Chem* 1996, **271**:3333-3335.
48. Raines CA, Longstaff M, Lloyd JC, Dyer TA: **Complete coding sequence of phosphoribulokinase: developmental and lightdependent expression of the mRNA.** *Mol Gen Genet* 1989, **220**:43-48.
49. Wedel N, Soll J: **Evolutionarily conserved light regulation of Calvin cycle activity by NADPH-mediated reversible phosphoribulokinase/ CP12/ glyceraldehyde-3-phosphate dehydrogenase complex dissociation.** *Proc Natl Acad Sci USA* 1998, **95**:9699-9704.
50. Marri L, Sparla F, Pupillo P, Trost P: **Co-ordinated gene expression of photosynthetic glyceraldehyde-3-phosphate dehydrogenase, phosphoribulokinase, and CP12 in *Arabidopsis thaliana*.** *J Exp Bot* 2005, **56**(409):73-80.
51. Lemaire SD, Stein M, Issakidis-Bourguette M, Keryer E, Benoit V, Pineau B, Gerard-Hirne C, Miginiac-Maslow M, Jacquot JP: **The**

- complex regulation of ferredoxin/thioredoxin-related genes by light and the circadian clock.** *Planta* 1999, **209**:221–229.
52. Putterill J, Robson F, Lee K, Simon R, Coupland G: **The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors.** *Cell* 1995, **80**:847–858.
53. Mouradov A, Cremer F, Coupland G: **Control of flowering time: interacting pathways as a basis for diversity.** *Plant Cell* 2002, Suppl S111–S130.
54. Simpson GG, Dean C: ***Arabidopsis*, the rosetta stone of flowering time.** *Science* 2002, **296**:285–289.
55. Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G: **Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*.** *Science* 2000, **288**:1613–1616.
56. Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G: **CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*.** *Nature* 2001, **410**:1116–1119.
57. Robson F, Costa MMR, Hepworth SR, Vizir I, Pineiro M, Reeves PH, Putterill J, Coupland G: **Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants.** *Plant J* 2001, **28**:619–631.
58. Borden KLB: **RING fingers and B-boxes: Zinc-binding protein-protein interaction domains.** *Biochem Cell Biol* 1998, **76**:351–358.
59. Tamagnone L, Merida A, Parr A, Mackay S, Culianez-Macia FA, Roberts K, Martin C: **The AmMYB308 and AmMYB330 transcription factors from *antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco.** *Plant Cell* 1998, **10**:135–154.
60. Payne CT, Clement J, Arnold D, Lloyd A: **Heterologous MYB genes distinct from GL1 enhance trichome production when**

- overexpressed in *Nicotiana tabacum*. *Development* 1999, 126:671–682.**
61. Li L, Osakabe Y, Joshi CP, Chiang VL: **Secondary xylem-specific expression of caffeoyl-coenzyme A 3-Omethyltransferase plays an important role in the methylation pathway associated with lignin biosynthesis in loblolly pine.** *Plant Mol Biol* 1999, **40**:555–565.
62. Araki S, Ito M, Soyano T, Nishihama R, Machida Y: **Mitotic cyclins stimulate the activity of c-Myb-like factors for transactivation of G2/M phase-specific genes in tobacco.** *J Biol Chem* 2004, **279**:32979–88.
63. Nagaoka S, Takano T: **Salt tolerance-related protein STO binds to a Myb transcription factor homologue and confers salt tolerance in *Arabidopsis*.** *J Exp Bot* 2003, **54**:2231–2237.
64. Wang ZY, Kenigsbuch D, Sun L, Harel E, Ong MS, Tobin EM: **A Myb-related transcription factor is involved in the phytochrome regulation of an *Arabidopsis* Lhcb gene.** *Plant Cell* 1997, **9**:491–507.
65. Newman LJ, Perazza DE, Juda L, Campbell MM: **Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the det3 mutant phenotype.** *Plant J* 2004, **37**:239–50.
66. Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q: **The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family.** *Plant Mol Biol* 2006, **60**:107–124.
67. Liang Y-K, Dubos C, Dodd IC, Holroyd GH, Hetherington AM, Campbell MM: **AtMYB61, an R2R3-MYB transcription factor controlling stomatal aperture in *Arabidopsis thaliana*.** *Curr Biol* 2005, **15**:1201–1206.
68. Suzuki K, Suzuki Y, Kitamura S: **Cloning and expression of a UDP-glucuronic acid decarboxylase gene in rice.** *J Exp Bot* 2003, **54**(389):1997-1999.

69. Moriarity JL, Hurt JH, Resnick AC, Storm PB, Laroy W, Schnaar RL, Snyder SH: **UDP-glucuronate decarboxylase, a key enzyme in proteoglycan synthesis.** *J Biol Chem* 2002, **277**:16968-16975.
70. Kobayashi M, Nakagawa H, Suda I, Miyagawa I, Match T: **Purification of UDP-D-glucuronate carboxy-lyase (UDP-D-xylose synthase) from pea seedlings.** *Plant Cell Physiol* 2002, **43**:1259-1265.
71. Harper AD, 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 Physiol* 2002, **130**:2188-2198.
72. Viikari L, Kantelinen A, Sundquist J, Linko M: **Xylanases in bleaching - from an idea to the industry.** *FEMS Microbiol Rev* 1994, **13**:335-350.
73. Bindschedler LV, Tuerck J, Maunders M, Ruel K, Petit-Conil M, Danoun S, Boudet AM, Joseleau JP, Bolwell GP: **Modification of hemicellulose content by antisense down-regulation of UDP-glucuronate decarboxylase in tobacco and its consequences for cellulose extractability.** *Phytochemistry* 2007, **68**:2635-2648.
74. Rosinski JA, Atchley WR: **Molecular evolution of the Myb family of transcription factors: evidence for polyphyletic origin.** *J Mol Evol* 1998, **46**:74-83.
75. Jin H, Martin C: **Multifunctionality and diversity within the plant MYB-gene family.** *Plant Mol Biol* 1999, **41**:577-585.
76. Stracke R, Werber M, Weisshaar B: **The R2R3- MYB gene family in *Arabidopsis thaliana*.** *Curr Opin Plant Biol* 2001, **4**:447-456.
77. Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B: **Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes.** *Plant Mol Biol* 2005, **57**:155-171.
78. Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B: **Differential regulation of closely related R2R3-**

- MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling.** *Plant J* 2007, **50**:60–677.
79. Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C: **Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis.** *Plant Cell* 2000, **12**:2383–2394.
80. Deluc L, Barrieu F, Marchive C, Lauvergeat V, Decendit A, Richard T, Carde JP, Merillon JM, Hamdi S: **Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway.** *Plant Physiol* 2006, **140**:499–511.
81. Patzlaff A, McInnis S, Courtenay A, Surman C, Newman LJ, Smith C, Bevan MW, Mansfield S, Whetten RW, Sederoff RR, Campbell MM: **Characterisation of a pine MYB that regulates lignification.** *Plant J* 2003, **36**(6):743–754.
82. Karpinska B, Karlsson M, Srivastava M, Stenberg A, Schrader J, Sterky F, Bhalerao R, Wingsle G: **MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen.** *Plant Mol Biol* 2004, **56**(2):255–270.
83. Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B, Verhaegen D, Chaubet-Gigot N, Grima-Pettenati J: **EgMYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis.** *The Plant J* 2005, **43**:553–567.
84. Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C: **Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*.** *EMBO J* 2000, **19**:6150–6161.
85. Preston J, Wheeler J, Heazlewood J, Li SF, Parish RW: **AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*.** *Plant J* 2004, **40**:979–995.
86. Fornale S, Sonbol FM, Maes T, Capellades M, Puigdomènech P, Rigau J, Caparrós-Ruiz D: **Down-regulation of the maize and *Arabidopsis thaliana* caffeic acid O-methyl-transferase genes by two new maize R2R3-MYB transcription factors.** *Plant Mol Biol* 2006, **62**:809–823.

87. Legay S, Lacombe E, Goicoechea M, Brière C, Séguin S, Mackay J, Grima-Pettenati J: **Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway.** *Plant Sci* 2007, **173**:542–549.
88. Zubieta C, Kota P, Ferrer J, Dixon RA, Noel JP: **Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase.** *Plant Cell* 2002, **14**:1265-1277.
89. Ye Z-H, Kneusel RE, Matern U, Varner JE: **An alternation methylation pathway in lignin biosynthesis in *Zinnia*.** *Plant Cell* 1994, **6**:1427-1439.
90. Ma QH, Xu Y: **Characterization of a caffeic acid 3-O-methyltransferase from wheat and its function in lignin biosynthesis.** *Biochimie* 2007, **xx**:1e10.
91. Pakusch A-E, Kneusel RE, Matern U: **S-Adenosyl-l-methionine: Trans-caffeoyl-coenzyme A 3-O-methyltransferase from elicitor-treated parsley cell suspension cultures.** *Arch Biochem Biophys* 1989, **271**:488–494.
92. Kühnl T, Koch U, Heller W, Wellmann E: **Elicitor induced S-adenosyl-l-methionine:caffeoyl-CoA 3-O-methyltransferase from carrot cell suspension cultures.** *Plant Sci* 1989, **60**:21–25.
93. Ye Z-H: **Association of caffeoyl coenzyme A 3-O-methyltransferase expression with lignifying tissues in several dicot plants.** *Plant Physiol* 1997, **115**:1341–1350.
94. Chen C, Meyermans H, Burggraeve B, De Rycke RM, Inoue K, De Vleeschauwer V, Steenackers M, Van Montagu MC, Engler GJ, Boerjan WA: **Cell-specific and conditional expression of caffeoyl-coenzyme A-3-O-methyltransferase in *Poplar*.** *Plant Physiol* 2000, **123**:853–867.
95. José M, Puigdomènech, P: **Hybrid-Proline rich and related proteins.** *Current Topic in Mol. Genet* 1994, **2**:109–126.
96. Xu Y, Buchholz WG, DeRose RT, Hall TC: **Characterization of a rice gene family encoding root-specific proteins.** *Plant Mol Biol* 1995, **27**:237–248.

97. Blanco-Portales R, López-Raé JA, Bellido ML, Moyano E, Dorado G, González-Reyes JA, Caballero JL, Muñoz-Blanco J: **A strawberry fruit-specific and ripening-related gene codes for a HyPRP protein involved in polyphenol anchoring.** *Plant Mol Biol* 2004, **55**:763–780.
98. John I, Wang H, Held BM, Wurtele E, Colbert JT: **An mRNA that specifically accumulates in maize roots delineates a novel subset of developing cortical cells.** *Plant Mol Biol* 1992, **20**:821–831.
99. Casab GJ: **Plant cell wall proteins.** *Ann. Rev. Physiol. Plant Mol Biol* 1998, **49**:281–309.
100. Festucci-Buselli RA, Otoni WC, Joshi CP: **Structure, organization, and functions of cellulose synthase complexes in higher plants.** *Braz J Plant Physiol* 2007, **19**(1):1-13.
101. Persson S, Paredez A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR: **Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*.** *Proc Natl Acad Sci USA* 2007, **104**(39):15566–15571.
102. Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR: **Interactions among three distinct Cesa proteins essential for cellulose synthesis.** *Proc Natl Acad Sci USA* 2003, **100**(3):1450-1455.
103. Ranik M, Myburg AA: **Six new cellulose synthase genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis.** *Tree Physiol* 2006, **26**:545-556.
104. Meins F, Fritig B, Linthorst HJM, Mikkelsen JD, Neuhaus J-M, Ryals J: **Plant chitinase genes.** *Plant Mol Biol Rep* 1994, **12**:22–28.
105. Levorson J, Chlan CA: **Plant chitinase consensus sequences.** *Plant Mol Biol Rep* 1997, **15**:122–133.
106. Zhong R, Ripberger A, Ye Z-H: **Ectopic deposition of lignin in the pith of stems of two *Arabidopsis* mutants.** *Plant Physiol* 2000, **123**:59–69.

107. Zhong R, Kays SJ, Schroeder BP, Ye Z-H: **Mutation of a chitinase-like gene causes ectopic deposition of lignin, aberrant cell shapes, and overproduction of ethylene.** *The Plant Cell* 2002, **14**:165–179.
108. Gachon C, Mingam A, Charrier B: **Real-time PCR: what relevance to plant studies?** *J Exp Bot* 2004, **55**(402):1445-1454.
109. Rider SD, Henderson JT, Jerome RE, Edenberg HJ, Romero-Severson J, Ogas J: **Coordinate repression of regulators of embryonic identity by PICKLE during germination in Arabidopsis.** *Plant J* 2003, **35**:33–43.
110. Dallas PB, Gottardo NG, Firth MJ, Beesley AH, Hoffmann K, Terry PA, Freitas JR, Boag JM, Cummings AJ, Kees UR: **Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR – how well do they correlate?** *BMC Genomics* 2005, **6**:59.
111. Yuen T, Wurmbach E, Pfeffer RL, Ebersole BJ, Sealfon SC: **Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays.** *Nucl Acids Res* 2002, **30**:e48.
112. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4**:249–264.
113. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**(9):5116–5121.
114. TIGR. [<http://www.tm4.org/mev>]
115. GenScript Corporation's Real-Time PCR Primer Design [<http://www.genscript.com/ssl-bin/app/primer>]
116. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using Real-Time quantitative PCR and the 2- $\Delta\Delta$ CT method.** *Methods* 2001, **25**:402–408.

Supplemental Figure 1.

Graph representation of selected genes relative expression profile by *in silico* analysis according to the SAM algorithm. The potentially codified protein names are indicated in the legend. Axis “y” represents the relative expression in log2. Axis “x” represents the analyzed samples, where:

Egr-cA - *Eucalyptus grandis*

r1 - clone 1

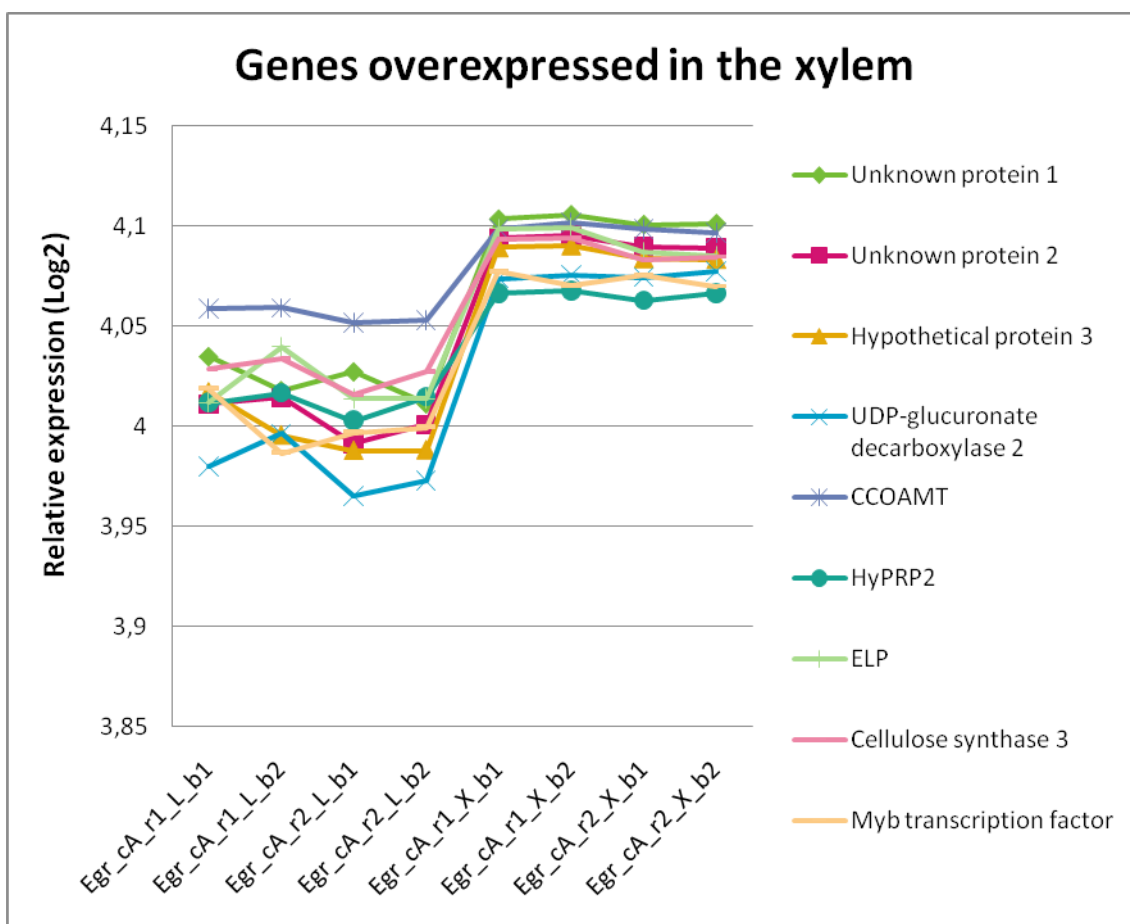
r2 - clone 2

L - leaves

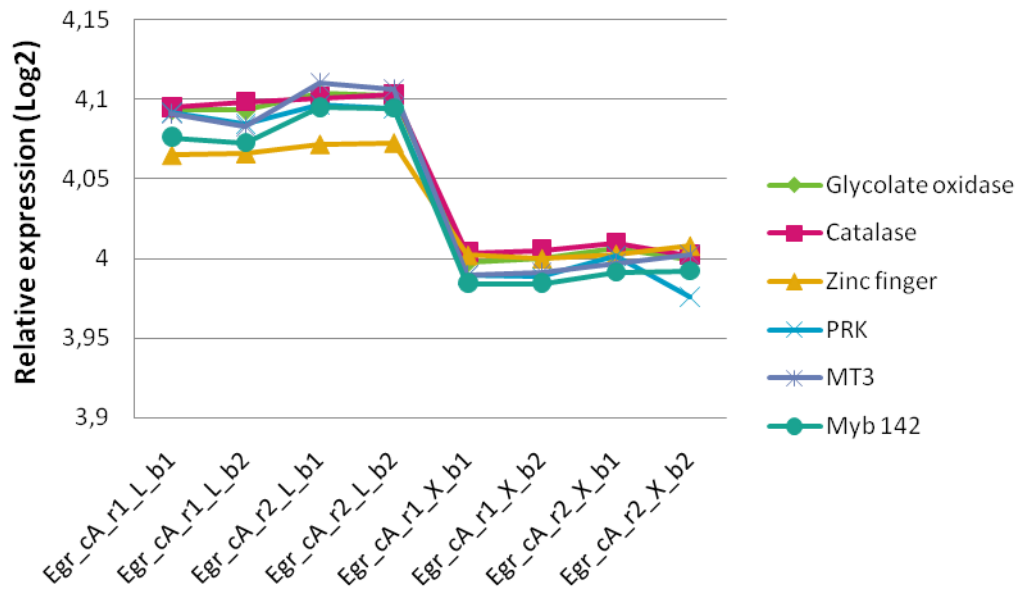
X - xylem

b1 - block 1 (slide)

b2 - block 2 (slide)

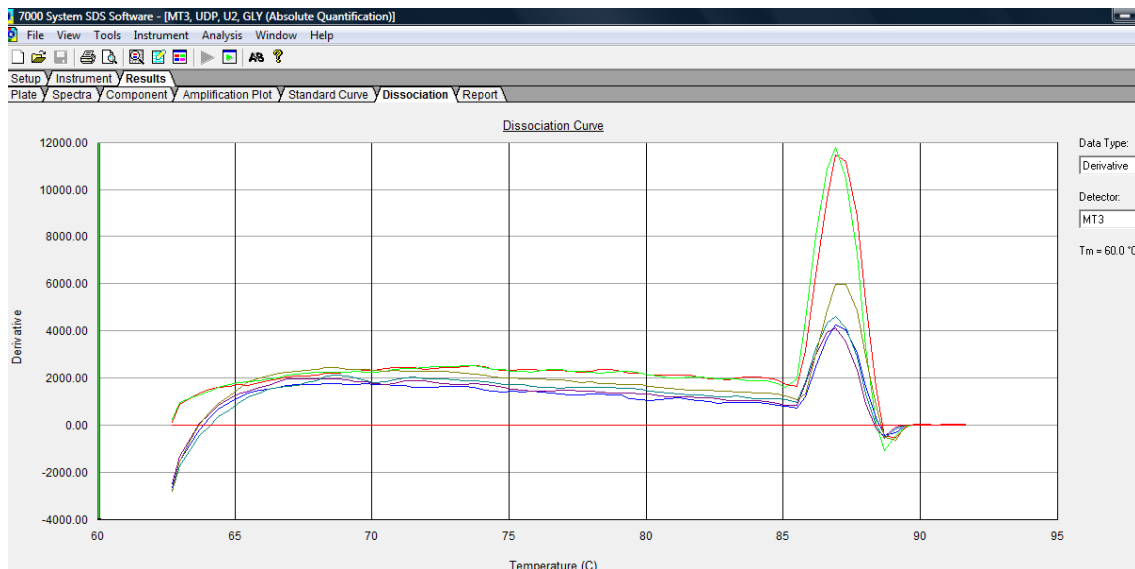


Genes overexpressed in the leaves

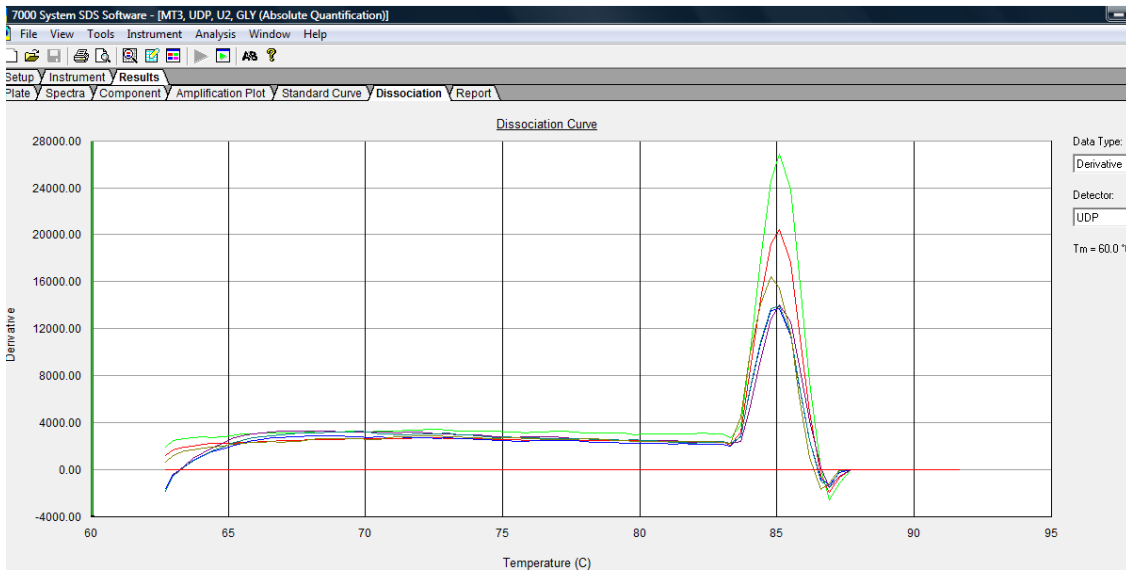


Supplemental Figure 2.

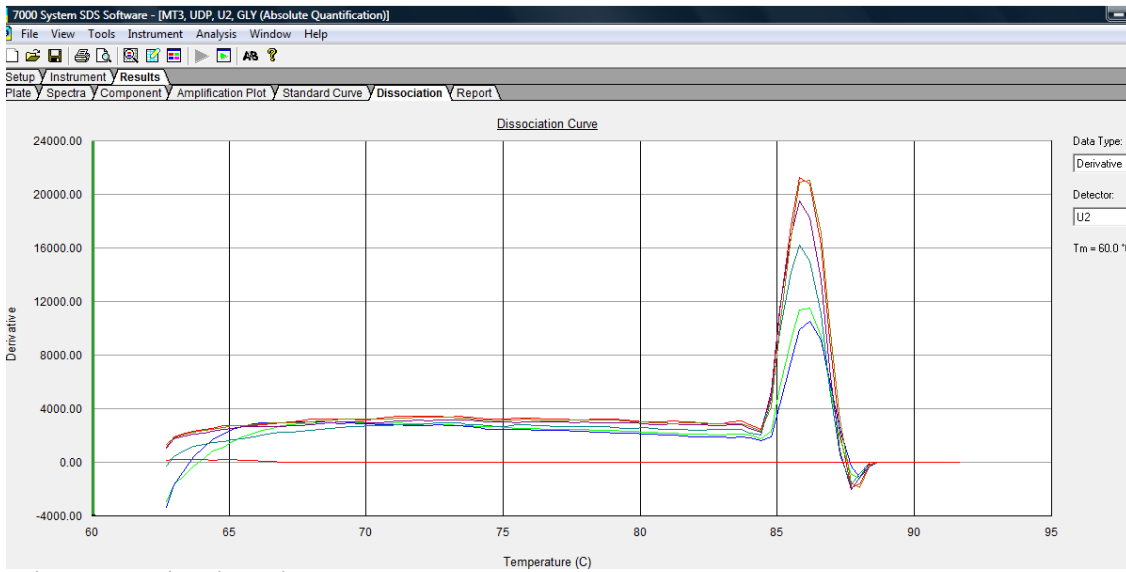
The dissociation curves from qRT-PCR amplification of primers designed were obtained after 40 cycles of amplification. The reaction specificity is observed by the presence of a single amplification peak. The occurrence of more than one peak, on the other hand, reveals the presence of unspecific products, and requires reaction optimization. The potentially codified protein names are indicated above each figure. The peaks correspond to each one of the six cDNAs from the tissues samples. A negative control was used in each reaction (line without peak).



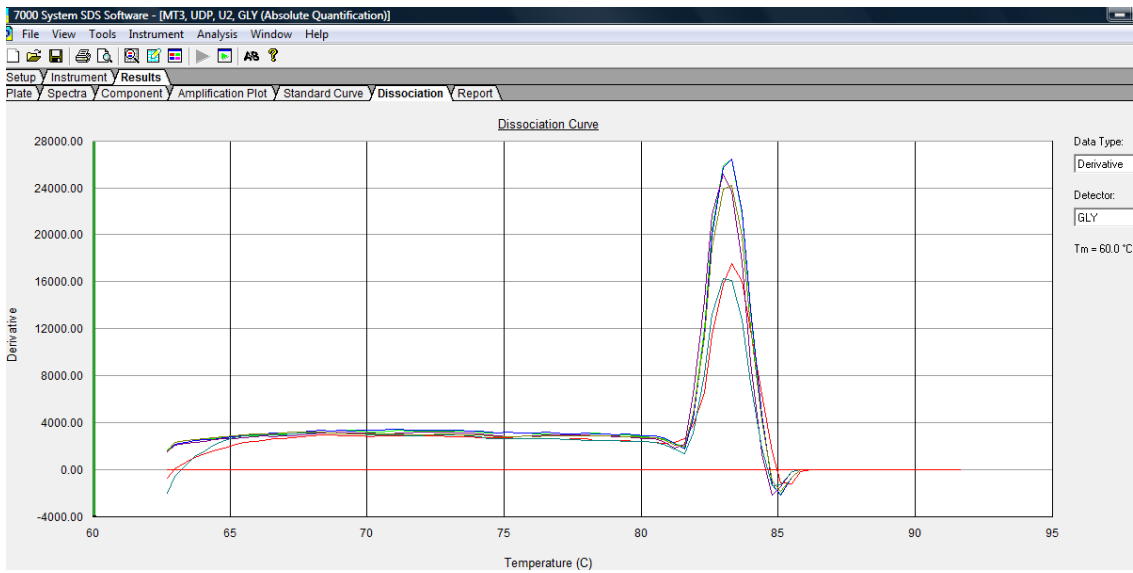
Metallothionein-like protein type 3 (MT3)



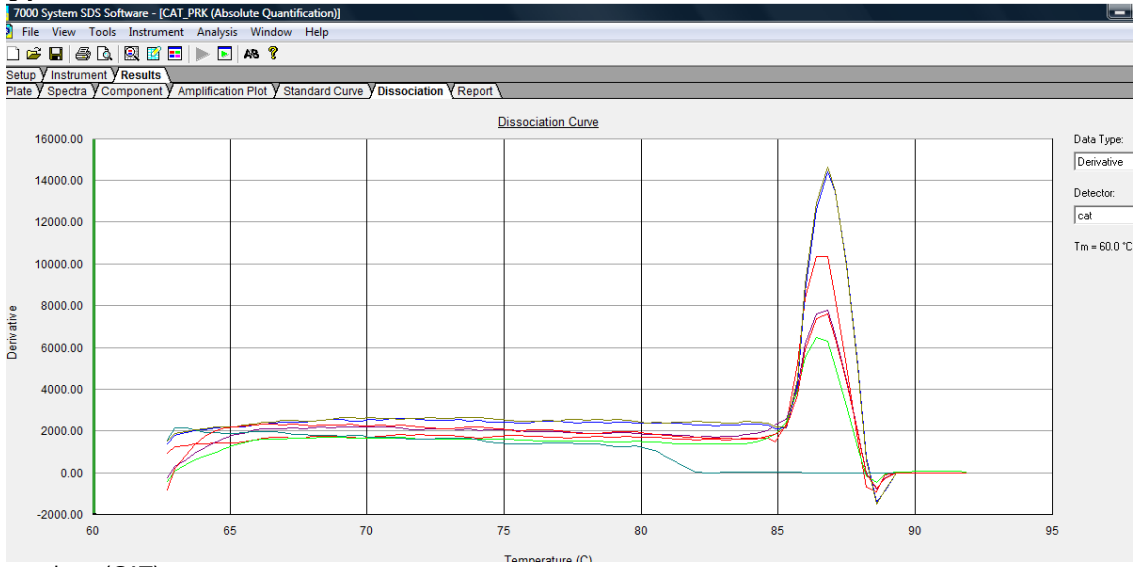
putative UDP-glucuronate decarboxylase 2 (UDP-gluc 2)



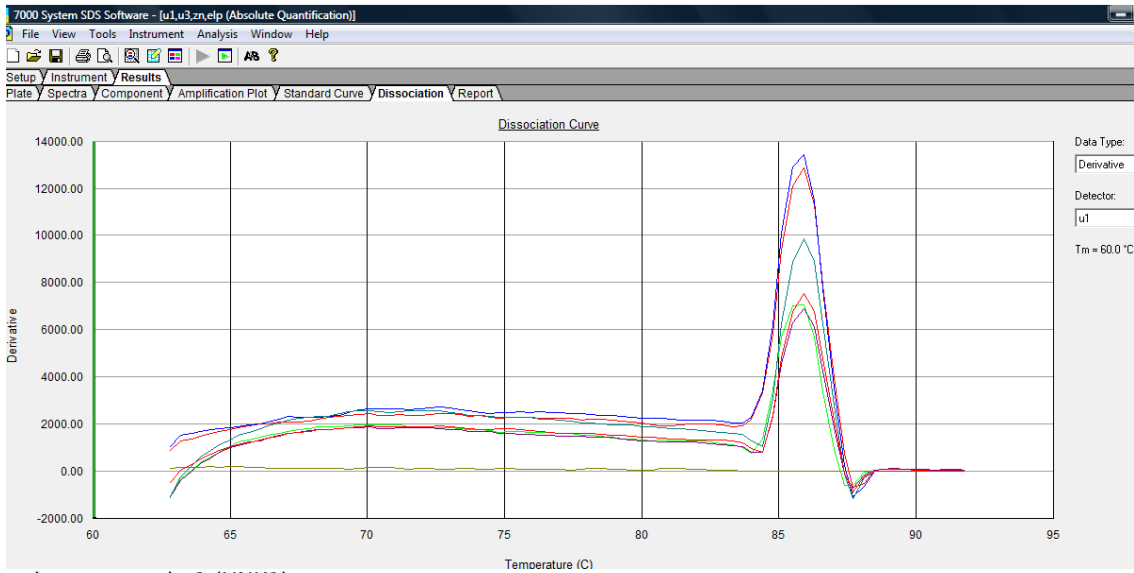
unknown protein 2 (UNK2)



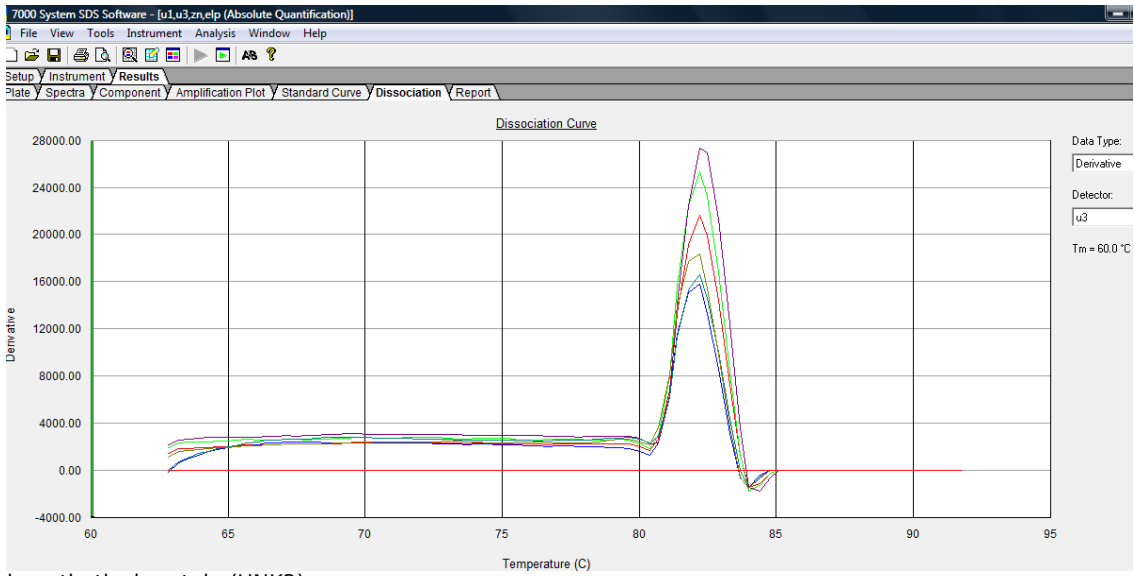
glycolate oxidase (GO)



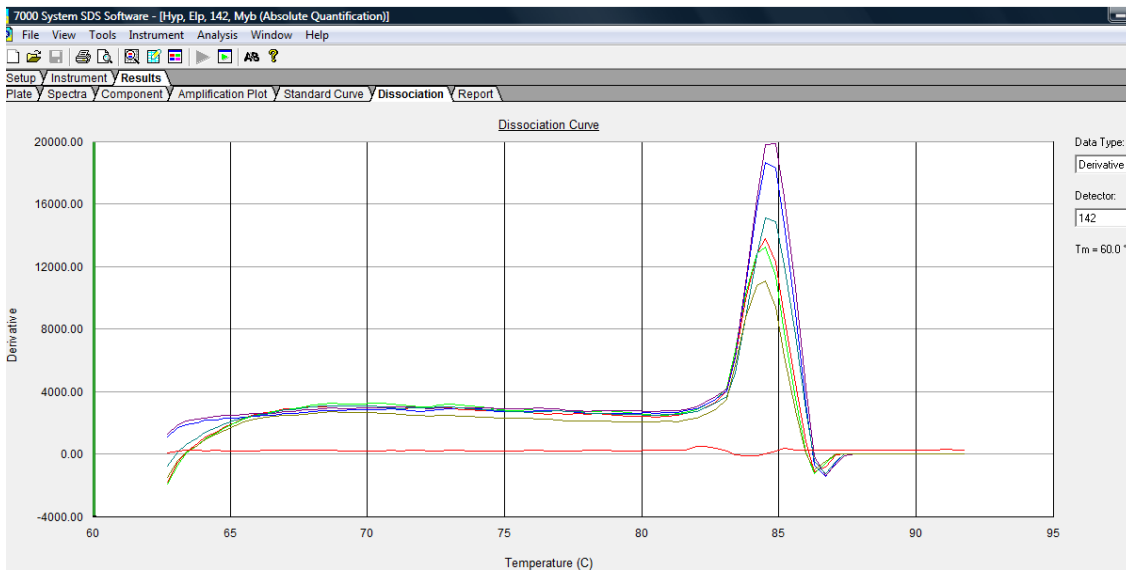
catalase (CAT)



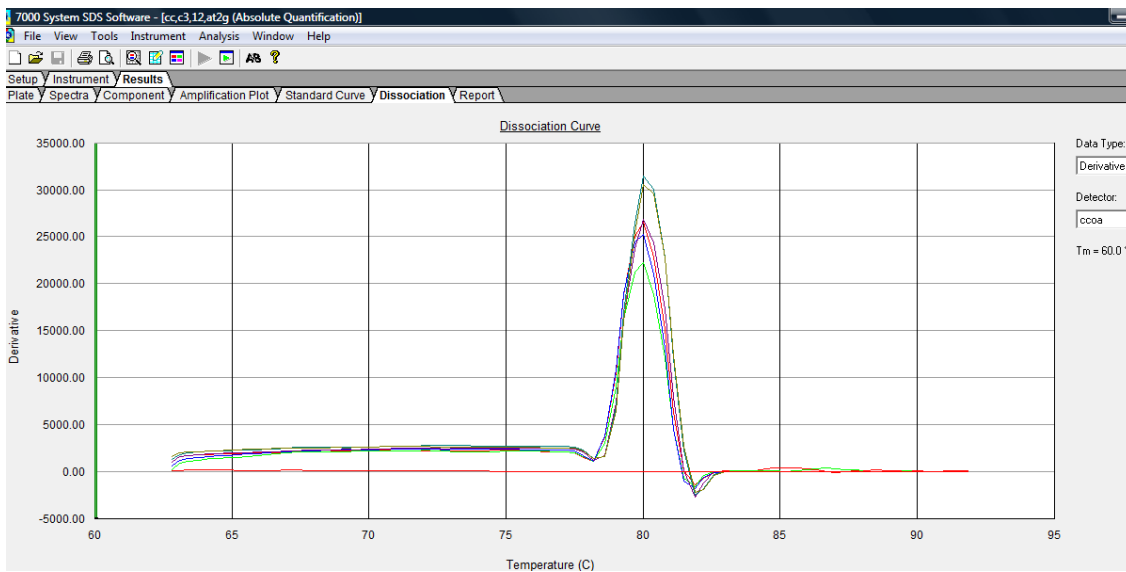
unknown protein 1 (UNK1)



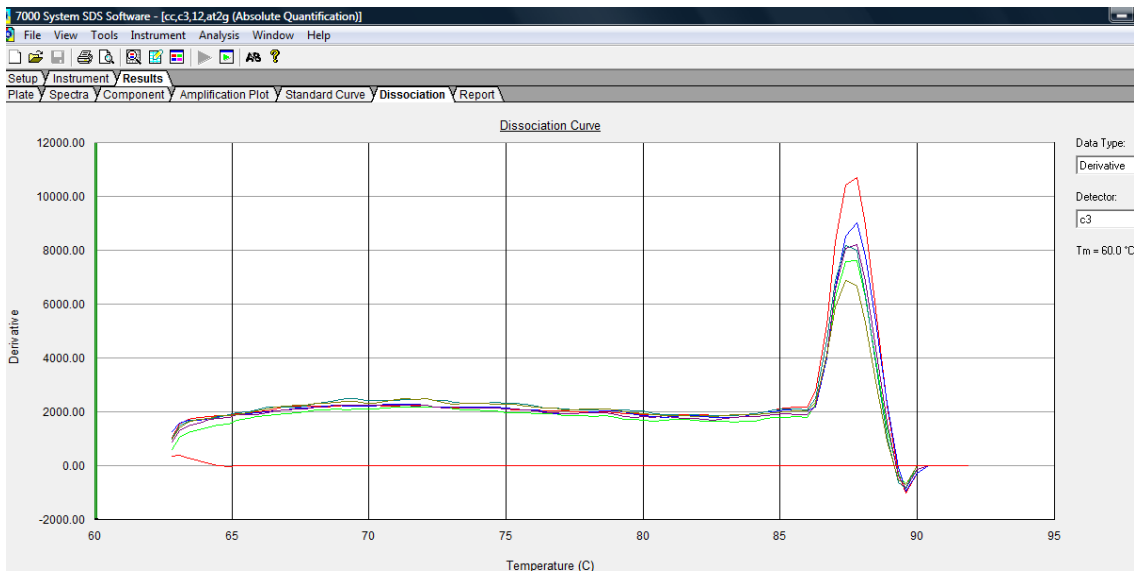
hypothetical protein (UNK3)



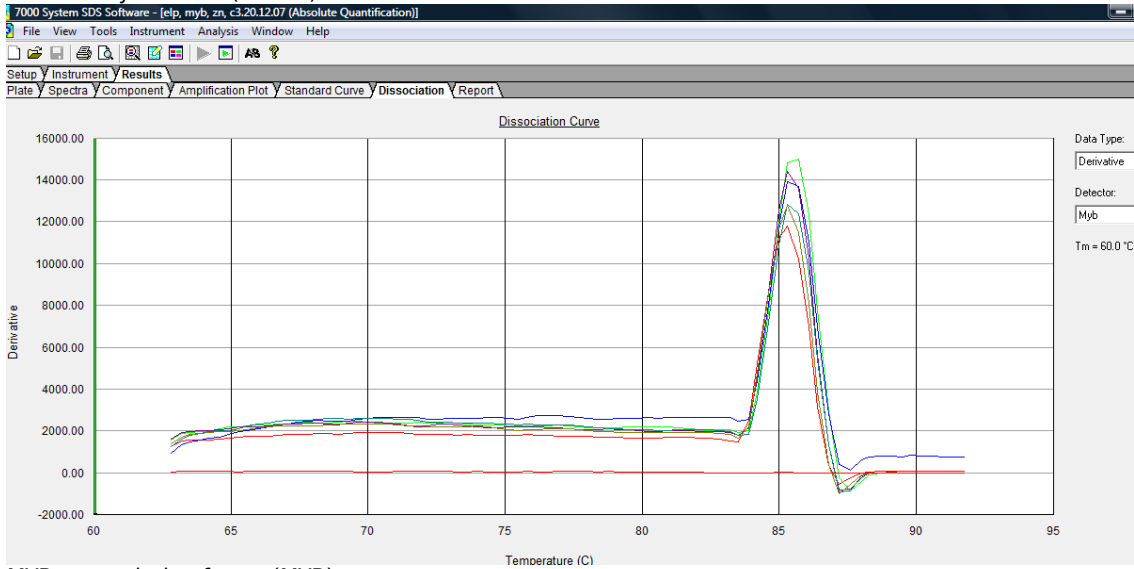
MYB transcription factor 142 (MYB142)



caffeoyl-CoA 3-O-methyltransferase (CCOAMT)



cellulose synthase 3 (CeSA3)



MYB transcription factor (MYB)

CAPÍTULO III

O PROJETO GENOLYPTUS E OS EXPERIMENTOS DE MICROARRANJOS DE DNA: CONSIDERAÇÕES EXTRAS, RESULTADOS E DISCUSSÃO

A idéia de realizar experimentos de microarranjos de DNA, assim como a definição dos objetivos, projeção dos experimentos e seleção do material biológico apropriado para testar as hipóteses de interesse (diferenças de expressão gênica entre tecido vascular de *E. grandis* e *E. globulus* e entre folhas e xilemas de *E. grandis*) foram realizados por membros do Projeto Genolyptus. Como descrito no Capítulo II, os ensaios de síntese de sondas oligonucleotídicas, marcação de cDNAs, hibridizações, lavagens, coleta de dados e análises preliminares foram desenvolvidos pela empresa NimbleGen Systems Inc. Os dados brutos foram, então, enviados ao grupo Genolyptus para a realização de análises mais refinadas e, finalmente, para a seleção de genes-candidatos. Neste Capítulo III estão descritas, de forma mais completa e justificada, as etapas realizadas desde a coleta de amostras biológicas até a seleção de genes-candidatos, apresentando-se os respectivos resultados e discussões pertinentes. A descrição técnica de “Materiais & Métodos” utilizados estão apresentados no Capítulo II.

Da coleta de amostras ao envio dos RNAs para os ensaios de microarranjo de DNA

As amostras coletadas para os experimentos de microarranjo de DNA estão enumeradas na Tabela 1. As extrações de RNA foram realizadas com o kit *PureLink Plant RNA Reagent* (Invitrogen) de acordo com o protocolo “*Miniprep*” fornecido pelo fabricante. A análise da

qualidade e da quantidade dos RNAs foi realizada por leitura em espectrofotômetro nos comprimentos de onda de 260 nm e de 280 nm. Os resultados (Tabela 2) confirmaram os baixos índices de polissacarídeos, compostos fenólicos e proteínas contaminantes, comprovados pelas razões das absorvâncias Abs_{260}/Abs_{280} superiores a 1,8 (Ausubel *et al.*, 1997). Na Figura 1 está representada a comprovação de que as amostras de RNA encontravam-se íntegras e em quantidades apropriadas, como pode ser observado pela localização e intensidade das bandas referentes aos rRNAs 28S e 18S, considerados bons indicadores da baixíssima ou ausente degradação das amostras (Ausubel *et al.*, 1997).

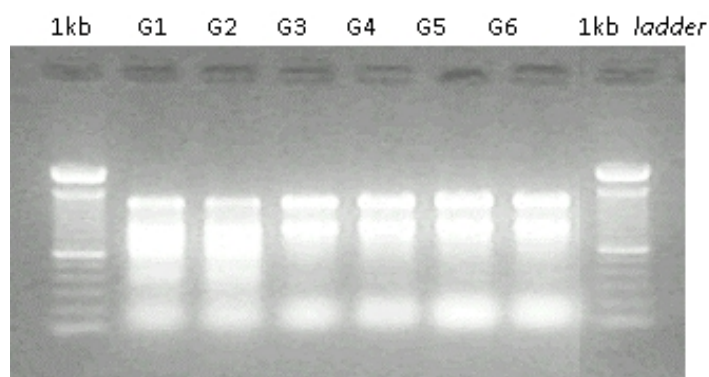


Figura 1. Imagem do resultado da eletroforese em gel de agarose a 1,5% em tampão TAE 1X corado com 0,01 mg/L de brometo de etídio. Foram aplicadas no gel amostras de 5 μ L (derivados de volumes totais de 330 μ L) de RNA total extraído de folhas maduras de dois clones de *E. grandis* (G1 e G2); xilemas de quatro clones de *E. grandis* (G3-G6). Como marcador de tamanhos de fragmentos de DNA foi utilizado o *GeneRuler 1 kb DNA Ladder* (Fermentas)

NOME DA AMOSTRA	TECIDO	CLONE
GR-XY-A1	Xilema	A1
GR-XY-A2		A2
GR-XY-B1		B1
GR-XY-B2		B2
GR-ML-A1	Folhas Maduras	B1
GR-ML-A2		B2

Tabela 1. Identidade das amostras biológicas coletadas para os ensaios de microarranjos de DNA do Projeto Genolyptus.

As amostras de RNA foram enviadas para a empresa NimbleGen, onde os microarranjos de DNA foram confeccionados. Para verificar a qualidade das amostras de RNA, técnicos da empresa as submeteram à análise com o equipamento *Bioanalyzer* (Agilent Technologies Inc.) baseado em eletroforese capilar. O resultado está representado na Figura 2. As múltiplas bandas de RNA foliar diferentes dos rRNAs 28S e 18S visualizadas nas canaletas 1 e 2 representam rRNAs de cloroplastos, 16S e 23S (Patterson & Smillie, 1971; Smart *et al.*, 1999; Tattersall *et al.*, 2005; Wang *et al.*, 2005).

Amostras de RNA	Abs ₂₆₀	Abs ₂₈₀	260/280	Concentração (ng/μL)	20 μg (μL)
G1	0.053	0.000	-----	212	94.339
G2	0.077	0.013	5.923	308	64.935
G3	0.117	0.036	3.250	468	42.735
G4	0.130	0.043	3.023	520	38.461
G5	0.211	0.081	2.604	844	23.693
G6	0.181	0.062	2.919	724	27.624

Tabela 2. Resultados das avaliações espectrofotométricas das amostras de RNA total preparadas para ensaios de hibridização de microarranjos de DNA.

Do experimento de microarranjo ao processamento dos dados brutos

As etapas subseqüentes do experimento de microarranjo foram realizadas na empresa NimbleGen. Estas incluíram: síntese de cDNAs com marcação com o grupo fluoróforo Cy3, hibridizações, lavagens, *scanning*, coleta de dados e normalizações preliminares dos dados, conforme metodologias da empresa.

Em relação à arquitetura dos microarranjos, cada gene foi avaliado por um conjunto de nove sondas, em duplicata, as quais apresentaram distintas constantes de afinidade pelo alvo e, conseqüentemente, diferentes níveis de intensidade do sinal. Para que a medida final de intensidade representasse uma quantificação dos níveis de mRNA presentes nas amostras hibridizadas, foi necessária a transformação dos valores de intensidade, referentes às sondas, em um único valor de intensidade final (Irizarry *et al.*, 2003). Para se chegar a esta etapa, primeiramente os dados foram processados pela NimbleGen. Os passos de processamento dos dados visaram eliminar o background, quantificar os sinais de fluorescência, examinar a qualidade dos dados e normalizá-los para a análise.

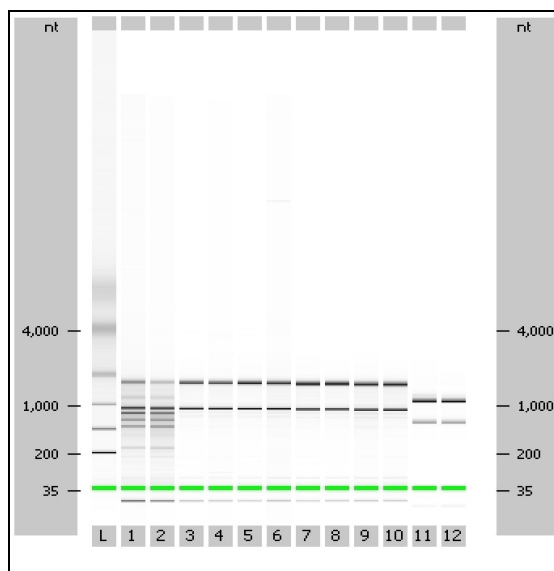


Figura 2. Imagem derivada da eletroforese capilar de fração das amostras de RNA total de *Eucalyptus* enviadas à NimbleGen e executada por técnicos daquela empresa com o equipamento *Bioanalyzer* (Agilent Technologies Inc.). A ordem de amostras aplicadas foi: (L) Marcador de tamanhos de fragmentos de DNA (origem não informada); (1 a 6) amostras G1 a G6, conforme legenda da Figura 1; (7-10) outras amostras de RNA de *Eucalyptus* (11 e 12); controles de RNAs ribossômicos de eucariotos (origem não informada).

Após o processamento, os dados foram enviados aos nossos laboratórios. Estes, quando recebidos, foram submetidos a nova normalização pelo método “Análise Robusta de Multi-arranjos” (do inglês, *Robust Multi-Array Analysis* ou RMA; Irizarry *et al.*, 2003). Somente após essas etapas, a identificação dos genes candidatos e a interpretação dos dados puderam ser realizadas, visando o reconhecimento dos padrões de expressão gênica e a compreensão do significado biológico de tais padrões (Clarke & Zhu, 2006).

Das análises preliminares aos resultados finais

As análises preliminares foram realizadas a fim de avaliar a consistência dos dados dos microarranjos fornecidos pela plataforma da NimbleGen. Para tanto, foram realizadas comparações entre diversos grupos de amostras pelo Teste *t* (resultados não apresentados). A consistência dos dados foi comprovada em termos da baixa variação observada entre os blocos do mesmo *slide* (*chip*) e entre as réplicas biológicas, ou seja, com uma elevada reprodutibilidade. Assim, como resultado, foi possível validar a robustez dos microarranjos realizados pela NimbleGen.

As análises dos dados já normalizados referentes aos quatro microarranjos de DNA específicos às folhas e xilemas dos clones 1 e 2 derivados de uma matriz de *E. grandis* foram realizadas pelo método “Análise da Significância de Microarranjos” (do inglês, *Significance Analysis of Microarrays* ou SAM; Tusher *et al.*, 2001). O algoritmo SAM faz parte do pacote MeV, disponibilizado pelo TIGR, e é uma alternativa ao Teste *t* de Student para a busca de expressão diferencial. Este teste leva em consideração o grande número de genes analisados e garante um nível de erro de 0%, diferentemente dos testes estatísticos comumente utilizados, como o Teste *t* ou a “Análise da Variância” (ANOVA, com porcentagens de erro de no mínimo 1%). No caso da análise desse experimento de microarranjo que incluiu um total de 21.442 seqüências únicas, um erro de 1% poderia significar 214 genes com falsa expressão diferencial.

O método SAM baseia-se na razão entre a mudança na expressão gênica entre duas condições e o erro padrão das medições. O algoritmo SAM permite identificar genes com mudanças estatísticas significantes na expressão por meio da realização de conjuntos de Testes *t* gene-específicos. Genes com valores maiores do que o ponto de corte são considerados potencialmente significantes. A porcentagem desses genes identificados ao acaso é a denominada “Falsa Razão de Descobertas” (do inglês, *False Discovery Rate* ou FDR) que, nestas análises, foi de 0%.

No algoritmo SAM, os dados de cada gene são permutados e um valor estatístico *d* (*d-value*) é computado para ambos os dados, o original e o permutado de cada gene. Nas análises finais, utilizamos o teste de duas classes não pareadas. Nesse tipo de análise, *d* é análogo ao valor estatístico *t* do Teste *t*, no qual é capturada a diferença entre os níveis de expressão média das condições experimentais, representados pela medida da variância dos dados. Na análise de duas classes, genes são

considerados positivos se a média de expressão do grupo B for significativamente maior que a do grupo A. São considerados negativos se o inverso ocorrer (MeV Manual, 2006).

Ao final das análises, o programa MeV gerou um gráfico interativo (Figura 1 do capítulo II) onde estão os valores observados *versus* os valores esperados (baseados nos *d-values*). O valor de delta foi ajustado manualmente no gráfico e o mesmo representa a distância vertical (em unidades de gráfico) da linha sólida que representa o grau de inclinação (quando o observado é igual ao esperado) e entre duas linhas paralelas pontilhadas que representam a região dentro da qual os genes são considerados sem expressão diferencial significativa. Saindo desta linha, os genes apresentam expressão diferencial significativa. Foi utilizado, ainda, o critério de *fold change* que determina o número de vezes que um gene é mais ou menos expresso em relação a outro (MeV Manual, 2006).

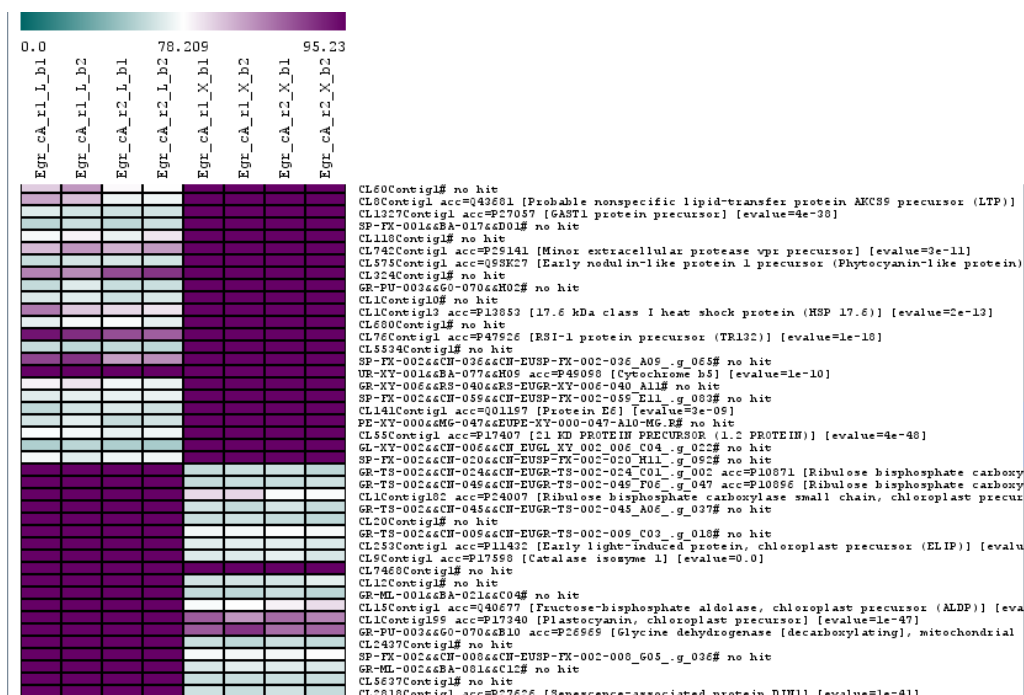


Figura 3. Representação gráfica de 42 dos 302 genes diferencialmente expressos entre folhas e xilema de *E. grandis* segundo análises dos resultados de microarranjos de DNA. Cada linha da figura representa um único gene que possui expressão variando entre baixa (verde) e alta (roxa) em folhas (esquerda) e xilema (direita). Onde: Egr_cA, *Eucalyptus grandis*; r1, clone 1; r2, clone 2; L, leaf; X, xylem; b1, block 1 (slide); b2, block 2.

Com o objetivo principal de identificar os genes mais diferencialmente expressos entre xilema e folhas de *E. grandis*, foram comparados os dados normalizados dos microarranjos hibridizados com cDNAs derivados dos xilemas e folhas dos clones 1 e 2 de uma única matriz de *E. grandis*, denominada matriz “A”, incluindo os dois blocos de cada *slide* (réplicas técnicas ou experimentais):

<u>Grupo 1: FOLHAS</u>	<u>Grupo 2: TECIDO</u>
VASCULAR	
Egr_cA_r1_L_b1	Egr_cA_r1_X_b1
Egr_cA_r1_L_b2	Egr_cA_r1_X_b2
Egr_cA_r2_L_b1	Egr_cA_r2_X_b1
Egr_cA_r2_L_b2	Egr_cA_r2_X_b2

Foram realizadas 70 permutações únicas entre os dois grupos e foi utilizado um *fold change* de 2, ou seja, apenas genes com valor de expressão diferencial mínimo de duas vezes foram selecionados. O valor de delta utilizado foi de 9,15. Como resultados desta análise foram obtidos 153 genes com expressão maior em folhas e 149 genes com maior expressão em tecido vascular de *E. grandis*. O perfil de expressão de todos os genes analisados estão demonstrados nas Figuras 3, 4 e 5.

Do total de aproximadamente 300 genes, com base nos maiores valores de *fold change*, foram selecionados os 30 genes candidatos mais expressos em folhas e 30 mais expressos em xilema de *E. grandis* e seus

perfis de expressão foram analisados *in silico*. A seguir, uma busca pelos clones de cDNA potencialmente mais completos e representativos de cada um dos agrupamentos representados pelos 60 supostos genes foi realizada no banco de dados de ESTs do Projeto Genolyptus. Em seguida, a anotação foi refeita (com a ferramenta BLASTx) submentendo as seqüências deduzidas de peptídeos dos 60 genes candidatos contra o banco de seqüências protéicas derivadas de plantas disponíveis pelo NCBI (dados não apresentados). Uma segunda seleção foi realizada baseada nas identidades moleculares e homologias dos genes candidatos e seus prováveis papéis no metabolismo em células vegetais. No total, 15 genes foram selecionados (Tabela 3), seu perfil de expressão analisado *in silico* e *primers* para qRT-PCR foram projetados, sendo seis mais expressos nas folhas e nove no tecido vascular (dados demonstrados no Capítulo II). Por meio dessa análise, foi possível verificar o alto nível de expressão diferencial entre os genes selecionados.

Abreviatura	Identidade (BLASTx)	Fold Change (log₁₀)
MT3	Proteína semelhante à metalotioneína do tipo 3 [<i>Carica papaya</i>] 6e-26	2,55
GO	Glicolate-oxidase [<i>Mesembryanthemum crystallinum</i>] 0.0	2,35
CAT	Catalase [<i>Prunus persica</i>] 0.0	2,30
PRK	Fosforribulocinase precursora de cloroplasto (PRK) [<i>Mesembryanthemum crystallinum</i>] 0.0	2,27
MYB142	Fator de transcrição MYB do tipo MYB142 [<i>Glycine max</i>] 3e-25	2,17
CO	Proteína tipo “dedo-de-zinco” CONSTANS-LIKE 16, [<i>Arabidopsis thaliana</i>] 5e-26	1,38
HyPRP2	Proteína rica em prolina híbrida do tipo 2 (HyPRP2) [<i>Gossypium hirsutum</i>] 9e-24	1,15
UNK1	Proteína expressa [<i>Arabidopsis thaliana</i>] 5e-11	2,09
UNK2	Proteína expressa [<i>Arabidopsis thaliana</i>] 3e-110	2,00

UNK3	Proteína hipotética [<i>Nicotiana benthamiana</i>] 4e-13	1,96
UDP-gluc 2	Descarboxilase de UDP-glicuronato do tipo 2, [<i>Nicotiana tabacum</i>] 1e-175	1,87
ELP	Quitinase do tipo ELP [<i>Arabidopsis thaliana</i>] 1e-20	1,79
CeSA3	Celulose-sintase tipo 3 [<i>Eucalyptus grandis</i>] 0.0	1,60
MYB	Fator de transcrição MYB [<i>Eucalyptus gunnii</i>] 7e-150	1,49
CCOAMT	Cafeoil-CoA-3-O-metiltransferase [<i>Linum usitatissimum</i>] 3e-15	1,28

Tabela 3. Lista dos genes selecionados de acordo com a análise dos microarranjos pelo método de SAM, com as respectivas abreviaturas. Os valores de identidade atribuídos por BLASTx estão representados pelos *E-values*.

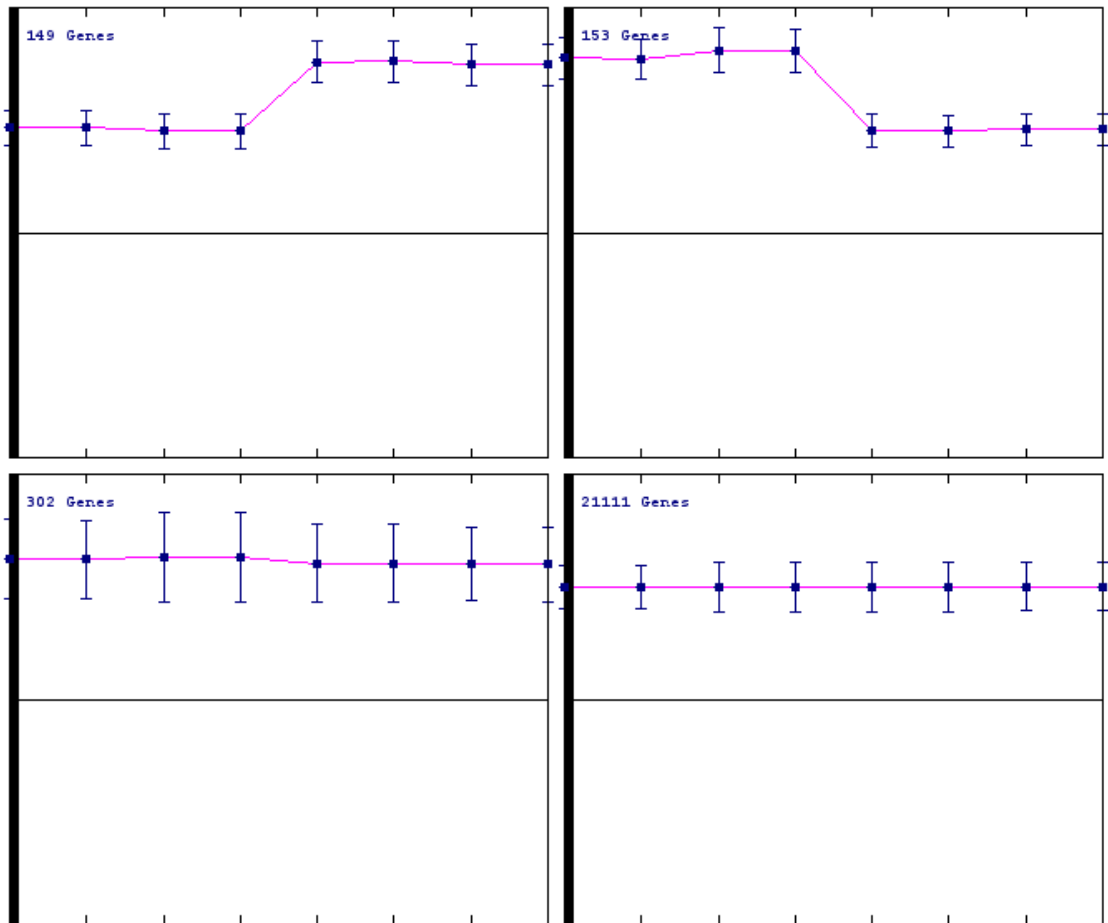


Figura 4. Gráfico gerado pelo algoritmo SAM. Correspondente à média e desvio-padrão das intensidades de sinal (em log₂) dos genes nas amostras de folhas (primeiras quatro à esquerda) e xilemas (últimas quatro à direita) do experimento de microarranjo. A: Genes mais expressos no tecido vascular. B: Genes mais expressos nas folhas. C: Todos os genes com expressão diferencial significativa (sobreposição de A e B). D: Genes sem expressão diferencial significativa.

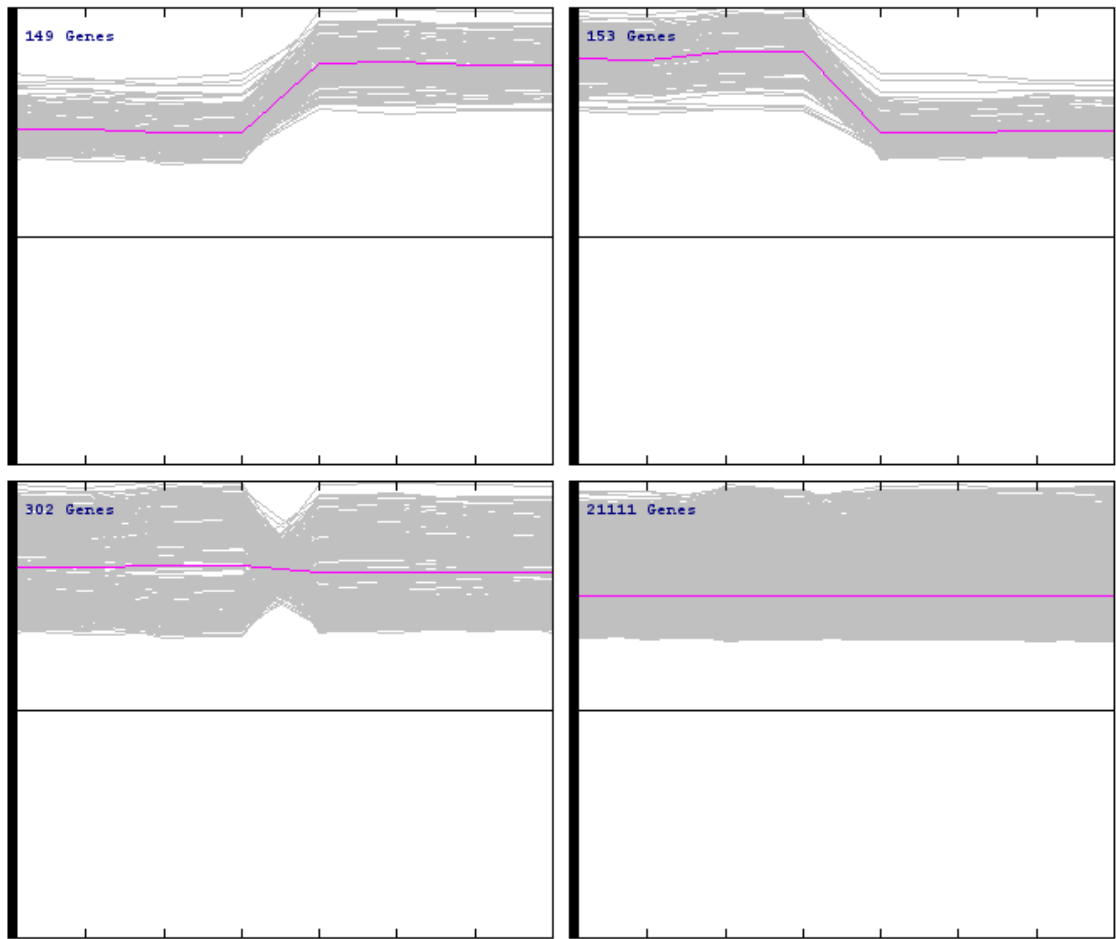


Figura 5. Gráfico gerado pelo algoritmo SAM. Cada linha representa a intensidade de sinal (em \log_2) de um único gene nas amostras de folhas (primeiras quatro à esquerda) e xilemas (últimas quatro à direita) do experimento de microarranjo. A linha rosa representa a média. A: Genes mais expressos no tecido vascular. B: Genes mais expressos nas folhas. C: Todos os genes com expressão diferencial significativa (sobreposição de A e B). D: Genes sem expressão diferencial significativa.

O pacote MeV (fornece, ainda, diversas outras ferramentas para a análise de dados de microarranjos de DNA. A formação de agrupamentos (*clusters*) de genes com perfis semelhantes de expressão é uma dessas ferramentas. Métodos para a análise de *clusters* fornecem uma maneira relativamente simples de organizar informações, pois permitem ao usuário ver, entender e analisar os dados com maior facilidade.

O grande número de genes e a complexidade das vias metabólicas aumentam em muito os desafios de compreender e interpretar os tantos dados resultantes da análise de microarranjos de DNA. O primeiro passo em direção a esse desafio é o uso de técnicas de *clustering* (“agrupar”), as quais são essenciais na identificação de padrões interessantes dos dados. No caso da análise do experimento de microarranjo como um todo, o agrupamento de genes em *clusters* permite separar, juntamente com genes de função conhecida e de interesse, muitos outros genes que possuem um padrão de expressão semelhante e provavelmente estejam envolvidos numa mesma rota metabólica ou exerçam função semelhante. *Clusters* apenas dos genes estudados por qRT-PCR foram criados neste trabalho, como exemplificado na Figura 6. O primeiro método utilizado foi o de “Análise Hierárquica de *Clusters*” (do inglês, *Hierarchical Cluster Analysis* ou HCL). Outro tipo de método utilizado para a análise dos dados foi o da “Técnica de Busca de *Clusters* por Afinidade” (do inglês, *Cluster Affinity Search Technique* ou CAST). Este método permitiu a geração de 5 *clusters* segundo os perfis de expressão gênica, conforme ilustrado na Figura 7. O perfil de expressão de cada *cluster* pode ser visualizado na Figura 8.

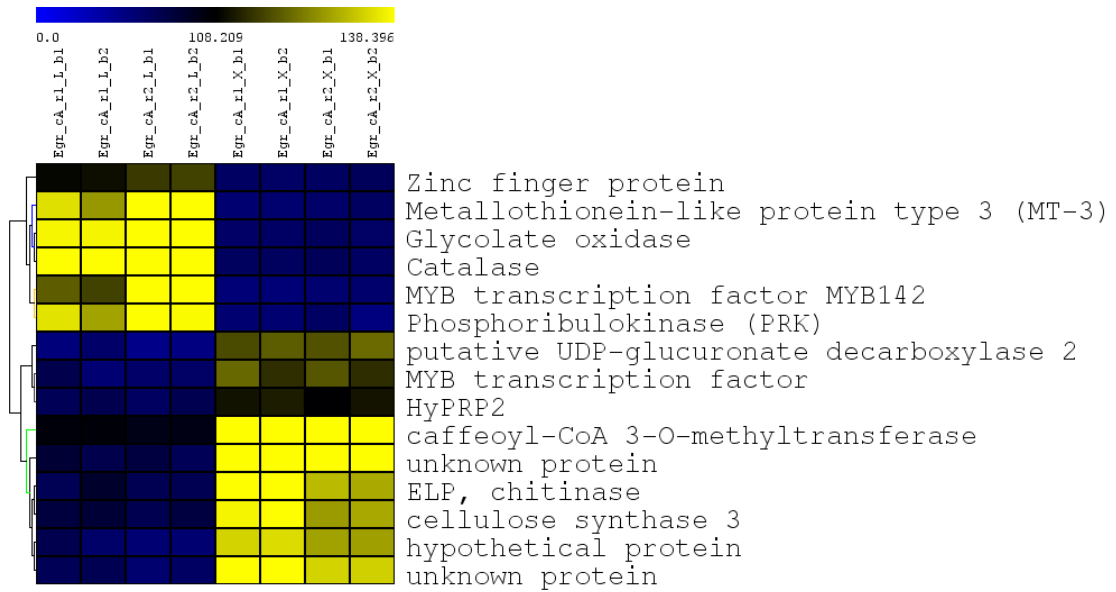
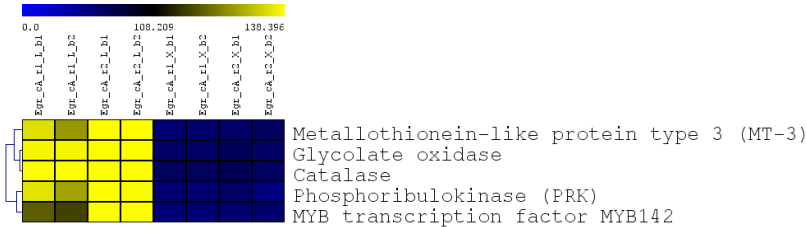
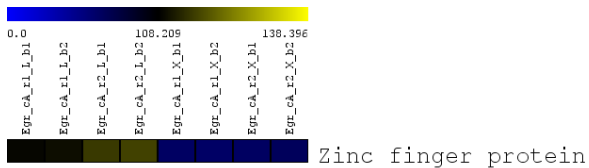


Figura 6. Análise Hierárquica de *Clusters* (HCL) de 15 genes de *E. grandis* selecionados segundo suas expressões diferenciais em folhas e xilema. A escala de cores na parte superior da figura demonstra a expressão baixa (azul) a alta (amarela) para os genes analisados. Onde: Egr_cA, *Eucalyptus grandis*; r1, clone 1; r2, clone 2; L, leaf; X, xylem; b1, block 1 (slide); b2, block 2.

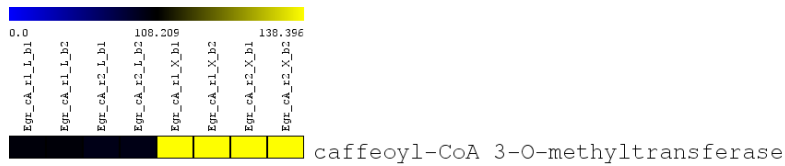
A - Cluster 1



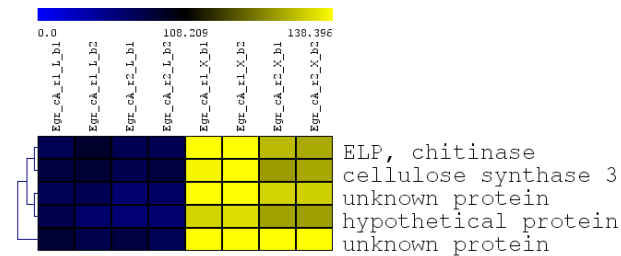
B - Cluster 2



C - Cluster 3



D - Cluster 4



E - Cluster 5

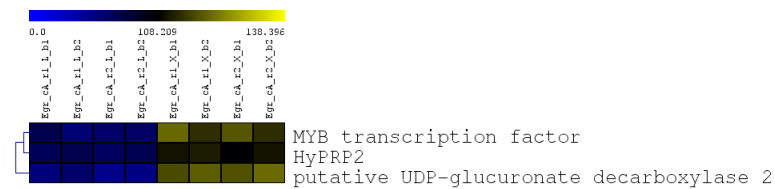


Figura 7: Clusters formados pela técnica CAST e os genes participantes de cada cluster. A-E: **Cluster 1:** MT3, GO, CAT, PRK e MYB142; **Cluster 2:** CO; **Cluster 3:** CCOAMT; **Cluster 4:** ELP, CeSA3, UNK1, UNK2 e UNK3; **Cluster 5:** MYB, UDP-gluc 2 e HyPRP2 (abreviações de acordo com Tabela 2 do Capítulo II). Onde: Egr_cA, *Eucalyptus grandis*; r1, clone 1; r2, clone 2; L, leaf; X, xylem; b1, block 1 (slide); b2, block 2.

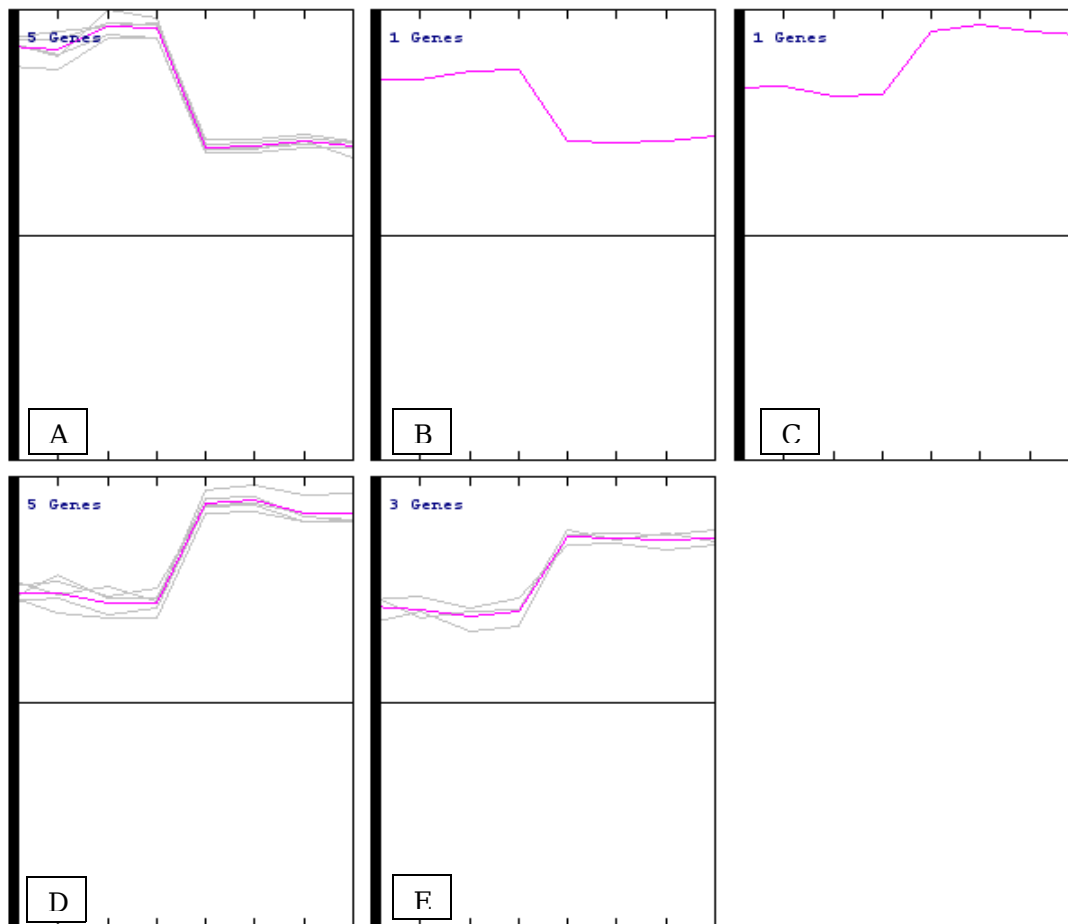


Figura 8: *Clusters* formados pela técnica CAST e o perfil de expressão dos genes de cada *cluster*. Valores de expressão em escala \log_2 . O nome do *cluster* está indicado no canto inferior esquerdo de cada quadro. A-E: **Cluster 1:** MT3, GO, CAT, PRK e MYB142; **Cluster 2:** CO; **Cluster 3:** CCOAMT; **Cluster 4:** ELP, CeSA3, UNK1, UNK2 e UNK3; **Cluster 5:** MYB, UDP-gluc 2 e HyPRP2 (abreviações de acordo com Tabela 2 do Capítulo II).

Estratégias genômicas baseadas em microarranjos e outras técnicas de larga-escala estão tornando-se cada vez mais importantes na biologia e química. Como resultado, é necessário desenvolver táticas para

“enxergar” a informação contida em um enorme número de tabelas de medidas quantitativas produzidas por este tipo de técnica. Uma premissa básica na organização de dados de expressão gênica é agrupar genes com padrões muito similares de expressão (Eisen *et al.*, 1998). Seria esperado que os genes agrupados num mesmo cluster exercessem a mesma função biológica ou que estivessem sob o controle dos mesmos fatores de transcrição. É importante ressaltar que o uso de técnicas de validação das análises de clusters é necessário para que as mesmas sejam confiáveis (Handl *et al.*, 2005).

Como resultado da análise pela técnica CAST, pode-se ver a visível separação de grupos de genes com padrões de expressão semelhantes. O *cluster 1* engloba os genes mais diferencialmente expressos nas folhas de *E. grandis* (MT3, GO, CAT, PRK e MYB142), que possuem uma expressão diferencial média de 2,3 (*Fold change* em escala \log_{10}). O *cluster 2* é composto somente pelo gene CO, que apresenta menor grau de expressão diferencial nas folhas (1,4). Os *clusters 3, 4 e 5* englobam somente genes diferencialmente expressos nos tecidos vasculares. Enquanto o *cluster 3* abriga o gene CCOAMT, que é o menos diferencialmente expresso, o *cluster 4* agrupa genes com a maior expressão diferencial no xilema (ELP, CeSA3, UNK1, UNK2 e UNK3), com média de *fold change* de 1,7. Já no *cluster 5*, estão genes com uma variação um pouco menor de expressão entre folhas e xilema. São estes: MYB, UDP-gluc 2 e HyPRP2, com média de *fold change* de 1,5.

A técnica de análise de cluster, como pôde ser visto neste caso, mostrou-se confiável. No caso do uso deste tipo de técnica para análise global de experimentos de microarranjo, é de extrema importância que os resultados obtidos possam e sejam validados assim como o são os resultados das análises de expressão gênica do experimento de microarranjo.

Selecionar uma grande quantidade de genes potencialmente interessantes a partir de experimentos de microarranjos de DNA e confirmar os candidatos por análise de RT-PCR quantitativo em tempo real é uma estratégia amplamente utilizada. Existem diversos trabalhos demonstrando a correlação entre experimentos de microarranjos e análise de qRT-PCR. Enquanto numerosos estudos obtiveram resultados similares por RT-PCR quantitativo em tempo real e experimentos de microarranjos, com uma correlação linear de até cinco ordens de magnitude, muitos outros artigos relataram que a variação, em número de vezes, medida pelo qRT-PCR foi geralmente menor (em até 10 vezes) do que a variação medida quando se utilizam experimentos de microarranjos, nos quais podem ocorrer hibridização cruzada para genes muito semelhantes, porém distintos (revisado em Gachon *et al.*, 2004). Em 2003, Wang *et al.* demonstraram que dados gerados por PCR em tempo real poderiam apresentar valores em escala maior do que os dados gerados por microarranjos, porém conservando ainda uma boa correlação entre as duas técnicas. Dados obtidos a partir do trabalho de Dallas *et al.* (2005) indicam que as correlações entre qRT-PCR e experimentos de microarranjos são geralmente fortes, um resultado um tanto encorajador para pesquisadores com acesso a pouca quantidade ou raro acesso a amostras. Nesse trabalho, é enfatizada a importância de garantir que a reação de PCR e o experimento de microarranjo reconheçam o mesmo transcrito. Em todos os casos, RT-PCR em tempo real é considerada a técnica mais confiável e discrepâncias podem frequentemente ser consideradas como resultado dos métodos de normalização e subtração de *backgrounds* utilizados na análise de microarranjos. Como resultado, análises de qRT-PCR estão sendo utilizadas como referência para comparar diferentes métodos de análises de experimentos de microarranjos (Puthoff *et al.*, 2003).

CONSIDERAÇÕES FINAIS

No presente trabalho, foi realizada a análise de experimentos de microarranjos de DNA com o objetivo de encontrar genes diferencialmente expressos entre folhas e xilema de *E. grandis*, assim como escolher genes para futuros caracterização e estudo. Por se tratar de uma tecnologia nova e nunca antes utilizada no Centro de Biotecnologia da UFRGS, a análise dos microarranjos requisitou o estabelecimento de técnicas de bioinformática e estatística. Após a primeira etapa de análise e validação de resultados, os dados gerados estão sendo de extrema utilidade para o projeto Genolyptus como um todo, fornecendo bases científicas que deverão ser testadas experimentalmente, dando origem a outros projetos em muitos dos laboratórios participantes. Tais projetos incluem a super-expressão e a inibição (RNAi) dos genes diferencialmente expressos em plantas de *A. thaliana*, *Nicotiana tabacum* e um modelo experimental de *Eucalyptus* sp. em desenvolvimento, assim como estudos de fatores de transcrição diferencialmente expressos em xilemas de *E. grandis* e *E. globulus*. A partir dos resultados gerados após esta primeira etapa, novos experimentos de microarranjos para indivíduos selecionados serão realizados com a finalidade de fornecer informação em abundância sobre os principais processos envolvidos na gênese da madeira de *Eucalyptus*.

REFERÊNCIAS BIBLIOGRÁFICAS (Capítulos I e III)

Abbott, A.; Georgi, L.; Yvergnaux, D.; Wang, Y.; Blenda, A.; Reighard, G.; Inigo, M. & Sosinski B. Peach the model genome for *Rosaceae*. *Acta Hort.*, 575: 145–155, 2002.

ABRAF. *Anuário estatístico da ABRAF: ano base 2006*. ABRAF Brasília, 2007.

Disponível em: [http: <www.abraflor.org.br/estatisticas/anuario-ABRAF-2007.pdf>](http://www.abraflor.org.br/estatisticas/anuario-ABRAF-2007.pdf). acesso em dezanove de dezembro de 2007.

Alba, R.; Payton, P.; Fei, Z.; McQuinn, R.; Debbie, P.; Martin, G. B.; Tanksley S. D. & Giovannoni, J. J. Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. *The Plant Cell*, 17: 2954–2965, 2005.

Allona, I.; Quinn, M.; Shoop, E.; Swope, K.; St.Cyr, S.; Carlis, J.; Riedl, J.; Retzel, E.; Campbell, M. M.; Sederoff, R. & Whetten, R. W. Analysis of xylem formation in pine by cDNA sequencing. *P Natl Acad Sci USA*, 95: 9693-9698, 1998.

An, G.; Ebert, P. R.; Yi, B. & Choi, C. H. Both tata box and upstream regions are required for the nopaline synthase promoter activity in transformed tobacco cells. *Mol Gen Genet*, 203: 245–250, 1986.

Andersson-Gunneras, 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. *Plant J*, 45: 144–165, 2006.

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 Physiol*, 137: 983–997, 2005.

Ausubel, F. M.; Brent, R.; Kingston, R. E.; Moore, D. D.; Smith, J. A.; Seidman, J.G. & Stuhl, K. *Current Prot Mol Biol*. New York: John Wiley and Sons, 1997.

Beers, E. P. & Zhao, C. *Arabidopsis* as a model for investigating gene activity and function in vascular tissues. Molecular breeding of wood plants. *Proceedings of the International Wood Biotechnology Symposium*, Narita, Japan, 43: 52, 2001.

Bergeron, D.; Boivin, R.; Baszczyński, C. L. & Bellemare, G. Roots specific expression of a glycine-rich protein in *Brassica napus*. *Plant Sci*, 96: 87–98, 1994.

Bhandari, S.; Fujino, T.; 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*, 224: 828–837, 2006.

Boudet, A. M.; Kajita, S.; Grima-Pettenati, J. & Goffner, D. Lignins and lignocellulosics: a better control of synthesis for new and improved uses. *TRENDS Plant Sci*, 8(12): 576-581, 2003.

Bourquin, V.; Nishikubo, N.; Abe, H.; Brumer, H.; Denman, S.; Eklund, M.; Christiernin, M.; Teeri, T. T.; Sundberg, B. & Mellerowicz, E. J. Xyloglucan endotransglycosylases have a function during the formation of secondary cell walls of vascular tissues. *The Plant Cell*, 14: 3073–3088, 2002.

BRACELPA - Associação Brasileira de Celulose e Papel: *Relatório anual de desempenho 2006*. Disponível em: <<http://www.bracelpa.org.br/bra/estatisticas/>> acesso em dezanove de dezembro de 2007.

Brito, J. O.; Barrichelo, L. E. G. & Seixas, F. Análise da Produção Energética e de Carvão Vegetal de Espécies de Eucalipto. *Instituto de Pesquisas e Estudos Florestais*, 23: 53-56, 1993.

Brunner, A. M.; Busov, V. B. & Strauss, S. H. Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends in Plant Sci*, 9(1): 49-56, 2004.

Calsa, Jr. T. & Figueira, A. Serial analysis of gene expression in sugarcane (*Saccharum* spp.) leaves revealed alternative C₄ metabolism and putative antisense transcripts. *Plant Mol Biol*, 63(6): 745-62, 2007.

Cassab, G. I. Plant cell wall proteins. *Annu Rev Plant Phys*, 49: 281–309, 1998.

Chaffey, N. Cambium: old challenges - new opportunities. *Trees-Struct Funct*, 13: 138-151, 1999.

Cheng, S. H.; Keller, B. & Condit, C. M. Common occurrence of homologues of petunia glycine-rich protein-1 among plants. *Plant Mol Biol*, 31: 163-168, 1996.

Chiang, V. L. Monolignol biosynthesis and genetic engineering of lignin in trees, a review. *Environ Chem Lett*, 4: 143–146, 2006.

Clarke, J. D. & Zhu, T. Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems: practical considerations and perspectives. *Plant J*, 45: 630–650, 2006.

Colodette, J. L.; Mokfienski, A.; Gomide, A. & Oliveira, R. Relative importance of wood density and carbohydrate content on pulping yield and product quality. *J Tianjin Univ Sci Technol*, 19(4): 71-80, 2004.

Dallas, P. B.; Gottardo, N. G.; Firth, M. J.; Beesley, A. H.; Hoffmann, K.; Terry, P. A.; Freitas, J. R.; Boag, J. M.; Cummings, A. J. & Kees, U. R. Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR – how well do they correlate? *BMC Genomics*, 6: 59, 2005.

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. *P Natl Acad Sci USA*, 99(24): 15794–15799, 2002.

Doblin, M.; Kurek, I.; Jacob-Wilk, D. & Delmer, D. P. Cellulose biosynthesis in plants: from genes to rosettes. *Plant Cell Physiol*, 43: 1407–1420, 2002.

Eisen, M. B.; Spellman, P. T.; Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, 95: 14863-14868, 1998.

Emmanuel, E. & Levy, A. A. Tomato mutants as tools for functional genomics. *Curr Opin Plant Biol*, 5: 112–117, 2002.

FAO – Food and Agriculture Organization. *Forest products and services*.

Disponível em: < <http://www.fao.org/forestry/index.jsp> > acesso em cinco de setembro de 2007.

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 Phytol*, 170: 739–752, 2006.

Fukuda, H. Tracheary element differentiation. *The Plant Cell*, 9: 1147-1156, 1997.

Gachon, C.; Mingam, A. & Charrier, B. Real-time PCR: what relevance to plant studies? *J Exp Bot*, 55(402): 1445-1454, 2004.

Girke, T.; Todd, J.; Ruuska, S.; White, J.; Benning, C. H. & Ohlrogge, J. Microarray analysis of developing *Arabidopsis* seed. *Plant Physiol*, 124: 1570-1581, 2000.

Gittins, J. R.; Pellny, T. K.; Biricolti, S.; Hiles, E. R.; Passey, A. J. & James, D. J. Transgene expression in the vegetative tissues of apple driven by the vascular-specific rolC and CoYMV promoters. *Transgenic Res*, 12: 391-402, 2003.

Goes da Silva, F.; Iandolino, A.; Al-Kayal, F.; Bohlmann, M. C.; Cushman, M. A.; Lim, H.; Ergul, A.; Figueroa, R.; Kabuloglu, E. K.; Osborne, C.; Rowe, J.; Tattersall, E.; Leslie, A.; Xu, J.; Baek, JM.; Cramer, G. R.; Cushman J. C. & Cook, D. R. Characterizing the grape transcriptome. Analysis of expressed sequence tags from multiple *Vitis* species and development of a compendium of gene expression during berry development. *Plant Physiol*, 139: 574-597, 2005.

Grima-Pettanti, J. & Goffner, D. Lignin genetic engineering revisited. *Plant Sci*, 145: 51-65, 1999.

Groover, A. T. What genes make a tree a tree? *TRENDS Plant Sci*, 10: 210-214, 2005.

Groover, A. & Robischon, M. Developmental mechanisms regulating secondary growth in woody plants. *Curr Opin Plant Biol*, 9: 55-58, 2006.

Haberer, G.; Young, S.; Bharti, A. K.; Gundlach, H.; Raymond, C.; Fuks, G.; Butler, E.; Wing, R. A.; Rounsley, S.; Birren, B.; Nusbaum, C.; Mayer, K. F. X. & Messing, J. Structure and architecture of the maize genome. *Plant Physiol*, 139: 1612-1624, 2005.

Handl, J.; Knowles, J. & Kell, D. B. Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21: 3201-3212, 2005.

Hertzberg, M.; Aspeborg, H.; Schrader, J.; Andersson, A.; Erlandsson, R.; Blomqvist, K.; Bhalerao, R.; Uhlén, M.; Teeri, T. T.; Lundeberg, J.;

Sundberg, B.; Nilsson, P. & Sandberg, G. A transcriptional roadmap to wood formation. *P Natl Acad Sci USA*, 98(25): 14732–14737, 2001.

IPEF- Instituto de Pesquisas e Estudos Florestais. *Chave de identificação de Espécies Florestais*. Piracicaba-SP. 2004. Disponível em: <<http://www.ipef.br/identificacao/cief/especies/grandis.asp>>. acesso em dezanove de dezembro 2007.

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 Mol Biol*, 52: 893–903, 2003.

Irizarry, R. A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y. D.; Antonellis, K. J.; Scherf, U. & Speed, T. P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4: 249–264, 2003.

Jaillon, O.; Aury, J. M.; Noel, B.; Policriti, A.; Clepet, C.; Casagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C.; Vezzi, A.; Legeai, F.; Hugueney, P.; Dasilva, C.; Horner, D.; *et al.* French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449: 463-467, 2007.

Joshi, C. P.; Bhandari, S.; Ranjan, P.; Kalluri, U. C.; Liang, X.; Fujino, T. & Samuga, A. Genomics of cellulose biosynthesis in poplars. *New Phytol*, 164: 53–61, 2004.

Keller, B.; Sauer, N. & Lamb, C. J. Glycine-rich cell wall proteins in bean: gene structure and association of the protein with the vascular system. *EMBO J*, 7: 3625–3633, 1988.

Ko, J. & Han, K. *Arabidopsis* whole-transcriptome profiling defines the features of coordinated regulations that occur during secondary growth. *Plant Mol Biol*, 55: 433–453, 2004.

Kubo, M.; Udagawa, M.; Nishikubo, N.; Horiguchi, G.; Yamaguchi, M.; Ito, J.; Mimura, T.; Fukuda, H. & Demura, T. Transcription switches for protoxylem and metaxylem vessel formation. *Gene Dev*, 19: 1855–1860, 2005.

Lawrence, C. J.; Dong, Q.; Polacco, M. L.; Seigfred, T. E. & Brendel, V. MaizeGDB, the community database for maize genetics and genomics. *Nucl Acids Res*, 32: D393–D397, 2004.

Lawton, M. A.; Tierney, M. A.; Nakamura, I.; Anderson, E.; Komeda, Y.; Dube, P.; Hoffman, N.; Fraley, R. T. & Beachy, R. N. Expression of a soybean beta-concglycinin gene under the control of the cauliflower mosaic virus-35s and 19s-promoters in transformed petunia tissues. *Plant Mol Biol*, 9: 315–324, 1987.

Li, L.; Lu, S. & Chiang, V. A Genomic and molecular view of wood formation. *Critical Rev Plant Sci*, 25(3): 215-233, 2006.

Li, L.; Osakabe, Y.; Joshi, C. P. & Chiang, V. L. Secondary xylem-specific expression of caffeoyl-coenzyme A 3-O-methyltransferase plays an important role in the methylation pathway associated with lignin biosynthesis in loblolly pine. *Plant Mol Biol*, 40: 555-565, 1999.

Liu, Z.; Wang, J.; Huang, X.; Xu, W.; Liu, Z. & Fang, R. The promoter of a rice glycine-rich protein gene, *Osgrp-2*, confers vascular-specific expression in transgenic plants. *Planta*, 216: 824-833, 2003.

Lorenz, W. W. & Dean, J. F. D. SAGE profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). *Tree Physiol*, 22: 301–310, 2002.

Lu, H.; Zeng, Q.; Zhao, Y.; Wang, S. & Jiang, X. Xylem-specific expression of a GRP1.8 promoter::4CL gene construct in transgenic tobacco. *Plant Growth Regul* 41: 279-286, 2003.

Lu, H.; Zhao, Y. & Jiang, X. Stable and specific expression of 4-coumarate:coenzyme A ligase gene (4CL1) driven by the xylem-specific Pto-4CL1 promoter in the transgenic tobacco. *Biotechnol Lett*, 26: 1147-1152, 2004.

Matsui, A.; Yokoyama, R.; Seki, M.; Ito, T.; Shinozaki, K.; Takahashi, T.; Komeda, Y. & Nishitani, K. AtXTH27 plays an essential role in cell wall modification during the development of tracheary elements. *Plant J*, 42(4): 525–534, 2005.

Matsuyama, T.; Satoh, H.; Yamada, Y. & Hashimoto, T. A maize glycine-rich protein is synthesized in the lateral root cap and accumulates in the mucilage. *Plant Physiol*, 120: 665–674, 1999.

MeV Manual - MultiExperiment Viewer, Version 4.0, 276p, 2006.
<<http://mev.tm4.org>>

Neilson, D. The global *Eucalyptus* resource and some solidwood-panel product developmental issues. IUFRO Conference, Launceston, Austrália, 2000.

No, E.; Zhou, Y. & Loopstra, C. Sequences upstream and downstream of two xylem-specific pine genes influence their expression. *Plant Sci*, 160: 77–86, 2000.

Nothnagel, E. A. Proteoglycans and related components in plant cells. *Int Rev Cytol*, 174: 95–291, 1997.

Ohashi-Ito, K. & Fukuda, H. HD-Zip III homeobox genes that include a novel member, ZeHB-13 (*Zinnia*)/ATHB-15 (*Arabidopsis*) are involved in procambium and xylem cell differentiation. *Plant Cell Physiol*, 44(12): 1350–1358, 2003.

Oh, S.; Park, S. & Han, K. Transcriptional regulation of secondary growth in *Arabidopsis thaliana*. *J Exp Bot*, 54(393): 2709-2722, 2003.

Patterson, B. D & Smillie, R. M. Developmental changes in ribosomal ribonucleic acid and fraction I protein in wheat leaves. *Plant Physiol*, 47: 196-198, 1971.

Paux, E.; Tamasloukht, M.; Ladouce, N.; Sivadon, P. & Grima-Pettenati, J. Identification of genes preferentially expressed during wood formation in *Eucalyptus*. *Plant Mol Biol*, 55: 263–280, 2004.

Paux, E.; Carocha, V.; Marques, C.; Sousa, A. M.; Borralho, N.; Sivadon, P. & Grima-Pettenati, J. Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *New Phytol*, 167: 89-100, 2005.

Pavy, N.; Johnson, J. J.; Crow, J. A.; Paule, C.; Kunau, T.; MacKay, J. & Retze, E. F. ForestTreeDB: a database dedicated to the mining of tree transcriptomes *Nucleic Acids Res*, 35: D888–D894, 2007.

Pereira, J. C. D.; Sturion, J. A.; Higa, A. R.; Higa, R. C. V. & Shimizu, J. Y. *Características da madeira de algumas espécies de eucalipto plantadas no Brasil*. In: Embrapa Florestas Documentos 38: 113p, 2000.

Persson, S.; Wei, H.; Milne, J.; Page, G. P. & Somerville, C. R. Identification of genes required for cellulose synthesis by regression

analysis of public microarray data sets. *P Natl Acad Sci USA*, 102(24): 8633–8638, 2005.

Plomion, C.; Leprovost, G. & Stokes, A. Wood formation in trees. *Plant Physiol*, 127: 1513–1523, 2001.

Poke, F. S.; Vaillancourt, R. E.; Potts, B. M. & Reid, J. B. Genomic research in *Eucalyptus*. *Genetica*, 125: 79–101, 2005.

Prassinis, C.; Ko, J.; Yang, J. & Han, K. Transcriptome profiling of vertical stem segments provides insights into the genetic regulation of secondary growth in hybrid aspen trees. *Plant Cell Physiol*, 46(8): 1213–1225, 2005.

Puthoff, D. P.; Nettleton, D.; Rodermeil, S. R. & Baum, T. J. *Arabidopsis* gene expression changes during cyst nematode parasitism revealed by statistical analyses of microarray expression profiles. *Plant J*, 33: 911–921, 2003.

Rafalski, A. Tagging the rice transcriptome. *Nat Biotechnol*, 25: 430–431, 2007.

Revista da Madeira, Curitiba, REMADE - Portal Nacional da madeira. No eucalipto a opção de futuro. n107, 2007. Disponível em: <<http://www.remade.com.br/pt/revista.php?ano=2007>> Acesso em doze de dezembro de 2007.

Schaffer, R.; Landgraf, J.; Accerbi, M.; Simon, V. V.; Larson, M. & Wisman, E. Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant Cell*, 13: 113–123, 2001.

Schena, M.; Shalon, D.; Davisl R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270: 467–470, 1995.

Schrader, J.; Nilsson, J.; Mellerowicz, E.; Berglund, A.; Nilsson, P.; Hertzberg, M. & Sandberga, G. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *Plant Cell*, 16: 2278–2292, 2004.

Smart, L. B.; Nall, N. M. & Bennett, A. B. Isolation of RNA and protein from guard cells of *Nicotiana glauca*. *Plant Mol Biol Rep*, 17: 371–383, 1999.

Sociedade Brasileira de Silvicultura. *Fatos e Números do Brasil Florestal*. 108p. São Paulo, 2006. Disponível em: <<http://www.sbs.org.br/FatoseNumerosdoBrasilFlorestal.pdf>> acesso em onze de julho de 2007.

Stalberg, K.; Ellerstrom, M.; Sjodahl, S.; Ezcurra, I.; Wycliffe, P. & Rask, L. Heterologous and homologous transgenic expression directed by a 2S seed storage promoter of *Brassica napus*. *Transgenic Res*, 7: 165–172, 1998.

Tattersall, E. A. R.; Ergul, A.; AlKayal, F.; DeLuc, L.; Cushman, J. C. & Cramer, J. R. Comparison of methods for isolating high-quality RNA from leaves of grapevine. *Am J Enol Viticult*, 56: 400-406, 2005.

Trainotti, L.; Bonghi, C.; Ziliotto, F.; Zanin, D.; Rasori, A.; Casadoro, G.; Ramina, A. & Tonutti, P. The use of microarray μ PEACH1.0 to investigate transcriptome changes during transition from pre-climacteric to climacteric phase in peach fruit. *Plant Sci*, 170: 606-613, 2006.

Tuskan, G. A.; DiFazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, N.; Ralph, S.; Rombauts, S.; Salamov, A.; Schein, J.; Sterck, L.; Aerts, A.; Bhalerao, R. R.; Bhalerao, R. P.; Blaudez, D.; *et al.* The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604, 2006.

Tusher, V. G.; Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *P Natl Acad Sci USA*, 98(9): 5116–5121, 2001.

Velasco, R.; Zharkikh, A.; Troggio, M.; Cartwright, D. A.; Cestaro, A.; Pruss, D.; Pindo, M.; Fitzgerald, L. M.; Vezzulli, S.; Reid, J.; Malacarne, G.; Iliev, D.; Coppola, G.; Wardell, B.; Micheletti, D.; Macalma, T.; *et al.* A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE*, 2(12): e1326. doi:10.1371/journal.pone.0001326, 2007.

Vettore, A. L.; da Silva, F. R.; Kemper, E. L.; Souza, G. M.; da Silva, A. M.; Ferro, M. I. T.; Henrique-Silva, F.; Giglioti, E. A.; Lemos, M. V. F.; Coutinho, L. L. *et al.* Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res*, 13: 2725–2735, 2003.

- Wang, R. C.; Okamoto, M.; Xing, X. J. & Crawford, N. M. Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol*, 132: 556–567, 2003.
- Wang, T.; Zhang, N. & Du, L. Isolation of RNA of high quality and yield from *Ginkgo biloba* leaves. *Biotechnol Lett*, 27: 629–633, 2005.
- Waters, D. L. E.; Holton, T. A.; Ablett, E. M.; Lee, L. S. & Henry, R. J. The ripening wine grape berry skin transcriptome. *Plant Sci*, 171(1): 132-138, 2006.
- Whetten, R.; Sun, Y. H.; Zhang, Y. & Sederoff, R. Functional genomics and cell wall biosynthesis in loblolly pine. *Plant Mol Biol*, 47: 275–291, 2001.
- Whetten, R. W.; MacKay, J. J. & Sederoff, R. R. Recent advances in understanding lignin biosynthesis. *Annu Rev Plant Phys*, 49: 585–609, 1998.
- White, J. A.; Todd, J.; Newman, T.; Focks, N.; Girke, T.; Martinez de Ilarduya, O.; Jaworski, J. G.; Ohlrogge, J. B. & Benning, C. H. A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiol*, 124: 1582–1594, 2000.
- Wissenbach, M.; Uberlacker, B.; Vogt, F.; Becker, D.; Salamini, F. & Rohde, W. Myb genes from *Hordeum vulgare*: tissue specific expression of chimeric Myb promoter/Gus genes in transgenic tobacco. *Plant J*, 4: 411–422, 1993.
- Xue, Y.; Li, J. & Xu, Z. Recent highlights of the China Rice Functional Genomics Program. *Trends Genet*, 19: 390–394, 2003.
- Yamamoto, K. & Sasaki, T. Large-scale EST sequencing in rice. *Plant Mol Biol*, 35: 135–144, 1997.
- Yang, J.; Park, S.; Kamdem, D. P.; Keathley, D. E.; Retzel, E.; Paule, C.; Kapur, V. & Han, K. 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 Mol Biol*, 52: 935–956, 2003.

Yang, S.; Zyl, L.; No, E. & Loopstra, C. A. Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*). *Plant Sci*, 166: 1185–1195, 2004.

Yazaki, J.; Kojima, K.; Suzuki, K.; Kishimoto, N. & Kikuchi, S. The Rice PIPELINE: a unification tool for plant functional genomics. *Nucleic Acids Res*, 32: D383-D387, 2004.

Ye, Z-H. Vascular tissue differentiation and pattern formation in plants. *Annu Rev Plant Biol*, 53: 183–202, 2002.

Zhao, C.; Craig, J. C.; Petzold, H. E.; Dickerman, A. W. & Beers, E. P. The xylem and phloem transcriptomes from secondary tissues of the *Arabidopsis* root-hypocotyl. *Plant Physiol*, 138: 803–818, 2005.